Weak Identification with Many Instruments

Anna Mikusheva[1] Liyang Sun[2]

[1]MIT
[2]CEMFI

April, 2023

Introduction

- We consider IV (instrumental variable) models with *many weak* instruments
  - Estimation with many instruments
  - How to determine that instruments are weak?
  - Weak identification robust inferences (some new results here)
  - Open questions

Introduction

- **Example 1**: Angrist and Krueger (1991)

$$wage_i = \beta \; education_i + controls + e_i,$$

- Instrument is quarter of birth
- First stage is heterogeneous: law depends on state and birth cohort
- Instruments used: QOB ($\times$ state dummy) ($\times$ year dummy)
    - year of birth (30)
    - year and state of birth (180)
    - year and state of birth, and their interactions (1530)
- Staiger and Stock (1997)- IV may be weak
- Hansen et al. (2008)- instruments are many

Introduction

- **Example 2**: 'Judges design'
- Bhuller, Dahl, Loken and Mogstad (JPE, 2020): "Incarceration, Recidivism, and Employment"

$$\mathit{recidivism}_i = \beta \;\; \mathit{incarceration}_i + \mathit{controls} + e_i,$$

- Instruments: "judge stringency" = the average incarceration rate in other cases a judge has handled
- This is a form of JIVE with instrument-dummies for judge assignment
- Sample size is roughly proportional to the number of judges

Setup

- Linear IV model with one endogenous variable:

$$\begin{cases} Y_i = \beta X_i + (\delta W_i) + e_i \\ X_i = \pi' Z_i + (\gamma W_i) + v_i \end{cases}$$

  where $Z_i \in \mathbb{R}^K$ s.t. $\mathbb{E}[e_i|Z_i, W_i] = \mathbb{E}[v_i|Z_i, W_i] = 0$
- Data is i.i.d., $i = 1, ..., N$
- Many instruments: $K \to \infty$ as $N \to \infty$ (up to $K = \lambda N$)
- Weak instruments: π is small in some sense
- For most results errors are heteroskedastic

Outline

1. Estimation
2. Weak Identification: detection
3. Weak IV robust inferences
   - AR test
   - Other tests: LM
4. Adding covariates
5. Conclusions and Open questions

Overview

1. Estimation
2. Weak Identification: detection
3. Weak IV robust inferences
4. Adding covariates
5. Conclusions and Open questions

Setup

- Assume away covariates (we will add them in the last section)

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

  where $Z_i \in \mathbb{R}^K$ s.t. $\mathbb{E}[e_i|Z_i] = \mathbb{E}[v_i|Z_i] = 0$
- Data is i.i.d., $i = 1, ..., N$
- For most results errors are heteroskedastic

TSLS

Most commonly known estimator is Two-Stage Least Squares (TSLS)

- First stage- finding the optimal instrument = best predictor

$$\widehat{X}_i = \widehat{\pi} Z_i$$

- Second stage: estimate structural equation using $\widehat{X}_i$ as the instruments
- Optimal instrument under homoskedasticity: $\mathbb{E}[X_i | Z_i]$ (Chamberlain, 1987)
- Concentration parameter $\frac{\pi' Z' Z \pi}{\sigma_v^2}$ plays as effective sample size (Stock and Yogo, 2005)

Estimation with Many IV

- First stage: $X_i = \pi' Z_i + v_i$
- If many regressors in the first stage, they might ‘overfit’ the noise
- Estimated optimal instrument is endogenous $\mathbb{E}[\widehat{X}_i e_i] \neq 0$
- For homoscedastic TSLS: $\widehat{X}_i = \pi' Z_i + v' Z (Z'Z)^{-1} Z_i$

$$\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N} \widehat{X}_i e_i\right] = K\sigma_{ev}$$

- Endogeneity is growing in K, leads to bias
- Bias of the IV estimator increases with the number of moment conditions/instruments (Bekker, 1994, Newey and Smith, 2004)

Estimation with Many IV

Suggestions on how to remove endogeneity:

- Sample splitting (Angrist and Krueger, 1995):
  - split sample to halves
  - select/estimate optimal instrument on one half
  - estimate β on the other half
- Jackknife (Angrist et al., 1999)
  - estimate optimal instrument for observation i on sample excluding i
  - use estimated optimal instrument

Estimation with Many IV

- $\widehat{\beta}_{TSLS} - \beta = \frac{X' P_Z e}{X' P_Z X}$, where $P_Z = Z(Z'Z)^{-1} Z'$
- Bias comes from $\mathbb{E}[X' P_Z e] = \mathbb{E}[v' P_Z e] = \sum_i P_{ii} \mathbb{E}[v_i e_i]$ the diagonal of the projection matrix, $trace(P_Z) = K$
- Idea: remove the diagonal

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

- It is very close to jackknife (numerical differences are tiny)
- Diagonal removal can be done to many estimators: JIVE-LIML and JIVE-Fuller (Hausman et al., 2012), JIVE-ridge (Hansen and Kozbur, 2014)

Estimation with Many IV

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

- TSLS is consistent when $\frac{\pi' Z' Z \pi}{K} \to \infty$ (Chao and Swanson, 2005)
- When $\frac{\pi' Z' Z \pi}{\sqrt{K}} \to \infty$, JIVE, JIVE-Fuller and JIVE-LIML are consistent (Hausman et al, 2012)
- When $\frac{\pi' Z' Z \pi}{\sqrt{K}} \to \infty$, JIVE, JIVE-Fuller and JIVE-LIML are asymptotically gaussian
    - Wald confidence sets and t-statistics can be used
    - Estimation of standard errors is non-trivial (Hausman et al, 2012)

Estimation with Many IV: Summary

- Many instruments can be hurtful if they do not extract additional information from the first stage
- Over-fitting creates a bias
- One should avoid using TSLS with many instruments
- Jack-knifing or diagonal removal is very fruitful idea

Overview

What is Weak Identification?

- If $\frac{\pi' Z' Z\pi}{\sqrt{K}} \to \infty$, then JIVE or JIVE-LIML are consistent and asymptotically gaussian
- What if there are better estimators (work well for weaker cases)?
  - 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 (Mikusheva and Sun, 2022)
- How to know in practice if $\frac{\pi' Z' Z\pi}{\sqrt{K}}$ is large enough to trust Wald confidence sets?

What is Weak Identification?

- Mikusheva and Sun (2022): pre-test for weak identification
  - Size distortions of JIVE-Wald depend on $\frac{\mu^2}{\Upsilon\sqrt{K}}$, where $\mu^2 = \sum_{i \neq j} P_{ij}(\pi' Z_i)(\pi' Z_j)$ and Υ- measure of the first stage uncertainty
  - Derived a cut-off, if $\frac{\mu^2}{\Upsilon\sqrt{K}}$ is above cut-off, then JIVE- Wald test has small size distortion
  - Have estimator $\widetilde{F}$ for $\frac{\mu^2}{\Upsilon\sqrt{K}}$
  - Got a cut-off for $\widetilde{F}$

Weak Identification: detection

- Our pre-test is based on the 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 $\widehat{\Upsilon}$ is an estimate of uncertainty in the first stage
- If $\widetilde{F} > 4.14$, then the JIVE- Wald test has less than 10 % size distortion
- Suggestion: if $\widetilde{F}$ is low, one should use "robust" tests
- Stata package implementing pre-test and robust tests: `manyweakiv` (beta version)

Re-visiting Angrist and Krueger (1991)

- Research question: returns 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	Robust AR	Robust LM
180 instruments	2.4	13.4	[0.066,0.132]	[0.008,0.201]	[0.067,0.135]
1530 instruments	1.3	6.2	[0.024,0.121]	[-0.047, 0.202]	[0.022,0.127]

Table: Angrist and Krueger (1991) 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 sets are constructed via analytical test inversion.

Overview

Weak IV-Robust Tests: Refresher, Fixed K

- $Y_i = \beta X_i + e_i$, Z_i-instrument ($\mathbb{E}[e_i|Z_i] = 0$)
- $H_0 : \beta = \beta_0$. 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

What Changes with $K \to \infty$?

- Homoskedastic AR statistics for fixed K:

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

- χ^2_K is a diverging distribution for large K
- $e(\beta_0)' P_Z e(\beta_0)$ has a non-zero mean $\mathbb{E} e' P_Z e = \sum_{i=1}^N P_{ii} \mathbb{E} e_i^2$
- Idea: remove the diagonal $\sum_{i \neq j} e_i(\beta_0) P_{ij} e_j(\beta_0)$
- Use CLT for quadratic forms (U-statistics)

AR test with many instruments

- 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 $\Phi_0 = \frac{2}{K} \sum_{i \neq j} P_{ij}^2 \sigma_i^2 \sigma_j^2$
- Under $H_0 : \beta = \beta_0$ we have $AR_0(\beta_0) \Rightarrow N(0, 1)$
- Need $K \to \infty$ for asymptotic distribution
- 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 (Crudu et al, 2021): $\widehat{\sigma}_i^2 = e_i^2(\beta_0)$
- It gives correct size, robust toward heteroscedasticity, but power is problematic at distant alternatives
- Residualizing $e(\beta_0)$ with respect to Z ($M = I - P_Z$)

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

- "Cross-fit" variance estimator (Newey et al, 2018; Kline et al., 2020):

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

AR: Variance Estimation

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

- Use proxy $\widehat{\sigma}_i^2 = \frac{1}{1-P_{ii}} e_i(\beta_0) M_i e(\beta_0)$
- Challenge is that we need a double sum:

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

- Our suggested estimator (Mikusheva and Sun, 2022):

$$\widehat{\Phi}_2 = \frac{2}{K} \sum_{i \neq j} \frac{P_{ij}^2}{M_{ii} M_{jj} + M_{ij}^2} \left[e_i(\beta_0) M_i e(\beta_0)\right] \left[e_j(\beta_0) M_j e(\beta_0)\right]$$

AR: Variance Estimation

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

- Anatolyev and Sølvsten (2020): one may get unbiased proxy for $\sigma_i^2 \sigma_j^2$ by using "leave-three-out" estimator

$$\widehat{\sigma_i^2 \sigma_j^2} = e_i(\beta_0) e_j(\beta_0) \sum_k \tilde{M}_{ik,-(ij)} e_k(\beta_0)[e_j(\beta_0) - Z_j' \widehat{\delta}_{-(ijk)}]$$

  where $\widehat{\delta}_{-(ijk)}$ is OLS from regressing $e(\beta_0)$ on Z leaving three observations (i, j, k) out

- There are explicit formulas for "leave-(one/two/three)-out" available, but numerical complexity increases

Feasible AR

$AR(\beta_0) = \frac{1}{\sqrt{K\widehat{\Phi}}} \sum_{i \neq j} e_i(\beta_0) P_{ij} e_j(\beta_0)$ rejects when $AR(\beta_0) > z_{1-\alpha}$

- Where $\hat{\Phi}$ can be:
  - Crudu et al (2021): $\widehat{\Phi}_1 = \frac{2}{K} \sum_{i \neq j} P_{ij}^2 e_i(\beta_0)^2 e_j(\beta_0)^2$
  - Mikusheva and Sun (2022):
    $\widehat{\Phi}_2 = \frac{2}{K} \sum_{i \neq j} \frac{P_{ij}^2}{M_{ii} M_{jj} + M_{ij}^2} \left[e_i(\beta_0) M_i e(\beta_0)\right] \left[e_j(\beta_0) M_j e(\beta_0)\right]$
  - Anatolyev and Sølvsten: $\widehat{\Phi}_3$ leave-three-out
- All three $\widehat{\Phi}$ are consistent for Φ_0 under the null
- Feasible tests with $\widehat{\Phi}_2$ and $\widehat{\Phi}_3$ achieve the same local power as the infeasible AR
- Feasible tests with $\widehat{\Phi}_2$ and $\widehat{\Phi}_3$ are consistent for distant alternatives
- Variance estimators are ordered in terms of increasing computational complexity

Power of AR

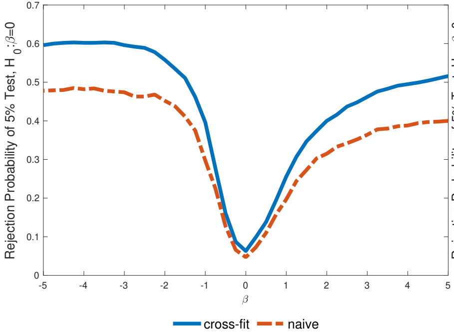
- Under the alternative $\beta = \beta_0 + \Delta$
- Define a leave-one-out information in the sample:

$$\mu^2 = \sum_{i \neq j} P_{ij}(\pi Z_i)(\pi Z_j) \asymp \pi' Z' Z \pi$$

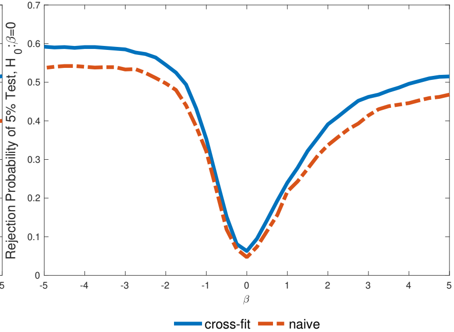
- Power statement: uniformly over a set of local alternative and a (reasonably restricted) set of μ^2 :

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

Variance Estimator is Important



(a) sparse, $\frac{\mu^2}{\sqrt{K}} = 2.5$



(b) dense, $\frac{\mu^2}{\sqrt{K}} = 2.5$

Power curves for leave-one-out AR tests with $\widehat{\Phi}_1$ (red dash) and $\widehat{\Phi}_2$ (blue line) and variance estimators under sparse vs. dense first stage. Instruments are $K = 40$ balanced group indicators, $N = 200$, based on 1,000 simulations

Weak IV-Robust Tests: LM

- Problem: AR is not efficient if identification is strong
- AR uses all instruments "equally"
- LM intends to test a "powerful" combination of instruments $e'Z\pi$,
- Idealistic LM is based on the linear combination $e'(\beta_0)Z\widehat{\pi} = e'(\beta_0)P_Z X$
- Leave-one-out gives us $LM^{1/2} \propto \sum_{i \neq j} e_i(\beta_0)P_{ij}X_j$

Robust LM

- The infeasible leave-one-out LM is

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

- Under $H_0 : \beta = \beta_0$ we have $LM^{1/2}(\beta_0) \Rightarrow N(0,1)$ as $N, K \to \infty$
- Reject when $\left| LM^{1/2}(\beta_0) \right|$ is large (two-sided test)
- Need an estimator for

$$\Psi = \frac{1}{K} \sum_{i=1}^{N} (\sum_{j \neq i} P_{ij} X_j)^2 \sigma_i^2 + \frac{1}{K} \sum_{i=1}^{N} \sum_{j \neq i} P_{ij}^2 \gamma_i \gamma_j,$$

$\sigma_i^2 = \mathbb{E} e_i^2, \gamma_i = \mathbb{E}[X_i e_i]$

LM: variance estimation

- Matsushita and Otsu (2022):

$$\widehat{\Psi}_1 = \frac{1}{K}\sum_i \widehat{\sigma}_i^2 (\sum_{j\neq i} P_{ij}X_j)^2 + \frac{1}{K}\sum_i \sum_{j\neq i} P_{ij}^2 \widehat{\gamma}_i \widehat{\gamma}_j,$$

  where $\widehat{\sigma}_i^2 = e_i^2(\beta_0)$ and $\widehat{\gamma}_i = X_i e_i(\beta_0)$
- $\widehat{\Psi}_1$ is consistent under the null $(H_0 : \beta = \beta_0)$
- There is a loss of power at alternative: $e(\beta_0) \neq e_i$ and $\widehat{\sigma}_i^2$ overstate the variance

LM: variance estimation (new result)

- We propose to use the ideas of double-cross-fit

$$\widehat{\Psi}_2 = \frac{1}{K}\sum_i \frac{e_i M_i e}{M_{ii}}(\sum_{j\neq i} P_{ij}X_j)^2 + \frac{1}{K}\sum_i\sum_{j\neq i}\widetilde{P}_{ij}^2 X_i M_i e X_j M_j e$$

- Here we use
  - $\widehat{\sigma}_i^2 = \frac{e_i M_i e}{M_{ii}}$ – unbiased proxy for σ_i^2
  - $\widehat{\gamma}_i = X_i M_i e$ – proxy for γ_i
  - re-weighting $\widetilde{P}_{ij}^2 = \frac{P_{ij}^2}{M_{ii}M_{jj}+M_{ij}^2}$ to correct for correlation in proxies

LM: variance estimation (new result)

- We propose to use the ideas of double-cross-fit

$$\widehat{\Psi}_2 = \frac{1}{K}\sum_i \frac{e_i M_i e}{M_{ii}}\Big(\sum_{j\neq i} P_{ij}X_j\Big)^2 + \frac{1}{K}\sum_i \sum_{j\neq i} \widetilde{P}_{ij}^2 X_i M_i e X_j M_j e$$

- Under assumptions needed for CLT and stricter moment condition, our estimator
  - Consistent under the null (when $\beta = \beta_0$)
  - Consistent for local alternative (when $\beta = \beta_0 + \Delta$ and $\Delta^2 \frac{\mu^2}{K} \to 0$)
  - Inconsistent for global alternatives but still deliver consistent LM test
- Leave-one-out LM with our variance estimation has the same power curves as "infeasible" LM test

Power of LM

- The infeasible leave-one-out LM is

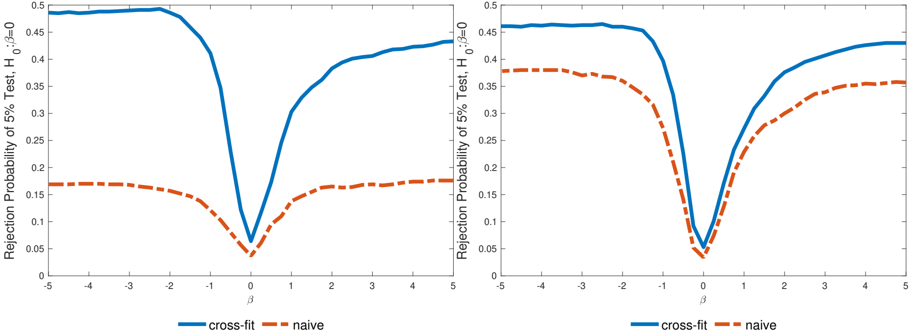
$$LM^{1/2}(\beta_0) = \frac{1}{\sqrt{K\Psi}} \sum_{i \neq j} e_i(\beta_0) P_{ij} X_j$$

- 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

Variance Estimator is Important



(c) sparse, $\frac{\mu^2}{\sqrt{K}} = 2.5$ (d) dense, $\frac{\mu^2}{\sqrt{K}} = 2.5$

Power curves for leave-one-out LM with $\widehat{\Psi}_1$ (red dash) and $\widehat{\Psi}_2$ (blue line) variance estimators under sparse vs. dense first stage. Instruments are $K = 40$ balanced group indicators, $N = 200$

Re-visiting Angrist and Krueger (1991)

	FF	$\widetilde{F}$	JIVE-Wald	Robust AR	Robust LM
180 instruments	2.4	13.4	[0.066,0.132]	[0.008,0.201]	[0.067,0.135]
1530 instruments	1.3	6.2	[0.024,0.121]	[-0.047, 0.202]	[0.022,0.127]

Table: Angrist and Krueger (1991) 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 sets are constructed via analytical test inversion.

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

- We may think about combining three statistics optimally

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

- AR and LM are for testing β_0 and $\widetilde{F}$ for assessing the strength of identification
- Lim, Wang and Zhang (2022) - suggests an optimal combination test
- Ayyar, Matsushita and Otsu (2022) - suggestions on how to build CLR test

Overview

Adding covariates: what is the problem?

- Linear IV model with one endogenous variable:

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

where $Z_i \in \mathbb{R}^K$ s.t. $\mathbb{E}[e_i | Z_i, W_i] = \mathbb{E}[v_i | Z_i, W_i] = 0$

- When there are no covariates (W_i) the bias was removed by removing a diagonal (JIVE)

$$\widehat{\beta}_{JIV} = \frac{\sum_{i \neq j} X_i P_{ij} Y_j}{\sum_{i \neq j} X_i P_{ij} X_j'} = \frac{X' P_{JIV} Y}{X' P_{JIV} X}$$

we denote $A_{JIV} = \{A_{ij} \cdot 1\{i \neq j\}\}$ for any matrix

- Could we do a similar thing: partial out covariates and remove the diagonal from P_Z?

Adding covariates: what is the problem?

- Let $M_W = I - W(W'W)^{-1}W'$ be partialling out operator
- $Y^{\perp} = M_W Y$, $X^{\perp} = M_W X$, $Z^{\perp} = M_W Z$, $P^{\perp} = P_{Z^{\perp}}$,
- Would the following estimator work?

$$\widehat{\beta} = \frac{\sum_{i \neq j} X_i^{\perp} P_{ij}^{\perp} Y_j^{\perp}}{\sum_{i \neq j} X_i^{\perp} P_{ij}^{\perp} Y_j^{\perp}} = \frac{(X^{\perp})' P_{JIV}^{\perp} Y^{\perp}}{(X^{\perp})' P_{JIV}^{\perp} X^{\perp}}$$

- No. This is the same as $\widehat{\beta} = \frac{X' M_W P_{JIV}^{\perp} M_W Y}{X' M_W P_{JIV}^{\perp} M_W X}$
- Matrix $M_W P_{JIV}^{\perp} M_W$ has a non-trivial diagonal and produces bias in the estimator

Adding covariates: what is the problem?

- What if we do this in opposite order:

$$\widehat{\beta} = \frac{X'(M_W P^{\perp} M_W)_{JIV} Y}{X'(M_W P^{\perp} M_W)_{JIV} X}$$

- It does not work either

$$(M_W P^{\perp} M_W)_{JIV} W \neq 0$$

it loses partialling out property

Adding covariates: estimation

- Solution proposed in Chao, Swanson and Woutersen (2023): find $\theta_1, ..., \theta_n$ and diagonal matrix D_θ:

$$M_W(P^\perp - D_\theta)M_W \text{ has zero diagonal}$$

  this problem is linear and solvable for well-balanced designs

- Suggested estimator

$$\widehat{\beta} = \frac{X' M_W (P^\perp - D_\theta) M_W Y}{X' M_W (P^\perp - D_\theta) M_W X}$$

- Chao, Swanson and Woutersen (2023) has proof of consistency and asymptotic gaussianity under some assumptions

Adding covariates: robust inference (new results)

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

- We can create a weak IV robust test for $H_0 : \beta = \beta_0$ using this idea

$$AR(\beta_0) = \frac{1}{\sqrt{K\Phi}}(Y - \beta_0 X)' M_W (P^{\perp} - D_{\theta}) M_W (Y - \beta_0 X)$$

- Under the null $AR(\beta_0) \Rightarrow N(0, 1)$, reject when $AR(\beta_0)$ is large

Adding covariates: robust inference (new results)

- Denote A_{ij} to be elements of $M_W(P^{\perp} - D_\theta)M_W$, then we can use

$$\widehat{\Phi} = \frac{2}{K}\sum_{i,j}\frac{A_{ij}^2}{M_{ii}M_{jj} + M_{ij}^2}\widehat{\sigma}_i^2\widehat{\sigma}_j^2$$

  where $\widehat{\sigma}_i^2 = \sum_k \frac{M_{ik}}{M_{ii}}(Y_i - \beta_0 X_i)(Y_k - \beta_0 X_k)$, while M_{ij} are elements of $M_{Z,W}$
- The "naive" variance estimator is inconsistent even under the null, since $(Y_i - \beta_0 X_i)^2$ overstates the variance drastically (it has part predictable by W_i)

Overview

Conclusions and Open Questions

- Many instruments come with costs - one needs to find an optimal way to combine them
- Uncertainty about the first stage produces biases of TSLS
- Jackknifing or deleting diagonal is productive idea for both estimation and inference
- The knife-edge case for consistency happens when $\frac{\pi' Z' Z \pi}{\sqrt{K}} \asymp const$
- There is a pre-test for weak identification robust to heteroscedasticity when $K \to \infty$, which depends on the estimator one uses with it
- Robust tests (AR and LM) use the idea of leave-one-out quadratic forms and cross-fit variance estimation
- Adding many covariates is non-trivial

Conclusions and open questions

- Open question: there is a pre-test for whether one can trust JIVE-Wald confidence set/ t-test. JIVE-LIML is more efficient (Hausman et al, 2012), but there is no pre-test for it
- Open question: there is no pre-test that accommodates many covariates either
- Open question: unclear what to do with inferences when there are multiple endogenous variables (sub-vector inference)
- Open question: many instruments framework accommodates well heterogeneous first stage, what to do about heterogeneous structural equation (non-parametric IV)

Reference

- Angrist, J.D., and Frandsen B. (2019). "Machine Labor" *NBER working paper* 26584
- Angrist, J.D., Imbens, G.W., and Krueger, A.B. (1999). "Jackknife instrumental variables estimation." *Journal of Applied Econometrics* 14, 57–67.
- Angrist, J.D., and Krueger, A.B. (1991). "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics* 106, 979–1014.
- Bekker, P.A. (1994). "Alternative Approximations to the Distributions of Instrumental Variable Estimators." *Econometrica* 62, 657–681.
- Chao, J.C., and Swanson, N.R. (2005). "Consistent Estimation with a Large Number of Weak Instruments." *Econometrica* 73, 1673–1692.

Reference

- Crudu, F., Mellace, G., and Sandor, Z. (2020). "Inference in Instrumental Variables Models with Heteroskedasticity and Many Instruments." *Econometric Theory*, forthcoming.
- Kline, P., Saggio, R., and Sølvsten, M. (2018). "Leave-out estimation of variance components."
- Stock, J.H., and Yogo, M. (2005). "Testing for weak instruments in Linear Iv regression. In Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg," pp. 80–108.
- Dennis Lim, Wenjie Wang, Yichong Zhang. (2022) "A Conditional Linear Combination Test with Many Weak Instruments"
- Sreevidya Ayyar, Yukitoshi Matsushita, Taisuke Otsu. (2022) "Conditional likelihood ratio test with many weak instruments"
- Matsushita, Y., and Otsu, T. (2022). "A JACKKNIFE LAGRANGE MULTIPLIER TEST WITH MANY WEAK INSTRUMENTS." *Econometric Theory*, 1-24.
- John C. Chao, Norman R. Swanson, Tiemen Woutersen, "Jackknife estimation of a cluster-sample IV regression model with many weak instruments," *Journal of Econometrics*, 2023