Canonical Research Designs IV: Instrumental Variables II

Paul Goldsmith-Pinkham

April 1, 2021

Roadmap for Today

- Key point from previous discussion is that the two stage least squares estimator (e.g. GMM with a particular weight matrix) gets at a particular causal estimand
- Today, we'll pivot to estimation issues
- This will show up in two ways:
  - Weak instruments
  - Many (weak) instruments

Weak instruments

$$Y_i = D_i\beta + W_i\gamma + \epsilon_i$$
$$D_i = Z_i\pi_1 + W_i\pi_2 + u_i$$

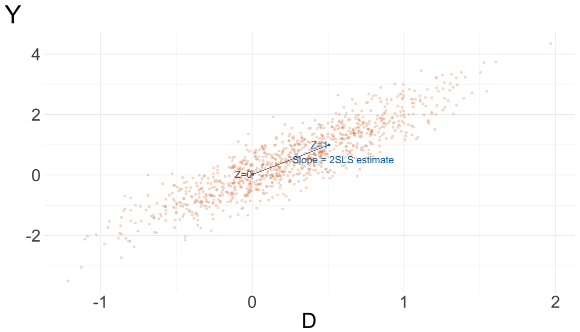
- Recall that one of the key assumptions for our estimation procedure was relevance
  - $\pi_1 \neq 0$, or $Cov(Z_i, D_i | W_i) \neq 0$
- Why is this necessary? Consider the 2SLS estimator for β_{IV} when W_i just includes a constant:

$$\hat{\beta} = \frac{Cov(Y_i, Z_i)}{Cov(D_i, Z_i)}$$

- If $Cov(D_i, Z_i) = 0$, this estimate is obviously undefined! But what about if it's very small?
  - Small variations in it will move around $\hat{\beta}$ in a big way. That's what statistical uncertainty will do
  - One easy way to see this: graphically

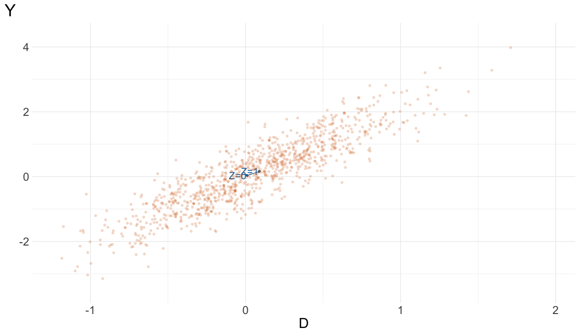
Weak instruments

- Simple 2SLS simulation, with binary instrument
    - First stage coef = 0.5, true beta = 2
- Note that the estimation on the x-axis comes from variation in the first stage
- The larger this is, the stronger the first stage
- However, if the first stage is weak, this interval is quite short, even if the variation in D stays the same



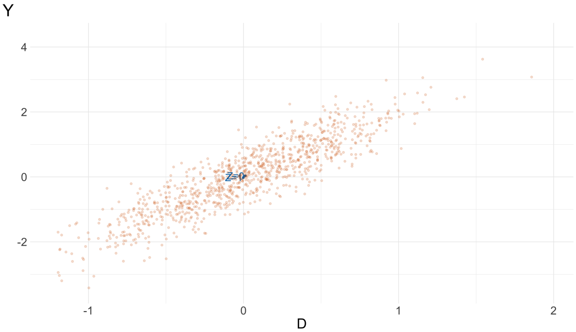
Weak instruments

- With a first stage coefficient of 0.1, it becomes hard to distinguish the points
  - Note: I hold fixed the overall variance of D here to keep the correct comparison!
- Given that the model is correctly specified, with enough data it should converge to the right β
- But small shifts in the x-axis will massively swing the estimate!



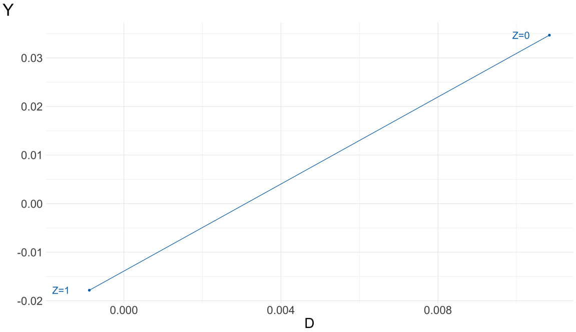
Weak instruments

- With a first stage coefficient of 0.01, the problem is even worse



Weak instruments

- With a first stage coefficient of 0.01, the problem is even worse
- We see that the relevant variation being exploited is tiny
- A small change in the x-axis points would even flip the sign!
- What does that do to our estimation procedure?



Weak instruments

- This graphical intuition should guide your understanding of the statistical problem

- For simplicity, assume the following: variables are demeaned (mean zero) and there are no additional controls (e.g. no constant). Hence,

$$\begin{aligned} Y_i &= D_i\beta + \epsilon_i \\ D_i &= Z_i\pi + u_i \\ \rightarrow Y_i &= Z_i \underbrace{\pi\beta}_{\delta} + u_i\beta + \epsilon_i \end{aligned}$$

- The 2SLS etimator (for single endog. variable) can then be written as:

$$\hat{\beta}_{2SLS} = \frac{D'Z(Z'Z)^{-1}Z'Y}{D'Z(Z'Z)^{-1}Z'D} = \frac{D'P_Z P_Z' Y}{D'P_Z P_Z D} = \frac{\hat{D}'\hat{Y}}{\hat{D}'\hat{D}} = \frac{\hat{\pi}'\hat{Q}\hat{\delta}}{\hat{\pi}'\hat{Q}\hat{\pi}} = \underbrace{\frac{\hat{\delta}}{\hat{\pi}},}_{\text{Single Instrument}} \qquad \hat{Q} = Z'Z$$

Weak instruments

- Intuitively, the 2SLS estimate is just the ratio of the reduced form and the first stage
    - This ratio can be highly non-linear with the denominator

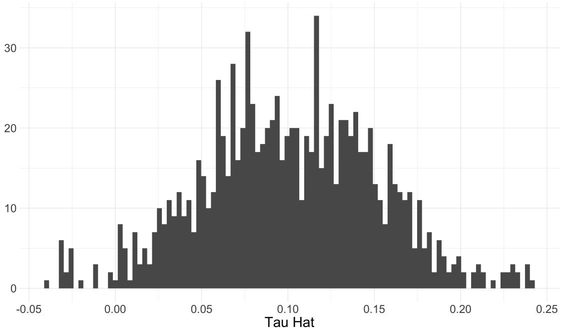
$$\hat{\beta}_{2SLS} = \frac{\hat{\delta}}{\hat{\pi}}$$

- Notice that under traditional asymptotic approximations, the small value for π is not a big deal.
    - Given a large enough sample, $\hat{\pi} \rightarrow \pi$, and you will consistently estimate β

- That's not really what we want to approximate though
    - In a finite sample, $\hat{\pi}$ is noisy, and if the s.e. of $\hat{\pi}$ is large relative to $\hat{\pi}$, that can cause very weird behavior in $\hat{\beta}$

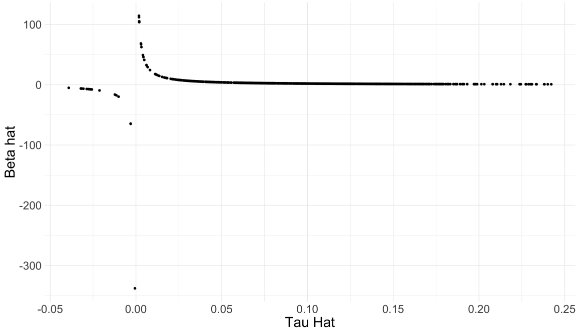
Weak instruments

- Imagine a marginally significant first stage (se = 0.05, estimate = 0.1)
- This estimator is normal, and reasonably well-behaved



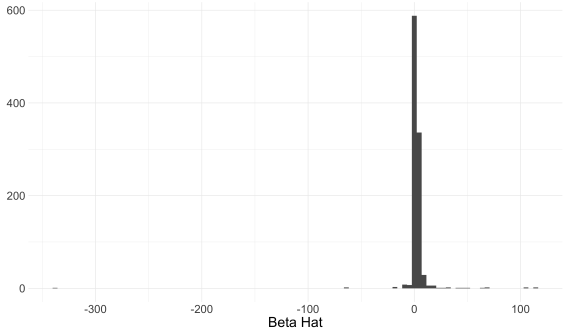
Weak instruments

- Imagine a marginally significant first stage (se = 0.05, estimate = 0.1)
- This estimator is normal, and reasonably well-behaved
- However, the relationship between $\hat{\tau}$ and $\hat{\beta}$ is highly nonlinear near zero



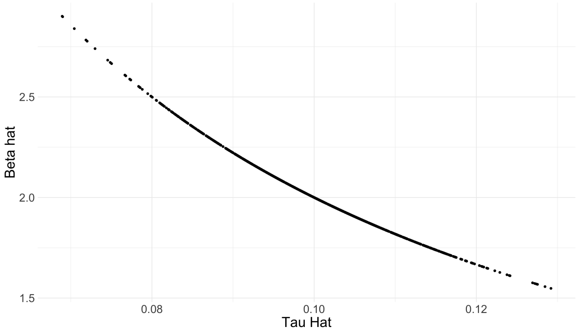
Weak instruments

- Imagine a marginally significant first stage (se = 0.05, estimate = 0.1)
- This estimator is normal, and reasonably well-behaved
- However, the relationship between $\hat{\tau}$ and $\hat{\beta}$ is highly nonlinear near zero
- This makes the distribution for $\hat{\beta}$ very non-normal
    - Asymptotic normality is a bad approximation!



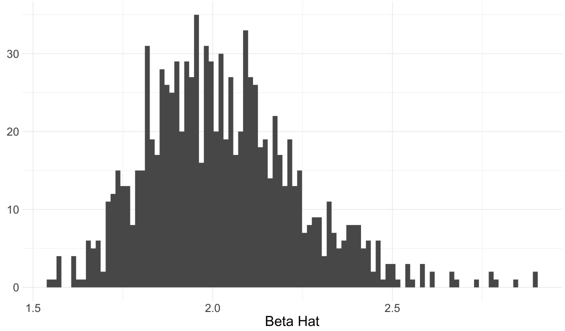
Weak instruments

- Interestingly, this is not the case if π is sufficiently large!
- Then the relationship is quite linear
  - Effectively, the delta method is a very good approximation



Weak instruments

- Interestingly, this is not the case if π is sufficiently large!
- Then the relationship is quite linear
  - Effectively, the delta method is a very good approximation
- This makes the distribution for $\hat{\beta}$ reasonably good
  - We have a problem about differentiating between these regimes



Weak instruments

- We want to account for this lack of normality. There are two ways to do this:
  - Test for a sufficiently strong first stage so we can ignore the issue
  - Use an approach that is robust to both

- To see these approaches, useful to have the following approximation for the bias in our estimate (Bekker 1994 group asymptotics):

$$\begin{aligned} E(\hat{\beta}_{2SLS} - \beta) &\approx (E(D'P_Z D))^{-1} E(u'P_Z \epsilon) \\ &= (E(\pi' Z Z' \pi) + E(u' P_Z u))^{-1} E(u' P_Z \epsilon) \\ &= (E(\pi' Z Z' \pi) + \sigma_u^2 K)^{-1} E(\sigma_{u\epsilon} K), \end{aligned}$$

  where the last step is a trick exploiting homoskedasticity in u and ϵ.
  - The trick: $E(u'P_Z u) = E(tr(u'P_Z u)) = tr(P_Z E(u'u)) = tr(P_Z \sigma_u^2) = K\sigma_u^2$, which exploits that 1) trace of a scalar is equal to the scalar 2) trace of expectation is expectation of trace 3) trace of idempotent matrix is the rank

Weak instruments

$$E(\hat{\beta}_{2SLS} - \beta) \approx \underbrace{\frac{\sigma_{u\epsilon}}{\sigma_u^2}}_{\text{OVB}} \left[\underbrace{\frac{E(\pi' Z' Z \pi)/K}{\sigma_u^2}}_{\text{First Stage F statistic}} + 1 \right]^{-1}$$

- In the end, you get a clear relationship between the bias in $\hat{\beta}_{2SLS}$ and the first stage F-statistic and the bias from OLS
- The first stage F is just the share explained in the first stage, relative to the “noise” in the first stage. As F increases, the bias decreases!
- If there is zero power, $F = 0$ and IV is just the OLS estimate
- Key point – when there are many instruments, the bias increases
    - This is essentially coming from “overfitting” in the first stage (recall where the K pops out)

Solution 1: Pretesting

- A natural solution to this is to just check if the F-statistic is large enough that these highlighted problems are not an issue.

- This is the approach initially developed by Staiger and Stock (1997) and Stock and Yogo (2005).
    - Typical rule of thumb: first-stage F-statistic above 10 means that bias won't be larger than 10% with size of 5%. Very popular!

- Key assumption: homoskedastic. This is a strong assumption!

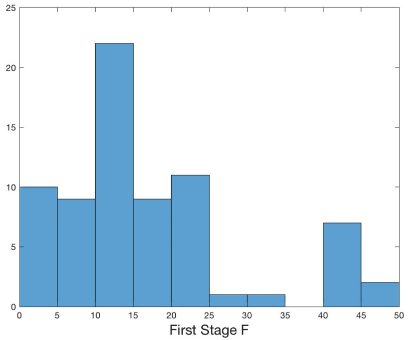


Figure 1: Distribution of reported first-stage F-statistics (and their non-homoskedastic generalizations) in 72 specifications with a single endogenous regressor and first-stage F smaller than 50. Total number of single endogenous regressor specifications reporting F-statistics is 108.

Solution 1: Pretesting

- We can do better, however. Montiel Olea and Pfluger(2013) have a heteroskedasticty-robust test, which proposes a more appropriate F statistic (allows for clustering, autocorrelation, etc.)
  - Cutoff is more like 23.1
  - An arms race in F-statistics!

- Stata package `weakivtest` here: `https://www.stata-journal.com/article.html?article=st0377`

A robust test for weak instruments in Stata

Carolin E. Pflueger
University of British Columbia
Vancouver, Canada
carolin.pflueger@sauder.ubc.ca

Su Wang
London School of Economics
London, UK
s.wang50@lse.ac.uk

Abstract. We introduce a routine, **weakivtest**, that implements the test for weak instruments by Montiel Olea and Pflueger (2013, *Journal of Business and Economic Statistics* 31: 358–369). **weakivtest** allows for errors that are not conditionally homoskedastic and serially uncorrelated. It extends the Stock and Yogo (2005, Testing for weak instruments in linear IV regression. In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, ed. D. W. K. Andrews and J. J. Stock, 80–108. [Cambridge University Press]) weak-instrument tests available in **ivreg2** and in the **ivregress** postestimation command **estat firststage**. **weakivtest** tests the null hypothesis that instruments are weak or that the estimator's Nagar (1959, *Econometrica* 27: 575–595) bias is large relative to a benchmark for both two-stage least-squares estimation and limited-information maximum likelihood with one endogenous regressor. The routine can accommodate Eicker–Huber–White heteroskedasticity robust estimates, Newey and West (1987, *Econometrica* 55: 703–708) heteroskedasticity- and autocorrelation consistent estimates, and clustered variance estimates.

Solution 1: Pretesting

- The arms race continues. Lee et al. (2020) point out that current practice focuses on the β term, rather than on the t-statistic
  - which is how we claim statistical significance
- Need a much stronger first stage (F = 104!) for this
- Highlights the challenge of using pre-testing
  - Moreover, pre-testing for IV, much like pre-testing in dind trends, cause distort inference for your parameters

Valid t-ratio Inference for IV[1]

David S. Lee[2]
Princeton University and NBER

Marcelo J. Moreira[4]
FGV EPGE

Justin McCrary[3]
Columbia University and NBER

Jack Porter[5]
University of Wisconsin

October 15, 2020

Abstract

In the single IV model, current practice relies on the first-stage F exceeding some threshold (e.g., 10) as a criterion for trusting t-ratio inferences, even though this yields an anti-conservative test. We show that a true 5 percent test instead requires an F greater than 104.7. Maintaining 10 as a threshold requires replacing the critical value 1.96 with 3.43. We re-examine 57 AER papers and find that corrected inference causes half of the initially presumed statistically significant results to be insignificant. We introduce a more powerful test, the tF procedure, which provides F-dependent adjusted t-ratio critical values.

Keywords: Instrumental Variables, Weak Instruments, t-ratio, First-stage F statistic

Solution 2: Robust confidence intervals

- With a just-identified single endogeneous regressor, Anderson-Rubin confidence intervals are valid, irrespective of the weakness of the first stage

- This is the easiest way to deal with this inference problem! These results are robust regardless of your first stage
    - Chernozhukov and Hansen (2008) discuss a very easy and simple way to implement these confidence intervals
    - Stata and R packages are also available

ELSEVIER
Available online at www.sciencedirect.com
ScienceDirect
Economics Letters 100 (2008) 68 – 71
economics letters
www.elsevier.com/locate/econbase

The reduced form: A simple approach to inference with weak instruments

Victor Chernozhukov [a], Christian Hansen [b,*]

[a] MIT, United States
[b] University of Chicago, Graduate School of Business, 5807 S. Woodlawn Ave., Chicago, IL 60637, United States

Received 30 July 2007; received in revised form 19 October 2007; accepted 14 November 2007
Available online 28 November 2007

Abstract

In this paper, we show that conventional heteroskedasticity and autocorrelation robust inference procedures based on the reduced form provide tests and confidence intervals for structural parameters that are valid when instruments are strongly or weakly correlated to the endogenous variables.

Keywords: Heteroskedasticity; Autocorrelation; Weak identification

JEL classification: C12C30

Solution 2: Robust confidence intervals

- With a just-identified single endogeneous regressor, Anderson-Rubin confidence intervals are valid, irrespective of the weakness of the first stage

- This is the easiest way to deal with this inference problem! These results are robust regardless of your first stage
  - Chernozhukov and Hansen (2008) discuss a very easy and simple way to implement these confidence intervals
  - Stata and R packages are also available

Repeating the testing procedure mentioned above for multiple values of β_0 allows the construction of confidence intervals which are robust to weak instruments and heteroskedasticity or autocorrelation through a series of conventional least squares regressions. The procedure for constructing a confidence interval is as follows:

1. Select a set, $\mathcal{B}$, of potential values for β.
2. For each $b \in \mathcal{B}$, construct $\tilde{Y} = Y - Xb$ and regress $\tilde{Y}$ on Z to obtain $\hat{\alpha}$. Use $\hat{\alpha}$ and the corresponding estimated covariance matrix of $\hat{\alpha}$, $\widehat{Var}(\hat{\alpha})$, to construct the Wald statistic for testing $\alpha = 0$, $W_S(b) = \hat{\alpha}' \left[\widehat{Var}(\hat{\alpha})\right]^{-1} \hat{\alpha}$. The use of a robust covariance matrix estimator in forming $\widehat{Var}(\hat{\alpha})$ will result in tests and confidence intervals robust to both weak instruments and heteroskedasticity and/or autocorrelation.
3. Construct the $1-p$ level confidence region as the set of b such that $W_S(b) \leq c(1-p)$ where $c(1-p)$ is the $(1-p)^{th}$ percentile of a χ^2_r distribution.

Many instruments

- Recall from our discussion above that even many instruments creates bias:

$$E(\hat{\beta}_{2SLS} - \beta) \approx \underbrace{\frac{\sigma_{u\epsilon}}{\sigma_u^2}}_{\text{OVB}} \left[\underbrace{\frac{E(\pi' Z' Z \pi)/K}{\sigma_u^2}}_{\text{First Stage F statistic}} + 1 \right]^{-1}$$

- This is due to "overfitting" in the projection of 2SLS

- This is very solveable. Use of jackknife IV (which leaves out the own observation) will address this issue
  - See Angrist, Imbens and Krueger (1999) for details
  - Inference methods in this setting are a little less well-developed, however

- We will revisit this issue when considering Judge IV settings