Part C: Instrumental Variables

C1: IV Idea and Mechanics; Weak IV

Kirill Borusyak

ARE 213 Applied Econometrics

UC Berkeley, Fall 2025

Part C outline

1. Assumptions, uses, and mechanics of IV; weak IV
2. IV with treatment effect heterogeneity
3. Examiner designs ("judge IVs")

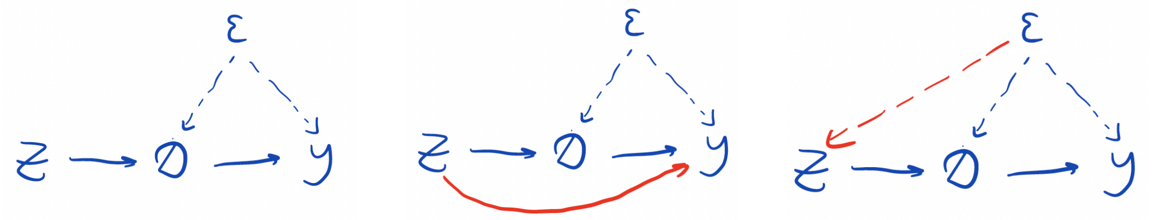
C1 outline

1. Setting and Examples
2. IV mechanics
3. IV for classical measurement error
4. Weak and Many (Weak) IVs

Readings:

- *IV Mechanics:* MHE (Ch. 4.1, 4.2.1, 4.6.4), Wooldridge (Ch. 5)
- *Weak IV:* Andrews et al. (2019, Annual Review of Economics), Imbens and Wooldridge (Lecture 15)

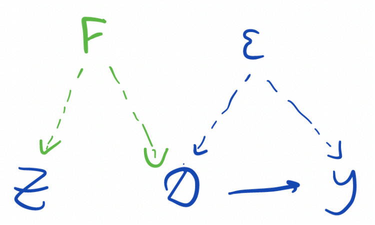
Instrumental variable DAGs



- **Exclusion**: $Y(d,z) = Y(d,z') \equiv Y(d)$ for all d, z, z'
- **Independence ("as-good-as-random assignment")**: $Z \perp\!\!\!\perp Y(d)$ for all d
- Exclusion + independence = instrument **exogeneity** (a.k.a. **orthogonality**)
  - Z is only correlated with Y if there is a causal effect of D on Y
  - Calling Z an "excluded instrument" is somewhat misleading

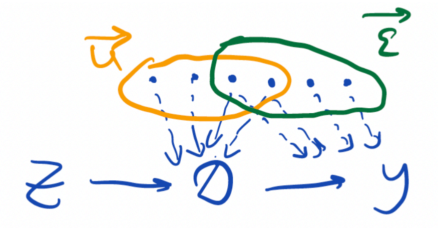
Instrumental variable DAGs (2)

- **Relevance**: $D(z) \neq D(z')$ with non-zero probability for some z, z'
- Exogeneity $+$ relevance $=$ instrument **validity**
- *Note*: non-causal first stage is also fine



  - E.g. $D_i =$ import penetration from China in industry i in the US
  - $Z_i =$ import penetration from China in Europe; $F_i =$?

IV with constant effects



- **Structural equation**: $Y_i = \mathcal{Y}(D_i, \vec{\varepsilon_i}) \implies Y_i = \tau D_i + \varepsilon_i$ (suppressing intercepts)
- **First stage**: $D_i = \mathcal{D}(Z_i, \vec{u_i}) \implies D_i = \pi Z_i + u_i$
- Relevance: $\pi \neq 0$ (assuming $\text{Cov}[Z_i, u_i] = 0$ by construction)
- Instrument exogeneity: $\text{Cov}[Z_i, \varepsilon_i] = 0$
- Treatment endogeneity: $\text{Cov}[\varepsilon_i, u_i] \neq 0$
- **Reduced-form** equation: $Y_i = \tau\pi Z_i + (\varepsilon_i + \tau u_i) \equiv \rho Z_i + e_i$

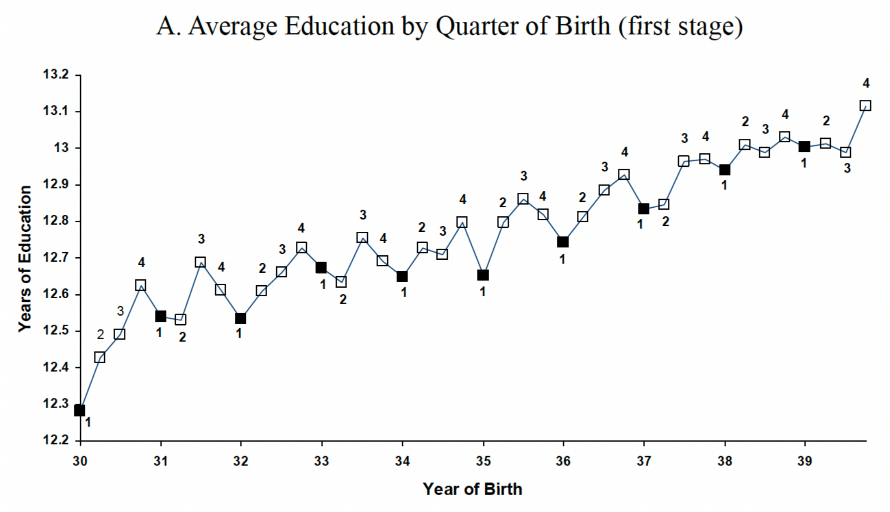
IV identification with constant effects

- Take 1:
  - $\mathrm{Cov}[Z_i, u_i] = 0 \Rightarrow$ First-stage OLS is consistent for π
  - $\mathrm{Cov}[Z_i, e_i] = 0 \Rightarrow$ Reduced-form OLS is consistent for $\rho = \tau\pi$
  - We have $\tau = \rho/\pi =$ Reduced-form/First-stage $= \mathrm{Cov}[Z_i, Y_i] / \mathrm{Cov}[Z_i, D_i]$
- Take 2: $\mathrm{Cov}[Z_i, \varepsilon_i] = 0$ where $\varepsilon_i = Y_i - \tau D_i$
  - Solving $\mathrm{Cov}[Z_i, Y_i] - \tau \mathrm{Cov}[Z_i, D_i] = 0$ yields the same answer

Example 1: Angrist and Krueger (1991)

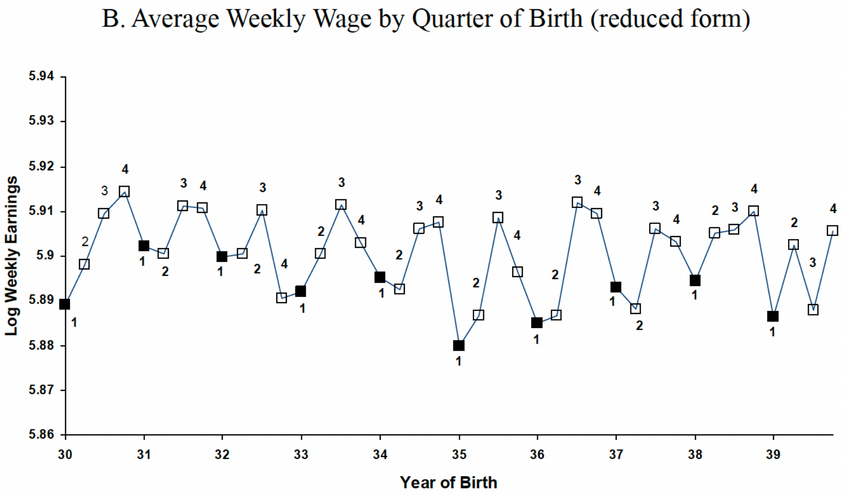
- Estimate returns to schooling: $D =$ years of schooling, $Y =$ log earnings
- OLS should be biased upwards: unobserved ability
- $Z =$ quarter of birth dummies
- Relevance: structure of compulsory schooling laws in the US
  - Children born in Q4 start school a year earlier than those born in Q1 next year
  - But can drop out at the same time upon reaching certain age, e.g. 16
  - Q1-born get less schooling

First stage



A. Average Education by Quarter of Birth (first stage)
Years of Education
Year of Birth
13.2
13.1
13
12.9
12.8
12.7
12.6
12.5
12.4
12.3
12.2
30
31
32
33
34
35
36
37
38
39

Reduced form



Exogeneity of QoB

- Exclusion: no other mechanism through which QoB affects earnings
    - Bound, Jaeger, and Baker (1995): QoB can affect school attendance, mental health issues
- Independence: QoB is not affected by factors correlated with potential earnings
    - Buckles and Hungerman (2013): Q1 births disproportionally happen to teenagers and unmarried mothers

Example 2: Encouragement designs

- $D =$ taking an experimental pill, $Y =$ health outcome
- $Z =$ dummy for being *invited* to receive the pill (with an option to decline)
- OLS bias?
- Relevance?
- Independence?
- Exclusion?

RCT with two-sided noncompliance

- Angrist (1990): The impact of serving in Vietnam war (D_i) on future earnings (Y_i)
  - Draft was based on televised lottery
  - Men with low lottery numbers were enlisted first
  - Still, some of them can avoid serving, while others can volunteer
  - Z_i = bins of lottery numbers

Example 3: Ingenious IVs from natural experiments

- Angrist and Evans (1998): How does the number of kids affect mother's labor supply?
  - Among mothers with 2+ kids, D_i = having 3+ kids
  - Z_{1i}: twins at second birth
  - Z_{2i}: first two kids are both boys or both girls

Example 4: Historical instruments

- Duranton and Turner (2012) study the impact of a region's connection to interstate highways in 1983 on urban growth 1983–2003

- They worry about strategic placement of highways:

> Our primary identification problem is the simultaneous determination of urban growth and transportation infrastructure. While one hopes that cities with high predicted employment and population growth receive new transportation infrastructure, we fear that such infrastructure is allocated to places with poor prospects. The resolution of this problem requires finding suitable instruments for transportation infrastructure. Our analysis suggests that such instruments should reflect either a city's level of transportation infrastructure at some time long ago or …

Example 4: Historical instruments (cont.)

- They use density of railroads in 1898 and highways in the plan of 1947 as IVs:

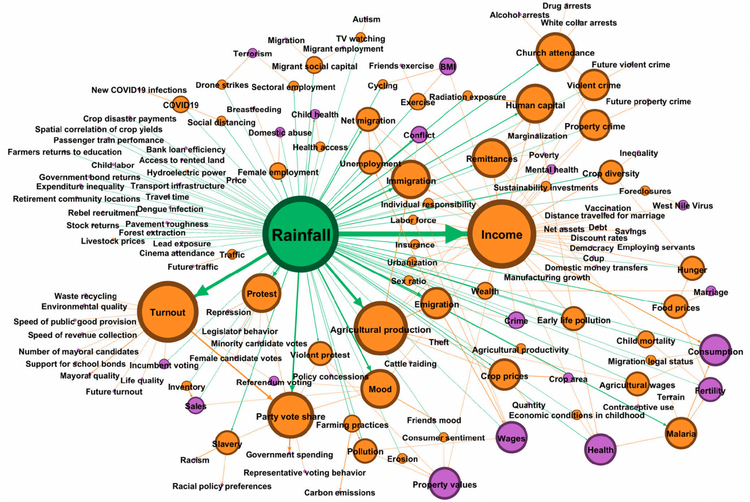
> The *a priori* case for thinking that 1898 railroad kilometres satisfy exogeneity condition (17) rests on the length of time since these railroads were built and the fundamental changes in the nature of the economy in the intervening years. The rail network was built, for the most part, during and immediately after the civil war, and during the industrial revolution.
>
> Furthermore, the rail network was constructed by private companies who were looking to make a profit from railroad operations in the not too distant future (Fogel, 1964; Fishlow, 1965). It is difficult to imagine how a rail network built for profit during the civil war and the industrial revolution could affect economic growth in cities 100 years later save through its effect on roads.

- Exclusion?
- Independence?

Example 5: Overusing an instrument

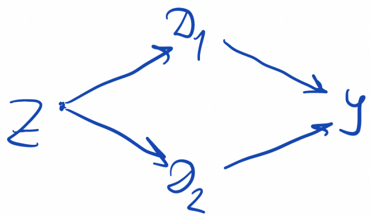
- Gallen and Raymond (2023) identify several instruments used for many endogenous variables
  - Local topography: presence of bodies of water and changes in elevation
  - Sibling characteristics in a family, e.g. having a twin
  - Ethnolinguistic fractionalization
  - Religion
  - Weather, e.g. rainfall

Usage of rainfall IV

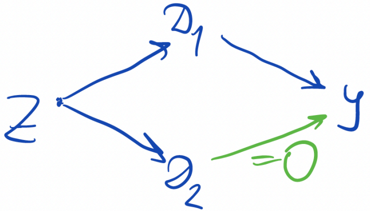


From Mellon (2024, AJPS); orange $= D$, purple $= Y$; circle size $= \#$ of studies

Often (but not always) a problem



Problem



OK for $D_1 \to Y$

OK (see Mellon (2024))

Failure of exogeneity

- Studies of the impacts of income on conflict often instrument for income with rainfall
- Sarsons (2015) shows that exogeneity is violated
    - Focuses on a subset of villages in India with a significant reduced-form
    - But no first-stage because of... dams

Example 6: IV for structural equations

- IV was developed to deal with **simultaneity** in a system of structural equations
  - Wright (1928) foundational study of the demand and supply for flaxseed
  - IV in the supply equation: demand curve shifter (price of cottonseed, a substitute)
  - IV in the demand equation: supply curve shifter (flaxseed yield per acre)
- IV exogeneity is about the error term, not the endogenous variable
  - Demand and supply have the same RHS variable (price) but require different IVs

$$Q = \tau_d P + \varepsilon_d, \qquad Q = \tau_s P + \varepsilon_s$$

  - Demand and inverse demand have different RHS variables but the same IV

$$Q = \tau_d P + \varepsilon_d \qquad \Longleftrightarrow \qquad P = \frac{1}{\tau_d} Q - \frac{1}{\tau_d} \varepsilon_d$$

Recap

- Exclusion and independence are distinct conditions
- Independence is guaranteed with random assignment, but exclusion may still fail
- Theory or common sense may promise exclusion, but independence may still fail
- Economically exogenous (e.g. historical and not strategically chosen) $\not\Longrightarrow$ econometrically exogenous
- If two studies use the same Z for different D but the same Y, exclusion likely fails for both
- Exogeneity requires thinking about ε (structural error or potential outcomes)

Outline

1. Setting and Examples
2. IV mechanics
3. IV for classical measurement error
4. Weak and Many (Weak) IVs

IV identification and estimation

- With constant effects, exogeneity implies moment condition: $\mathrm{Cov}\left[Z_i, Y_i - \tau D_i\right] = 0$

- Thus, if $\mathrm{Cov}\left[Z_i, D_i\right] \neq 0$,

$$\tau = \frac{\mathrm{Cov}\left[Z_i, Y_i\right]}{\mathrm{Cov}\left[Z_i, D_i\right]} = \frac{\mathrm{Cov}\left[Z_i, Y_i\right] / \mathrm{Var}\left[Z_i\right]}{\mathrm{Cov}\left[Z_i, D_i\right] / \mathrm{Var}\left[Z_i\right]} \equiv \frac{\text{Reduced-form}}{\text{First-stage}}$$

- If Z_i is binary, IV $=$ **Wald estimand**:

$$\tau = \frac{\mathbb{E}\left[Y_i \mid Z_i = 1\right] - \mathbb{E}\left[Y_i \mid Z_i = 0\right]}{\mathbb{E}\left[D_i \mid Z_i = 1\right] - \mathbb{E}\left[D_i \mid Z_i = 0\right]}$$

- Estimator $=$ sample analog

Angrist and Krueger with single binary IV

Table 4.1.2: Wald estimates of the returns to schooling using quarter of birth instruments

	(1) Born in the 1st or 2nd quarter of year	(2) Born in the 3rd or 4th quarter of year	(3) Difference (std. error) (1)-(2)
ln (weekly wage)	5.8916	5.9051	-0.01349 (0.00337)
Years of education	12.6881	12.8394	-0.1514 (0.0162)
Wald estimate of return to education			0.0891 (0.0210)
OLS estimate of return to education			0.0703 (0.0005)

(Reproduced from MHE)

Two-stage least squares (2SLS)

- Take first-stage fitted values $\hat{D}_i = \pi Z_i$ and regress Y_i on $\hat{D}_i$ ("**second stage**"):

$$\tau_{2SLS} = \frac{\mathrm{Cov}\left[Y_i, \hat{D}_i\right]}{\mathrm{Var}\left[\hat{D}_i\right]} = \frac{\mathrm{Cov}\left[Y_i, \hat{D}_i\right]}{\mathrm{Cov}\left[D_i, \hat{D}_i\right]} = \frac{\mathrm{Cov}\left[Y_i, Z_i\right]}{\mathrm{Cov}\left[D_i, Z_i\right]} = \tau_{IV}$$

- Avoid doing 2SLS manually:
  - SE are incorrect because they don't account for estimation noise of π
  - Easy to make mistakes, e.g. with controls

IV with included controls

- Expand structural equation, first-stage, and reduced-form to:

$$
\begin{aligned}
Y_i &= \tau D_i + \kappa'_{SE} X_i + \varepsilon_i \\
D_i &= \pi Z_i + \kappa'_{FS} X_i + u_i \\
Y_i &= \rho Z_i + \kappa'_{RF} X_i + e_i
\end{aligned}
$$

- Assume exogeneity $\mathbb{E}[Z_i \varepsilon_i] = 0$
- Without loss, we have $\mathbb{E}[X_i \varepsilon_i] = \mathbb{E}[X_i u_i] = \mathbb{E}[Z_i u_i] = 0$
- Denote $\tilde{Z}$ residuals from regressing Z on X; similar for $\tilde{Y}$, $\tilde{D}$
- There are several equivalent ways to get the IV estimand/estimator...

IV with included controls (2)

1. Moment conditions: for $\boldsymbol{D}_i = (D_i, X_i)'$, $\boldsymbol{\tau} = (\tau, \kappa_{SE})'$, and $\boldsymbol{Z}_i = (Z_i, X_i)'$,

$$\mathbb{E}\left[\boldsymbol{Z}_i(Y_i - \boldsymbol{D}_i'\boldsymbol{\tau})\right] = 0 \quad \Longrightarrow \quad \boldsymbol{\tau} = \mathbb{E}\left[\boldsymbol{Z}_i\boldsymbol{D}_i'\right]^{-1}\mathbb{E}\left[\boldsymbol{Z}_i Y_i\right]$$

   if $\mathbb{E}\left[\boldsymbol{Z}_i\boldsymbol{D}_i'\right]$ is full-rank (relevance)

2. Run first-stage and reduced-form regressions, take $\hat{\tau} = \hat{\rho}/\hat{\pi}$

3. By FWL, using $\tilde{Z}$ as an IV

$$\tau = \frac{\mathrm{Cov}\left[\tilde{Z}_i, Y_i\right]}{\mathrm{Cov}\left[\tilde{Z}_i, D_i\right]} = \frac{\mathrm{Cov}\left[\tilde{Z}_i, Y_i\right]/\mathrm{Var}\left[\tilde{Z}_i\right]}{\mathrm{Cov}\left[\tilde{Z}_i, D_i\right]/\mathrm{Var}\left[\tilde{Z}_i\right]} \equiv \frac{\text{Reduced-form with controls}}{\text{First-stage with controls}}$$

IV with included controls (3)

4. Also by FWL, regressing $\tilde{Y}$ and $\tilde{D}$ and IV'ing with Z or $\tilde{Z}$:

$$\tau = \frac{\operatorname{Cov}\left[Z_i, \tilde{Y}_i\right]}{\operatorname{Cov}\left[Z_i, \tilde{D}_i\right]} = \frac{\operatorname{Cov}\left[\tilde{Z}_i, \tilde{Y}_i\right]}{\operatorname{Cov}\left[\tilde{Z}_i, \tilde{D}_i\right]}$$

5. 2SLS: take first-stage fitted values $\hat{D}_i = \hat{\pi} Z_i + \hat{\kappa}'_{FS} X_i$ and regress Y_i on $\hat{D}_i$ and X_i

Multiple endogenous variables & IVs

- Suppose we have $\dim(Z_i) = K$ instruments for $\dim(D_i) = J \leq K$ endogenous variables
  - *Note:* "included" controls are added to both lists
- We have K moment conditions for J unknowns:

$$\mathbb{E}\left[Z_i\left(Y_i - D_i'\tau\right)\right] = 0$$

- Any J linear combinations $\Pi' Z_i$ of IVs (for $K \times J$ matrix Π with full rank J) would produce a consistent estimator:

$$\mathbb{E}\left[\Pi' Z_i\left(Y_i - D_i'\tau\right)\right] = 0 \quad \Longrightarrow \quad \tau = \left(\Pi'\mathbb{E}\left[Z_i D_i'\right]\right)^{-1}\Pi'\mathbb{E}\left[Z_i Y_i\right]$$

- In the **just-identified** case $K = J$, Π cancels out
  - There is a single IV estimator

Overidentified case

- In the **overidentified** case $K > J$, we have many consistent estimators
- 2SLS uses the first-stage matrix as Π:

$$\Pi = \mathbb{E}\left[Z_i Z_i'\right]^{-1} \mathbb{E}\left[Z_i D_i'\right]$$

  i.e. using as instruments the first-stage fitted values $\Pi' Z_i$
  - *Note:* there are J first-stages, each with all instruments on the RHS
- Instrumenting D_i with $\Pi' Z_i$ = regressing Y on $\Pi' Z_i \implies$ 2SLS procedure
- We can also test that all combinations of IVs yield the same estimate (up to noise)
  - Hansen's J-statistic for overidentifying restrictions
  - If the test rejects, you don't know which IVs violate exogeneity
  - And the test can reject because of heterogeneous effects

2SLS as a control function estimator

- We have two strategies for dealing with endogenous D_i: adding controls and IV
  - Are they related to each other?
- IV isolates "clean" variation in D_i
  - 2SLS estimates first-stage $D_i = \hat{\pi}' Z_i + \hat{u}_i$ and regresses Y_i on $\hat{\pi}' Z_i$
- What if we control for "dirty" variation in D_i instead (**control function** estimator)

$$Y_i = \tau D_i + \mu \hat{u}_i + \text{error}$$

  - *Exercise*: the two estimators are numerically the same (even with covariates)
  - Converse is also true: controlling for X_i = instrumenting D_i with its FWL residual

Control function approach to IV

- Is the control function approach to IV useful?
    - Equivalence breaks with nonlinearities: e.g. first-stage is probit or has square terms
    - Helpful when combining IVs with parametric restrictions: e.g. Heckman selection model that imposes joint normality of ε_i and latent first-stage error

Outline

IV for classical measurement error

- Another use of IV is when your RHS variable has **classical measurement error**:

$$\begin{aligned} Y_i &= \beta X_i^* + \varepsilon_i, & X_i^* &\text{ is what you'd like to measure} \\ X_i &= X_i^* + \xi_i, & X_i &\text{ is what you observe} \end{aligned}$$

- Classical measurement error: $\operatorname{Cov}\left[\xi_i, X_i^*\right] = 0$
- Caveat: can only happen with continuous variables
  - **Misclassification** of discrete variables is never classical: e.g. when $X_i^*, X_i \in \{0, 1\}$,

$$\begin{aligned} X_i^* &= 0 & \implies & & \xi_i \geq 0 \\ X_i^* &= 1 & \implies & & \xi_i \leq 0 \end{aligned}$$

    so $\operatorname{Cov}\left[\xi_i, X_i^*\right] < 0$

Bias of OLS

- Suppose OLS would have no bias absent measurement error: $\text{Cov}\left[X_i^*, \varepsilon\right] = 0$
- OLS with mismeasured X_i suffers from **attenuation bias**:

$$\beta_{OLS} = \frac{\text{Cov}\left[Y, X\right]}{\text{Var}\left[X\right]} = \frac{\text{Cov}\left[\beta X^* + \varepsilon, X^* + \xi\right]}{\text{Var}\left[X^* + \xi\right]} = \beta \frac{\text{Var}\left[X^*\right]}{\text{Var}\left[X^*\right] + \text{Var}\left[\xi\right]}$$

  with $\frac{\text{Var}[X^*]}{\text{Var}[X^*]+\text{Var}[\xi]} \in (0, 1)$ determined by the signal-to-noise ratio

- Can also see this from

$$Y_i = \beta\left(X_i - \xi_i\right) + \varepsilon_i = \beta X_i + \left(\varepsilon_i - \beta \xi_i\right)$$

  where $\text{Cov}\left[X_i, \xi_i\right] = \text{Cov}\left[X_i^* + \xi_i, \xi_i\right] = \text{Var}\left[\xi_i\right] \neq 0$

IV estimation is immune

- Classical error in the RHS variable: if $\mathrm{Cov}\left[Z, \varepsilon\right] = 0$ and $\mathrm{Cov}\left[Z, \xi\right] = 0$,

$$\beta_{IV} = \frac{\mathrm{Cov}\left[Y, Z\right]}{\mathrm{Cov}\left[X, Z\right]} = \frac{\mathrm{Cov}\left[\beta X^* + \varepsilon, Z\right]}{\mathrm{Cov}\left[X^* + \xi, Z\right]} = \beta$$

- Can also have non-classical measurement error in the IV: if also $Z_i = Z_i^* + \zeta_i$

$$\beta_{IV} = \frac{\mathrm{Cov}\left[Y, Z\right]}{\mathrm{Cov}\left[X, Z\right]} = \frac{\mathrm{Cov}\left[\beta X^* + \varepsilon, Z^* + \zeta\right]}{\mathrm{Cov}\left[X^* + \xi, Z^* + \zeta\right]} = \beta$$

assuming $\mathrm{Cov}\left[Z^*, \varepsilon\right] = \mathrm{Cov}\left[Z^*, \xi\right] = \mathrm{Cov}\left[\varepsilon, \zeta\right] = \mathrm{Cov}\left[\xi, \zeta\right] = \mathrm{Cov}\left[X^*, \zeta\right] = 0$

Lessons

1. Do not need to worry about classical error if already using IVs
   - Measurement error is a possible reason why OLS estimates tend to be smaller than IV
2. When using OLS, can solve measurement error by using a second imperfect measurement of X_i as an IV, as long as the errors are independent
   - E.g. Chalfin and McCrary (2018) instrument police-per-capita from Uniform Crime Reports data with the same from Annual Survey of Government
   - And use the 2nd measure to quantify measurement error in the 1st
   - *Exercise:* We can use X_{1i} as IV for X_{2i} or the other way round. What is better? Is there something better yet?

Outline

Weak IV: Overview

What is the weak IV situation?

- Instruments do not provide enough variation in the endogenous variables
- $K = J = 1$: low (partial) correlation of D_i and Z_i (even if $\hat{\pi}$ is significant)
- $K > J = 1$: low (partial) correlation of D_i with all or most Z_{ki}
- $J > 1$:
  - First-stage is weak for at least one endogenous variable
  - Or all individual first stages are strong, yet the same combinations of instruments predict multiple endog.vars

    $\Rightarrow$ 2nd stage of 2SLS is close to perfect multicollinearity
  - General condition: the 1st stage matrix Π is close to rank-deficient

Weak IV: Overview (2)

Problems:

- Large sensitivity to violations of IV exogeneity
- Large variance of IV estimates
- Failure of the asymptotic approximation (and bootstrap, too):
  - Finite-sample bias of IV estimates in the direction of OLS
  - Non-Gaussian distributions of IV estimates and t-statistics
  - Distorted coverage of tests and confidence intervals (both asymptotic and bootstrap)

Weak IV: Overview (3)

- Solutions:
  - Pre-screening: checking that IVs are strong enough (usually via first-stage F-stats)
  - Using "weak-robust" tests and confidence intervals (e.g. Anderson-Rubin)
  - With many (even strong) IVs: replacing 2SLS with more well-behaved estimators (LIML, JIVE)
- Severity of the problem and available solutions vary by:
  - Number of endogenous vars: $J = 1$ vs. $J > 1$
  - Just-identified vs. overidentified case: $K = J$ vs. $K > J$
  - Few or many IV
- See Andrews, Stock, Sun (ARE 2019) but the debate continues

Sensitivity to exogeneity violations

- Recall $Y_i = \tau D_i + \varepsilon_i$ and $D_i = \pi Z_i + u_i$ with $\mathrm{Cov}[Z_i, u_i] = 0$
- When $\mathrm{Cov}[Z_i, \varepsilon_i] \neq 0$, it is not guaranteed that IV is less biased than OLS, *especially* when the IV is weak: for a scalar IV,

$$\tau_{OLS} - \tau = \frac{\mathrm{Cov}[\varepsilon_i, D_i]}{\mathrm{Var}[D_i]}, \qquad \tau_{IV} - \tau = \frac{\mathrm{Cov}[\varepsilon_i, Z_i]}{\mathrm{Cov}[D_i, Z_i]} = \frac{\mathrm{Cov}[\varepsilon_i, \pi Z_i]}{\mathrm{Cov}[D_i, \pi Z_i]}$$

  - Here $\mathrm{Var}[\pi Z_i] / \mathrm{Var}[D_i] = R^2_{FS}$ from the first-stage
  - If $R^2_{FS} = 0.01$, the covariance with the error term to go down by more > 100 times to reduce bias

Large asymptotic variance of IV

- For a scalar IV and under homoskedasticity $\mathrm{Var}\left[\varepsilon_i \mid Z_i\right] = \sigma^2$,

$$\mathrm{Var}\left[\sqrt{N}\left(\hat{\tau}_{IV} - \tau\right)\right] \approx \frac{\sigma^2}{\mathrm{Var}\left[\pi Z_i\right]} = \frac{\sigma^2}{\mathrm{Var}\left[D_i\right] \cdot R^2_{FS}}$$

- Compare with OLS when OLS is unbiased:

$$\mathrm{Var}\left[\sqrt{N}\left(\hat{\tau}_{OLS} - \tau\right)\right] \approx \frac{\sigma^2}{\mathrm{Var}\left[D_i\right]}$$

F-statistics

First-stage F-statistics play a key role when $J = 1$

- They test for $\pi = 0$ (and not also $\kappa_{FS} = 0$!) in
$$D_i = \pi' Z_i + \kappa_{FS}' X_i + u_i$$
- $F = \frac{1}{K}\hat{\pi}' \cdot \hat{\sum}_{\pi\pi}^{-1} \cdot \hat{\pi}$ where $\hat{\sum}_{\pi\pi}$ is the estimate of $\mathrm{Var}\left[\hat{\pi}\right]$
    - With homoskedastic $\hat{\sum}_{\pi\pi}$: "Cragg–Donald F-statistic"
    - With robust $\hat{\sum}_{\pi\pi}$: "Kleibergen–Paap F-statistic"
    - When $K = 1$, $F = t^2$

Finite-sample bias towards OLS

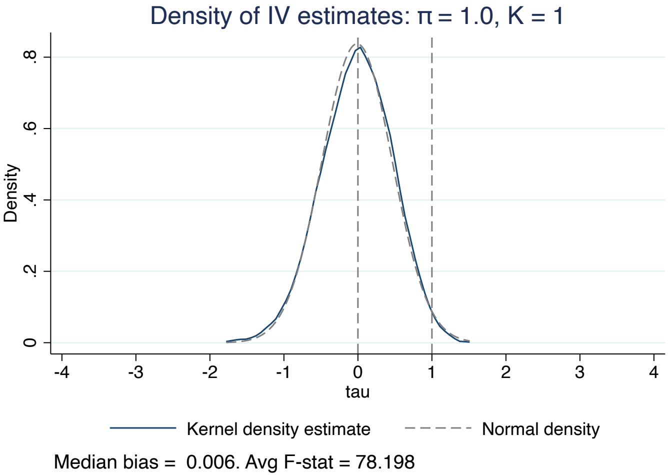
- Bound et al. (1995): with scalar D_i and homoskedasticity, IV is biased towards OLS in finite samples when the first-stage F-stat is small:

$$\mathbb{E}\left[\hat{\tau}_{IV} - \tau\right] \approx \frac{\mathrm{Cov}\left[\varepsilon_i, u_i\right]}{\mathrm{Var}\left[u_i\right]} \cdot \frac{1}{F+1},$$

  where the first factor equals OLS bias when $\pi = 0$ (and so $D_i = u_i$)
- Intuition:
  - With many IVs, even if all of them are completely irrelevant, the first-stage overfits and produces fitted values $\hat{D}_i$ approximating D_i
  - With few IVs, this will happen only a bit — but the F-stat still reflects that
- *Note:* In just-identified cases with Gaussian u_i, $\mathbb{E}\left[\hat{\tau}_{IV}\right]$ doesn't exist (but think median bias instead)

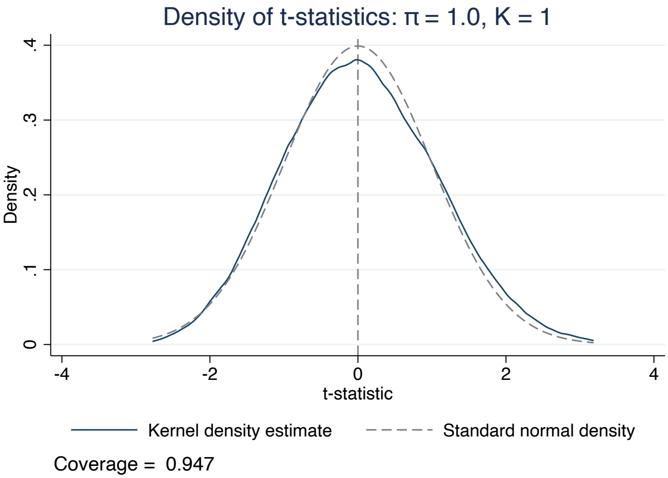
Monte Carlo: Single strong IV

Set $N = 300$, true effect $\tau = 0$, OLS $= 1$. Strong IV: $\pi = 1$



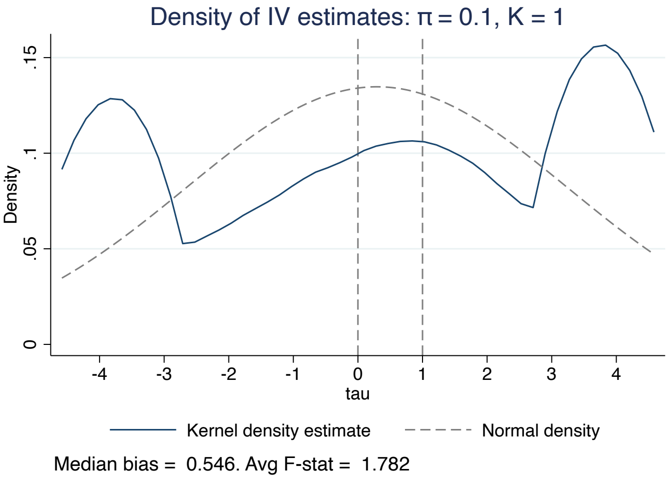
Tests and coverage: Single strong IV

With single strong IV, distribution of t-stats $\approx \mathcal{N}(0, 1)$



Monte Carlo: Weak IV

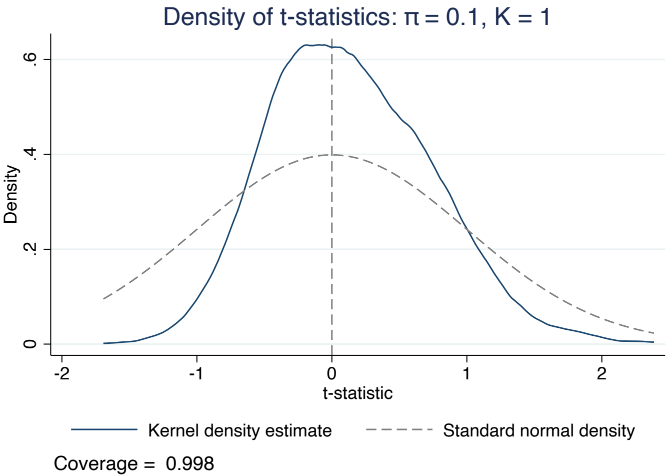
Now set $\pi = 0.1 \implies$ Non-Gaussian distribution and bias



(The distribution is winsorized at $|\hat{\tau}| = 4$)

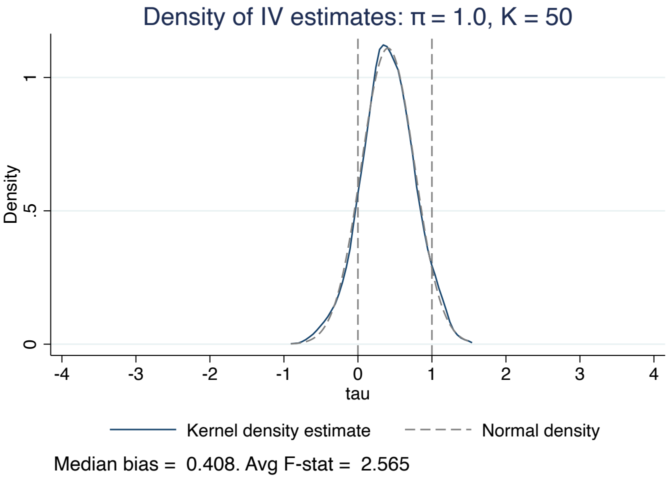
Tests and coverage: Weak IV

Angrist and Kolesár (2024): with $K = J = 1$, unless endogeneity is super strong, SE will be very high and coverage is conservative



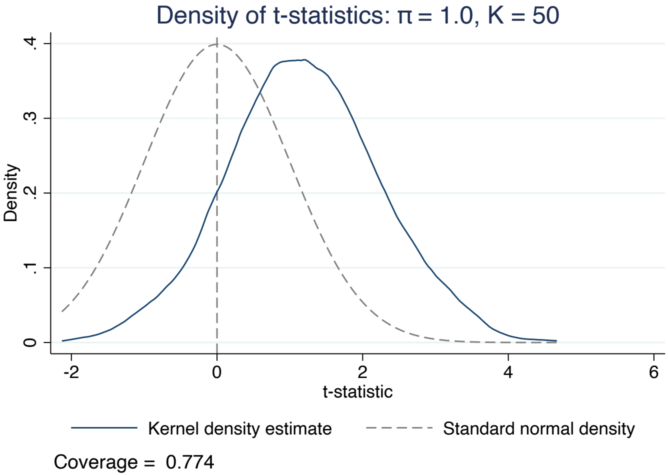
Monte Carlo: Many IV

Now set $\pi = 1$ but add $K - 1$ irrelevant IVs for $K = 50 \implies$ Also bias



Tests and coverage: Many IV

With many IV, first-stage overfits and appears strong $\Longrightarrow$ undercovered confidence intervals



Many IV bias in practice

- Bound et al. (1995) reanalyze many-instrument 2SLS estimates of Angrist and Krueger (1991)
    - 30 IVs: QoB interacted with year of birth
    - 180 IVs: additionally, QoB interacted with state of birth
- They use randomly generated QoB instead of actual ones
    - Find coefs centered near OLS, possibly significant

What to do?

- Report OLS as a reference point
- Report and visualize first-stage estimates
- Report the appropriate first-stage F-statistic on the instruments
- Report and visualize reduced-form estimates: *"if the reduced form estimates are not significantly different from zero, ... the effect of interest is either absent or the instruments are too weak to detect it"* (Angrist and Krueger, 2001)
- Report "weak-robust" confidence intervals
- With $K > J$, try alternative estimators: just-identified IV; LIML or JIVE

Pre-testing for weak IVs

- Test for weak IV for $J = 1$ (Staiger and Stock (1997), Stock and Yogo (2005))
  - Understood as worst-case bias (in % of OLS bias) or worst-case distortion of Wald test coverage
  - Test: $F >$ special critical value. Heuristic: $F > 10$. Requires homoskedasticity
- Montiel Olea and Pflueger (2013) extend to heteroskedastic case:
  - $K = 1$: use robust (a.k.a. Kleibergen–Paap) F-stat, with Stock & Yogo critical values
  - $K > 1$: use "effective F-stat" they propose ($\neq$ homoskedastic or robust), with different critical values (but not too different from Stock and Yogo)

Pre-testing for weak IVs (2)

- Angrist and Kolesár (2024) for $J = K = 1$, no need to worry about F-stats
  - The worst-case is too extreme for most practical situations
- For $K > 1$, Sanderson and Windmeijer (2016) test for homoskedastic case
  - Can correctly reject even if each first-stage has a high F-stat
  - See Lewis and Mertens (2024) on the worst-case with $K > 1$ and heteroskedasticity

Weak-robust inference

- Consider $J = 1$ but any K
- Exogeneity implies $\rho = \tau\pi$ even with weak IVs
  - $(\hat{\rho}, \hat{\pi})$ are jointly asy. normal; can get robust variance estimate $\hat{\Sigma}$
- To test $\tau = b$, use the Anderson-Rubin statistic

$$AR = (\hat{\rho} - \hat{\pi} b)' \, \hat{V}^{-1}(b) \, (\hat{\rho} - \hat{\pi} b) \, , \qquad \hat{V}(b) = \hat{\Sigma}_{\rho\rho} - b \left(\hat{\Sigma}_{\rho\pi} + \hat{\Sigma}_{\pi\rho} \right) + b^2 \hat{\Sigma}_{\pi\pi}$$

  - Under the null, $AR \sim \chi^2_K \implies$ Reject if $AR > \chi^2_{K,1-\alpha}$
  - Construct confidence interval by test inversion: collect all b that are not rejected
  - With homoskedastic $\hat{\Sigma}$ or with $K = 1$, there is an analytical formula for the CI (Mikusheva (2010))

Properties of AR confidence intervals

- CI can be infinite ($= \mathbb{R}$ or, weirdly, $\mathbb{R} \backslash [a_1, a_2]$)
  - Not surprising: when $\pi = 0$ there is no information about τ
  - Indeed, CI is infinite whenever can't reject $\pi = 0$ by Wald test
- With $K = J = 1$:
  - CI $\neq \emptyset$: $\hat{\tau} = \hat{\rho}/\hat{\pi}$ is always in it (since $AR(\hat{\tau}) = 0$), but needn't be in the middle
  - Test is asymptotically efficient (see Andrews, Stock, Sun, Section 5.1.1)
- With $K > J = 1$:
  - CI is empty if overidentifying restrictions are rejected ($\hat{\rho}$ is far from $\tau\hat{\pi}$ for all τ)
  - Since the test is sensitive to overid. restrictions, it's inefficient for testing $\tau = b$
  - Improvements have been proposed (see Andrews et al. Section 5.2)
- With $J > 1$, test inversion produces inconvenient J-dimensional confidence sets
  - See Andrews et al. (Section 5.3) on CIs for individual coefficients

Alternative estimators with overidentification: LIML

Limited information maximum likelihood (LIML):

- Proposed as joint MLE estimator for the structural equation + first-stage, assuming homoskedastic normal errors
- Minimizes (homoskedastic) $AR(b) \Longrightarrow$ lies within the AR conf. interval (whenever non-empty)
- Has much smaller median bias than 2SLS under weak IV
- But higher variance $\Longrightarrow$ MSE comparison unclear (e.g. Blomquist and Dahlberg (1999))
- Doesn't have moments $\Longrightarrow$ can have extreme outliers
- Doesn't work well with heterogeneous effects (Kolesar (2013))

Alternative estimators: JIVE

Jackknife IV estimator (JIVE; Angrist, Imbens, and Krueger (1999)):

- Avoids overfitting in the first-stage by a leave-out procedure
- For each i estimate the first-stage $\hat{\pi}_{-i}$ excluding observation i
- Obtain fitted values $\tilde{D}_i = \hat{\pi}'_{-i} Z_i$ and set $\hat{\tau}_{\text{jackknife}} = (\tilde{D}' D)^{-1} \tilde{D} Y$
- There is a way to obtain $\hat{\tau}_{\text{jackknife}}$ without actually running the first-stage N times
- Small median bias but variance is typically even worse than LIML

Aside: Two-Sample IV

- For simplicity, consider just-identified IV, perhaps with covariates
- Usual situation: observe $(\boldsymbol{Z}, \boldsymbol{D}, Y)$, compute $\hat{\boldsymbol{\tau}} = (\boldsymbol{Z}'\boldsymbol{D})^{-1}\boldsymbol{Z}'Y$
- Angrist (1990) Vietnam lottery paper:
  1. Social Security data: earnings Y, date of birth $\rightarrow$ lottery number Z, covariates (race, year of birth) — but not veteran status D
  2. Veteran records: Z, D, covariates — but not Y

  - *Note*: Y and D never observed together
- Two-sample IV: $\hat{\boldsymbol{\tau}} = \left(\frac{1}{N_2}\boldsymbol{Z}_2'\boldsymbol{D}_2\right)^{-1}\frac{1}{N_1}\boldsymbol{Z}_1'Y$
- Two-sample 2SLS: $\hat{\boldsymbol{\tau}} = \left(\hat{\boldsymbol{D}}_1'\hat{\boldsymbol{D}}_1\right)^{-1}\hat{\boldsymbol{D}}_1'Y$ where $\hat{\boldsymbol{D}}_1 = \boldsymbol{Z}_1\left(\boldsymbol{Z}_2'\boldsymbol{Z}_2\right)^{-1}\boldsymbol{Z}_2\boldsymbol{D}_2$ are fitted values from the 2nd-sample first-stage computed for the 1st sample
- Inoue and Solon (2010): they differ; two-sample TSLS is more efficient; derive SE

References I

Andrews, I., J. Stock, and L. Sun (2019): “Weak Instruments in Instrumental Variable Regression: Theory and Practice,” *Annual Review of Economics*.

Angrist, J. and M. Kolesár (2024): “One Instrument to Rule Them All: The Bias and Coverage of Just-ID IV,” *Journal of Econometrics*, arXiv:2110.10556 [econ, stat].

Angrist, J. D. (1990): “Lifetime earnings and the Vietnam era draft lottery: evidence from social security administrative records,” *American Economic Review*, 80, 313–336.

Angrist, J. D. and W. N. Evans (1998): “Children and Their Parents ' Labor Supply : Evidence from Exogenous Variation in Family,” *American Economic Association*, 88, 450–477.

Angrist, J. D., G. W. Imbens, and A. Krueger (1999): “Jackknife instrumental variables estimation,” *Journal of Applied Econometrics*, 67, 57–67.

Angrist, J. D. and A. B. Krueger (1991): “Does Compulsory School Attendance Affect Schooling and Earnings?” *Quarterly Journal of Economics*.

——— (2001): “Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments,” *Journal of Economic Perspectives*.

References II

Blomquist, S. and M. Dahlberg (1999): "Small sample properties of LIML and jackknife IV estimators: experiments with weak instruments," *Journal of Applied Econometrics*, 14, 69–88.

Bound, J., D. a. Jaeger, and R. M. Baker (1995): "Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak," *Journal of the American Statistical Association*, 90, 443–450, iSBN: 01621459.

Buckles, K. S. and D. M. Hungerman (2013): "Season of Birth and Later Outcomes: Old Questions, New Answers," *Review of Economics and Statistics*, 95, 711–724.

Chalfin, A. and J. McCrary (2018): "Are U.S. Cities Underpoliced? Theory and Evidence," *Review of Economics and Statistics*, 100, 167–186.

Duranton, G. and M. A. Turner (2012): "Urban Growth and Transportation," *Review of Economic Studies*, 79, 1407–1440.

Gallen, T. and B. Raymond (2023): "Broken Instruments," *Working Paper*.

Inoue, A. and G. Solon (2010): "Two-Sample Instrumental Variable Estimators," *Review of Economics and Statistics*, 92, 557–561.

References III

Kolesar, M. (2013): "Estimation in an Instrumental Variables Model With Treatment Effect Heterogeneity," 1–45.

Lewis, D. J. and K. Mertens (2024): "A Robust Test for Weak Instruments for 2SLS with Multiple Endogenous Regressors," *Working Paper*.

Mellon, J. (2024): "Rain, Rain, Go Away: 194 Potential Exclusion-Restriction Violations for Studies Using Weather as an Instrumental Variable," *American Journal of Political Science*.

Mikusheva, A. (2010): "Robust confidence sets in the presence of weak instruments," *Journal of Econometrics*, 157, 236–247, publisher: Elsevier B.V.

Montiel Olea, J. L. and C. Pflueger (2013): "A robust test for weak instruments in Stata," *Journal of Business and Economic Statistics*, 31, 358–369, arXiv: The Stata Journal ISBN: 0735-0015\r1537-2707.

Sanderson, E. and F. Windmeijer (2016): "A weak instrument F -test in linear IV models with multiple endogenous variables," *Journal of Econometrics*, 190, 212–221.

Sarsons, H. (2015): "Rainfall and Conflict: A Cautionary Tale," *Journal of Development Economics*, 115, 62–72.

References IV

Staiger, D. and J. H. Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586.

Stock, J. H. and M. Yogo (2005): "Testing for Weak Instruments in Linear IV Regression," in *Identification and Inference for Econometric Models*, 80–108.