Sargan Lectures: Weakly Identified Econometric Models

Anna Mikusheva

MIT

June, 2024

1. Linear IV model with weak instruments
2. Many Instruments
3. Weak GMM and other structural models

Sargan Lecture 1
Weak Instruments

Anna Mikusheva

MIT

June, 2024

Outline

1. Classic asymptotic results
2. Weak IV phenomenon
3. Identification Robust Inference
   - Full parameter inference
   - Sub-vector inference
4. Estimation
5. Summary and Open Questions

Introduction

- Linear regressions with i.i.d. data:

$$Y_i = \beta' X_i + \gamma' W_i + e_i$$

  can be estimated by OLS when we have exogeneity
  - $\mathbb{E}[e_i(X_i'; W_i')'] = 0$
  - sometimes we assume $\mathbb{E}[e_i | X_i, W_i] = 0$
- In observational data exogeneity often does not hold
- **Example 1**: Angrist and Krueger (1991)

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

- Reasons for violations of exogeneity include missing confounding factors, simultaneity, mis-measurement

Introduction

$$Y_i = \beta' X_i + \gamma' W_i + e_i$$

- Common strategy in applied economics - Instrumental Variables
- Instrument Z_i is a variable the produces exogenous variation in the regressor (conditional on controls)
  - Z moves X (relevance)
  - $\mathbb{E}[Ze] = 0$ (exogeneity)
- **Example 1**: Angrist and Krueger (1991)

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

- Instrument is quarter of birth

Overview

1. Classic asymptotic results
2. Weak IV phenomenon
3. Identification Robust Inference
4. Estimation
5. Summary and Open Questions

Classic IV results

$$Y_i = \beta' X_i + \gamma' W_i + e_i, \text{ with } \mathbb{E}[W_i e_i] = 0$$

- Instruments Z_i (K-dimensional) satisfy:
  - exogeneity: $\mathbb{E}[Z_i e_i] = 0$
  - relevance: $X_i = \pi' Z_i + \delta' W_i + v_i$, $rank(\pi) = dim(X_i)$
- GMM moment condition

$$\mathbb{E}\left[(Y_i - \beta' X_i - \gamma' W_i)(Z_i', W_i')'\right] = 0$$

Classic IV results

$$Y_i = \beta' X_i + \gamma' W_i + e_i, \text{ with } \mathbb{E}[W_i e_i] = 0$$

- Instruments Z_i (K-dimensional)
- In case of small dimensional W_i asymptotic analysis is equivalent to considering partialled out model

$$\tilde{Y}_i = \beta' \tilde{X}_i + \tilde{e}_i$$

  with instrument $\tilde{Z}_i$, where variables with tilde $\tilde{Y} = (I - W(W'W)^{-1}W')Y$ are residuals from regression on W.

Classic IV results

- From now we concentrate on model without controls (controls are partialled out):
$$Y_i = \beta' X_i + e_i,$$
- Instruments Z_i (K-dimensional) satisfy:
  - exogeneity: $\mathbb{E}[Z_i e_i] = 0$
  - relevance: $X_i = \pi' Z_i + v_i$, $rank(\pi) = dim(X_i)$
- GMM estimator:
$$\widehat{\beta}(S) = \arg\min_{\beta} (Y - X\beta)' Z S Z' (Y - X\beta)$$

  where S is some positive-definite weight matrix

Classic IV results

- In just identified case weight matrix is unimportant

$$\widehat{\beta} = (Z'X)^{-1}Z'Y$$

- The logic of consistency and asymptotic gaussianity:

$$\sqrt{n}(\widehat{\beta} - \beta_0) = \left(\frac{1}{n}Z'X\right)^{-1}\frac{1}{\sqrt{n}}Z'e$$

  - Law of Large Numbers $\frac{1}{n}Z'X \to^p \mathbb{E}[Z_iX_i']$
  - CLT: $\frac{1}{\sqrt{n}}Z'e \Rightarrow N(0, \mathbb{E}[e_i^2 Z_i Z_i'])$

- If $\mathbb{E}[Z_iX_i']$ is a full rank matrix (relevance), then $\widehat{\beta}$ is consistent and asymptotically gaussian

$$\sqrt{n}(\widehat{\beta} - \beta_0) \Rightarrow N\left(0, \left(\mathbb{E}[X_iZ_i']\right)^{-1}\mathbb{E}[e_i^2 Z_iZ_i']\left(\mathbb{E}[Z_iX_i']\right)^{-1}\right)$$

Classic IV results

- GMM estimator in an over-identified case:

$$\widehat{\beta}(S) = \arg\min_{\beta}(Y - X\beta)'ZSZ'(Y - X\beta)$$

where S is some positive-definite weight matrix

- Under some conditions (relevance is one of them):

$$\widehat{\beta}(S) = \left(X'ZSZ'X\right)^{-1} X'ZSZ'Y$$

is consistent and asymptotically gaussian

- Optimal choice of weight matrix $S^* = (\mathbb{E}[e_i^2 Z_i Z_i'])^{-1}$

Classic IV results

- GMM estimator (let X be d-dimensional, $d < K$):

$$\widehat{\beta}(S) = \arg\min_{\beta}(Y - X\beta)' Z S Z' (Y - X\beta)$$

- First order condition:

$$\frac{X'Z}{n} S Z'(Y - X\beta) = 0$$

combines instruments to a just identified set $Z_s = Z \cdot S \cdot \frac{(Z'X)}{n}$

- Optimal instrument

$$Z^* = Z \cdot (\mathbb{E}[e_i^2 Z_i Z_i'])^{-1} \cdot \mathbb{E}[Z_i X_i']$$

weights signal and noise optimally

Classic IV results

- GMM estimator: $\widehat{\beta}(S) = \arg\min_\beta (Y - X\beta)' Z S Z' (Y - X\beta)$
- Infeasible optimal choice of weight matrix $S^* = (\mathbb{E}[e_i^2 Z_i Z_i'])^{-1}$
- We need $\hat{\Sigma}(\beta)$ a consistent estimator of asymptotic variance of $Z'(Y - X\beta)$
- Feasible realizations of efficient IV:
  - 2 step efficient GMM
    $\widehat{\beta}_{2GMM} = \arg\min_\beta (Y - X\beta)' Z \left[\hat{\Sigma}(\hat{\beta}_0)\right]^{-1} Z'(Y - X\beta)$
  - Continuously updated GMM
    $\widehat{\beta}_{CUE} = \arg\min_\beta (Y - X\beta)' Z \left[\hat{\Sigma}(\beta)\right]^{-1} Z'(Y - X\beta)$
- Under homoskedasticity $\Sigma(\beta_0) \propto (Z'Z)$
  - 2 Step GMM= TSLS
  - CUE=LIML

Summary

Classical asymptotics is based on two ideas:

1. A properly selected just-identified IV is consistent and asymptotically gaussian because:
   - Numerator is gaussian (exogeneity important here)
   - Denominator is consistent and full rank (relevance!)
2. Signal is strong enough that one can select consistently an optimal instrument

Overview

1. Classic asymptotic results
2. Weak IV phenomenon
3. Identification Robust Inference
4. Estimation
5. Summary and Open Questions

Examples of Weak IV

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

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

- Instrument is quarter of birth
- Empirical sample
  - very large sample ($>$ 300K observations)
  - very low first stage R^2 (between .0001 and .0002)
  - quite tight standard errors for IV estimate
- Bound, Jaeger and Baker (1995) randomly simulated quarter of birth and re-run estimation - nothing in their IV estimates suggested that they are invalid (!)

Examples of Weak IV

- Euler equation

$$\mathbb{E}[(\Delta c_{t+1} - \tau - \psi r_{t+1}) Z_t] = 0$$

- Can estimate it two different ways:

$$\Delta c_{t+1} = \tau + \psi r_{t+1} + e_t \text{ using instruments } Z_t$$

$$r_{t+1} = \mu + \gamma \Delta c_{t+1} + v_t \text{ using instruments } Z_t$$

- We have $\psi = 1/\gamma$
- In data the standard confidence sets for ψ obtained in two ways do not intersect
- Problem: Δc_{t+1} is hard to forecast (leads to weak instruments)

Weak IV with 1 regressor and 1 instrument

$$Y_i = \beta' X_i + e_i, \text{ with instrument } Z_i$$

- Estimator does not depend on weight matrix (in just id case)

$$\widehat{\beta} = \frac{\sum_i Z_i Y_i}{\sum_i Z_i X_i}; \quad \sqrt{n}(\widehat{\beta} - \beta_0) = \frac{\frac{1}{\sqrt{n}} \sum_i Z_i e_i}{\frac{1}{n} \sum_i Z_i X_i}$$

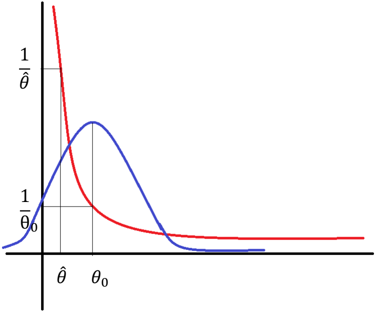
- Exogeneity means $\mathbb{E}[Z_i e_i] = 0$, CLT holds under quite wide set of assumptions: $\frac{1}{\sqrt{n}} \sum_i Z_i e_i \Rightarrow N(0, \Sigma)$
- $\frac{1}{n} \sum_i Z_i X_i \rightarrow^p \mathbb{E}[Z_i X_i]$, by relevance $\mathbb{E}[Z_i X_i] \neq 0$

Weak IV with 1 regressor and 1 instrument

$$\sqrt{n}(\widehat{\beta} - \beta_0) = \frac{\frac{1}{\sqrt{n}} \sum_i Z_i e_i}{\frac{1}{n} \sum_i Z_i X_i}$$

- CLT for denominator $\frac{1}{\sqrt{n}} \sum_i (Z_i X_i - \mathbb{E}[Z_i X_i]) \Rightarrow N(0, \Omega)$
- Denominator $\frac{1}{n} \sum_i Z_i X_i \approx \mathbb{E}[Z_i X_i] + \frac{1}{\sqrt{n}} N(0, \Omega)$
- If noise $\frac{1}{\sqrt{n}} N(0, \Omega)$ is comparable in size to the signal $\mathbb{E}[Z_i X_i]$, the numerator may be close to zero

Weak identification: explanation



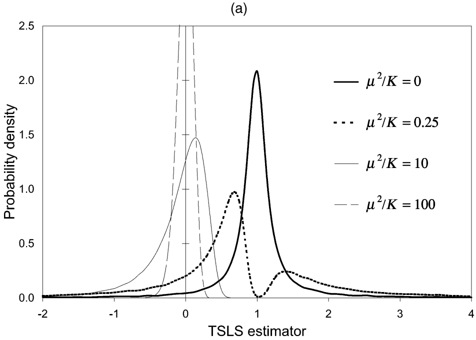
- Here $\theta_0 = \mathbb{E}[Z_i X_i]$ and $\widehat{\theta} = \frac{1}{n} \sum_i Z_i X_i$
- Concentration parameter $\mu^2 = \frac{(\mathbb{E}[Z_i X_i])^2}{Var(Z_i X_i)/n}$, signal-to-noise ratio measures the extent of problem

Weak identification: explanation

$$\mu(\widehat{\beta} - \beta_0) = \sqrt{\frac{\Sigma}{\Omega}} \cdot \frac{\xi_e}{1 + \frac{\xi_x}{\mu}}$$

- $\xi_e = \frac{1}{\sqrt{n\Sigma}} \sum_i Z_i e_i \Rightarrow N(0,1)$
- $\xi_x = \frac{1}{\sqrt{n\Omega}} \sum_i (Z_i X_i - \mathbb{E}[Z_i X_i]) \Rightarrow N(0,1)$
- ξ_e and ξ_x are correlated (endogeneity of X_i), this creates a bias
- If the instrument is irrelevant ($\mu = 0$), then $\widehat{\beta} - \beta_0 = \sqrt{\frac{\Sigma}{\Omega}} \cdot \frac{\xi_e}{\xi_x}$ is asymptotically centered at $\widehat{\beta}_{OLS}$
- Concentration parameter μ^2 serves as effective sample size

Weak identification: explanation



- Simulations from Stock et al (2002). Homoskedastic just identified ($K = 1$) case, $\frac{\Sigma}{\Omega} = 1$, correlation of errors is 0.99

Weak identification: pre-test

- Concentration parameter $\mu^2 = \frac{(\mathbb{E}[Z_i X_i])^2}{Var(Z_i X_i)/n}$, signal-to-noise ratio measures the extent of problem
- If μ^2 is small, $\widehat{\beta}$ is biased towards OLS, t-statistics tests/confidence sets are not reliable
- First stage regression $X_i = \pi Z_i + v_i$
- First stage F-statistics for testing $H_0 : \pi = 0$ is a good proxy for μ^2 in just-identified case with 1 endogenous regressor
- Empirical rule of thumb: $F > 10$ gives reliable inference
- This rule is not satisfactory outside of 1-regressor, 1-instrument settings

Overview

1. Classic asymptotic results
2. Weak IV phenomenon
3. Identification Robust Inference
   - Full parameter inference
   - Sub-vector inference
4. Estimation
5. Summary and Open Questions

AR test

$$Y_i = \beta' X_i + e_i$$

- Assumption is exogeneity: $\mathbb{E}[Z_i e_i] = 0$
- Want to test $H_0 : \beta = \beta_0$ in a reliable way
- Idea: if β_0 is true parameter value then $e_i = Y_i - \beta_0' X_i$ is uncorrelated with Z_i
- Under minor assumptions: $\frac{1}{\sqrt{n}} \sum_i Z_i (Y_i - \beta_0' X_i) \Rightarrow N(0, \Sigma)$
- We can construct a consistent estimator of Σ

$$AR(\beta_0) = \frac{1}{n}(Y - \beta_0 X)' Z \widehat{\Sigma}^{-1} Z'(Y - \beta_0 X) \Rightarrow \chi^2_K$$

AR test

- Intuition of AR test in a just identified case
- $\theta_1 = \mathbb{E}[Z_i Y_i]$ and $\widehat{\theta}_1 = \frac{1}{n}\sum_i Z_i Y_i$ and $\theta_2 = \mathbb{E}[Z_i X_i]$ and $\widehat{\theta}_2 = \frac{1}{n}\sum_i Z_i X_i$
- $\sqrt{n}(\widehat{\theta}_1 - \theta_1) \Rightarrow N(0, \Sigma)$ and $\sqrt{n}(\widehat{\theta}_2 - \theta_2) \Rightarrow N(0, \Omega)$
- Parameter of interest $\beta = \frac{\theta_1}{\theta_2}$ and $\widehat{\beta} = \frac{\widehat{\theta}_1}{\widehat{\theta}_2}$
- By delta-method $\sqrt{n}(\widehat{\beta} - \beta) \Rightarrow N(0, \Sigma_\beta)$
- Classic asymptotics test for $H_0 : \beta = \beta_0$ using t-statistics : $\sqrt{\frac{n}{\Sigma_\beta}}(\widehat{\beta} - \beta_0) \Rightarrow N(0, 1)$
- AR reformulates a hypothesis as $H_0 : \theta_1 = \beta_0\theta_2$ and tests this hypothesis:

$$AR(\beta_0) = \frac{\left(\widehat{\theta}_1 - \beta_0\widehat{\theta}_2\right)^2}{Var(\widehat{\theta}_1 - \beta_0\widehat{\theta}_2)}$$

AR test

- One can test $H_0 : \beta = \beta_0$ identification robust way accepting $AR(\beta_0) < \chi^2_{k,1-\alpha}$
- Optimal set in a class of robust test with some invariance property for a just-identified case
- Applicable in several regressors and or several instruments cases
- Confidence set can be produced as a set of β_0 accepted by AR test
- Confidence set can be infinite (good feature)
- Confidence set can be empty (has power against mis-specification)

Robust inference in over-identified case

$$Y_i = \beta' X_i + e_i$$

- Over-identified case: instrument Z_i is K-dimensional and $K > d$
- AR test is still robust towards weak identification, but has low power when identification is strong
- When instruments are strong, we can combine them in the efficient way, and do AR test using only efficient instrument
- This is the idea of LM test. Infeasible version (with weight π):

$$LM^*(\beta_0) = \frac{1}{n}(Y - \beta_0 X)'(Z\pi)\left(\pi'\widehat{\Sigma}\pi\right)^{-1}(Z\pi)'(Y - \beta_0 X) \Rightarrow \chi^2_d$$

LM test

- The optimal combination is $\pi = \left(\mathbb{E}[e_i^2 Z_i Z_i']\right)^{-1} \mathbb{E}[Z_i X_i]$
- When signal-to-noise ratio is low, it is hard to estimate the signal $\mathbb{E}[Z_i X_i]$ well, getting variance is a challenge as well
- We give up on variance - find an optimal combination for homoskedatsic case only
- Idea: use OLS for combining $\widehat{\pi} = (Z'Z)^{-1} Z'X$
- Problem: $\widehat{\pi}$ is very volatile under weak IV and is correlated with $Z'e$.
- Naive LM test does not have correct size under weak IV:

$$LM_{naive}(\beta_0) = \frac{1}{n}(Y - \beta_0 X)'(Z\widehat{\pi})\left(\widehat{\pi}'\widehat{\Sigma}\widehat{\pi}\right)^{-1}(Z\widehat{\pi})'(Y - \beta_0 X)$$

LM test

- Kleibergen (2002): create a new estimator $\tilde{\pi}$
  - consistent for π if identification is strong
  - asymptotically independent from $Z'e$ if identification is weak

$$KLM(\beta_0) = \frac{1}{n}(Y - \beta_0 X)'(Z\tilde{\pi})\left(\tilde{\pi}'\widehat{\Sigma}\tilde{\pi}\right)^{-1}(Z\tilde{\pi})'(Y - \beta_0 X)$$

- $\tilde{\pi} = \widehat{\pi} - \widehat{cov}(\widehat{\pi}, Z'e)\left(\widehat{Var}(Z'e)\right)^{-1} Z'e$
- KLM test for $H_0 : \beta = \beta_0$
  - has correct size irrespective of identification strength
  - asymptotically efficient if identification is strong

Conditional inference

- Under strong identification there are asymptotically uniformly most powerful tests for $H_0 : \beta = \beta_0$ (Wald, LM, LR)
- Under weak identification, there is no asymptotically uniformly most powerful (Andrews et al 2006)
- KLM may have very low power under weak identification (Moreira et al 2024)
- Can we use any other test statistics for $H_0 : \beta = \beta_0$, like Wald or LR?
- Problem: asymptotic distributions of Wald or LR depend on π
- Solution: Moreira (2003) - conditional inference

Conditional inference

- Problem: You want to test $H_0 : \beta = \beta_0$
  - distribution of test statistics $\mathcal{S}$ depends on (nuisance) parameter π
  - critical values should depend on π
  - you cannot estimate π with good enough precision
- Solution (Moreira, 2003): there is (asymptotically) sufficient statistics $\mathcal{T}$ for π
  - distribution of data (or any statistics) conditionally on $\mathcal{T}$ does not depend on π
  - create critical values depending on $\mathcal{T}$ (random critical values!!!) to control conditional size

$$\mathbb{P}\{\mathcal{S} > cv_\alpha(t) | \mathcal{T} = t\} = \alpha$$

- Original idea: use simulations conditioning on samples with given value of $\mathcal{T}$
- For some statistics like LR, no simulations are needed- there are analytic way of calculating conditional p-value, (called CLR test)

Summary

- If you want to either test $H_0 : \beta = \beta_0$ or construct a confidence set for β
    - you should NOT use a pre-test for weak identification
    - you should use identification robust test (or invert it for a confidence set)
- Arguments against pre-test
    - first stage F works only for 1 regressor-1 instrument case, or over-identified homoskedastic case
    - even in simplest case it may create selection bias (Angrist and Kolesar, 2022)
- Arguments for identification robust testing
    - Control size irrespective of the identification strength
    - Asymptotically efficient if identification is strong (KLM, CLR, conditional Wald)

Outline

1. Classic asymptotic results
2. Weak IV phenomenon
3. Identification Robust Inference
   - Full parameter inference
   - Sub-vector inference
4. Estimation
5. Summary and Open Questions

Projection method

$$Y_i = \beta_1 X_{1,i} + \beta_2 X_{2,i} + e_i$$

- Instruments Z_i are at least $d_1 + d_2$-dimensional
- β_1 is the parameter of interest, β_2 is nuisance
- How to test $H_0 : \beta_1 = \beta_{1,0}$? Confidence set for β_1 only?
- Projection method:
  - Hypothesis $H_0 : \beta_1 = \beta_{1,0}$ is accepted if $\tilde{H}_0 : \beta = (\beta_{1,0}, \beta_{2,0})$ is accepted for some $\beta_{2,0}$ (search among all potential $\beta_{2,0}$)
  - Create a joint confidence set for β, project it on β_1 space
  - Test (confidence set) for β should be done in identification robust way

Projection method and power loss

- Assume that test is done by using KLM statistics
- Projection method test for $H_0 : \beta_1 = \beta_{1,0}$ accepts if

$$\min_{\beta_{2,0}} KLM(\beta_{1,0}, \beta_{2,0}) = KLM(\beta_{1,0}, \widehat{\beta}_2) \leq \chi^2_d$$

  where $d = d_1 + d_2$
- If all parameters were strongly identified, and we knew that, we would adjust degrees of freedom and used $\chi^2_{d_1}$
- Indeed,

$$\min_{\beta_{2,0}} KLM(\beta_{1,0}, \beta_{2,0}) \leq KLM(\beta_{1,0}, \beta_{2,0}) \sim \chi^2_d$$

  so, projection test use conservative critical values (is not as powerful as it could have been)

Improvements over projection method

- Want to test $H_0 : \beta_1 = \beta_{1,0}$
  - If nuisance parameter is strongly identified then we can adjust critical values for the degrees of freedom
  - If nuisance parameter is weakly identified then distribution of most tests are harder to assess, and they depend on nuisance parameters
  - Projection method controls size (but typically not similar)
- Need to 'pre-test' whether nuisance parameter is strong or weak
- No perfect pre-test exists, but anything that improves power of two-step procedure over projection method is good.
- There are proposals of this type: Chaudhuri and Zivot (2011), Andrews (2018), Guggenberger et al (2012)

Overview

1. Classic asymptotic results
2. Weak IV phenomenon
3. Identification Robust Inference
4. Estimation
5. Summary and Open Questions

Estimation

- When identification is weak, there is no consistent estimator exists
- Under weak identification in just identified case, TSLS does not have any moments, is median biased (towards OLS)
- Under weak identification in over-identified case, TSLS is biased towards OLS
- LIML and TSLS behaves differently under weak identification
- Porter and Hirano (2015) if instruments can be arbitrary weak, no asymptotically mean-unbiased, no asymptotically median-unbiased estimator exists
- Andrews and Armstrong (2017) if the sign of first stage is known, one can create an asymptotically unbiased estimator

Overview

1. Classic asymptotic results
2. Weak IV phenomenon
3. Identification Robust Inference
4. Estimation
5. Summary and Open Questions

Summary

- Classical results in IV rely on relevance in two ways: (i) denominator stabilize at a full rank matrix; (ii) consistency of optimal combination of instrument
- Weak identification distort both
- Inferences: AR test solves problem (i) by re-formulating hypothesis
- Inferences: (ii) is 'solved' by KLM, CLR tests
- Estimation is hard, because of not well-defined criteria of quality

Open questions

- (Very hard) Estimation - what can be or reported as an estimator? Or should it be reported at all?
- (Hard) Seems that some pre-test for identification strength is needed to improve performance for a sub-vector tests