Part D: Instrumental Variables

D1: IV Idea and Mechanics; Weak IV

Kirill Borusyak

ARE 213 Applied Econometrics

UC Berkeley, Fall 2023

IV outline

1. IV idea, mechanics, weak IV
2. IV with treatment effect heterogeneity
3. Shift-share IV designs, formula instruments, recentering
4. Examiner designs ("judge IVs")
5. A bit on control function approaches

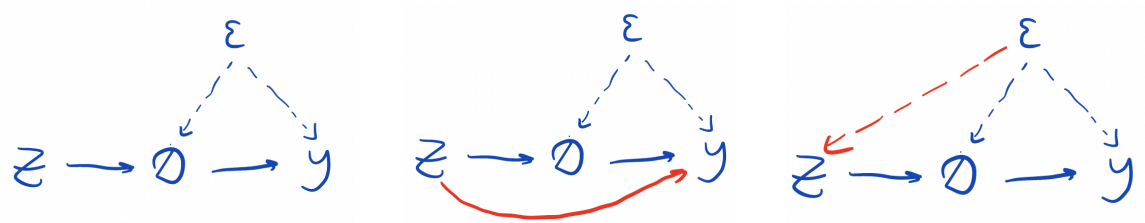
D1 outline

1. Setting and Examples
2. IV Mechanics
3. Weak and Many (Weak) IVs

Readings:

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

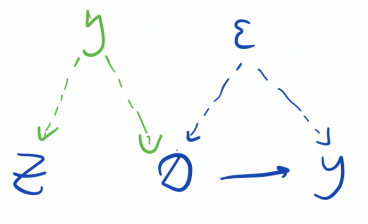
Instrumental variable DAGs



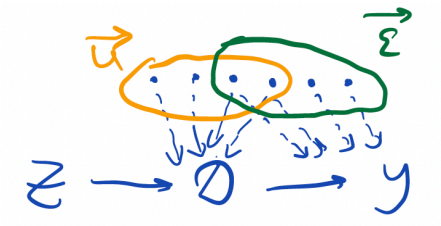
- **Exclusion**: $Y(d, z) = Y(d, z') \equiv Y(d)$ for all d, z, z'
- **Independence**: $Z \perp\!\!\!\perp Y(d)$ for all d
- Exclusion + independence = instrument **exogeneity** (a.k.a. **orthogonality**)
  - Z is only correlated with Y because of the causal effect of D on Y
  - Calling Z an "excluded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

Instrumental variable DAGs (2)

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



IV with constant effects



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

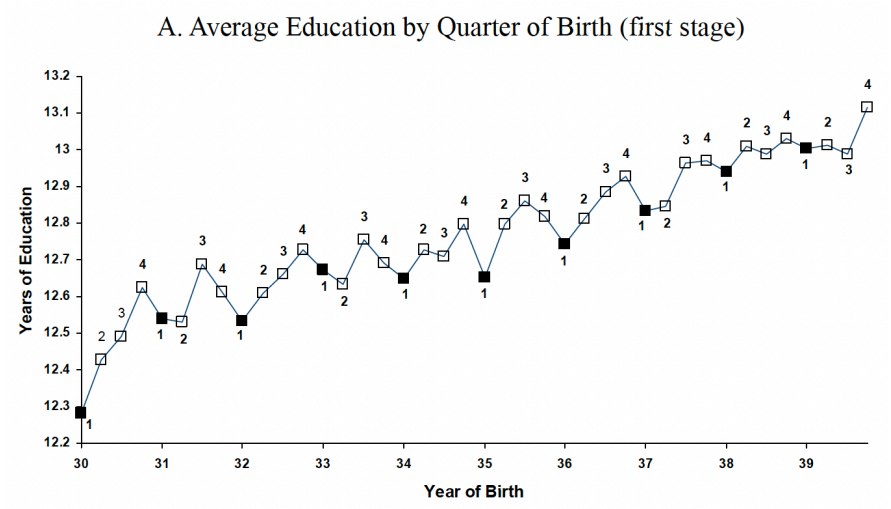
Example 1: Randomized encouragement designs

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

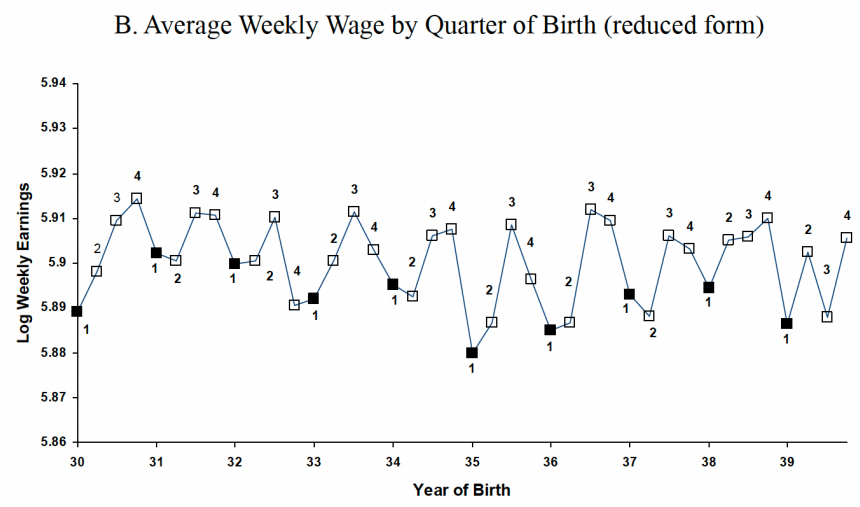
Example 2: Angrist and Krueger (1991)

- Estimate returns to schooling: $D =$ years of schooling, $Y =$ log earnings
- $Z =$ quarter of birth dummies
- Relevance: structure of compulsory schooling laws
  - Children born in Q4 start school a year earlier 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
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
Year of Birth

Reduced form



Exogeneity of QoB

- Exclusion: no other mechanism through which QoB affects earnings
  - Bound, Jaeger, 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 3: 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 ...

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

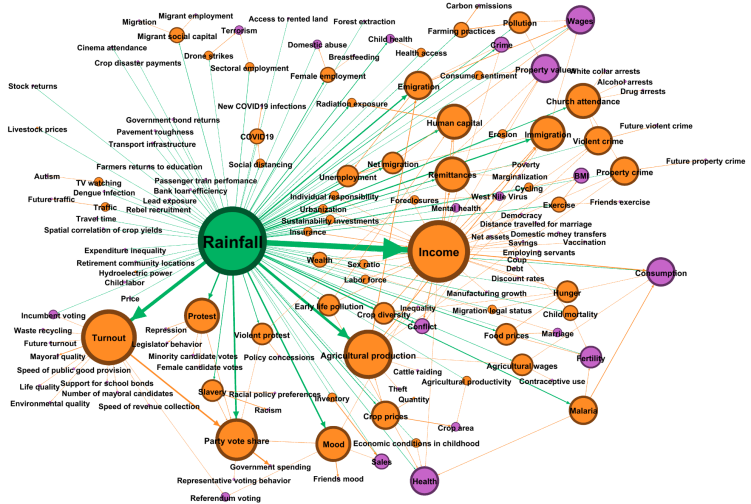
> The *a priori* case for thinking that 1898 railroad kilometres satisfy exogeneity condition (17) rests on the length of time since these railroads were built and the fundamental changes in the nature of the economy in the intervening years. The rail network was built, for the most part, during and 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 4: Overusing an instrument

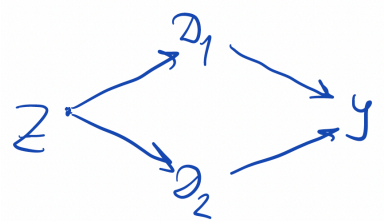
- Gallen and Raymond (2023) identify several instruments used with many endogenous variables and outcomes
  - Presence of bodies of water and changes in elevation
  - Sibling characteristics, e.g. having a twin
  - Ethnolinguistic fractionalization
  - Religion
  - Weather, e.g. rainfall

Usage of rainfall IV

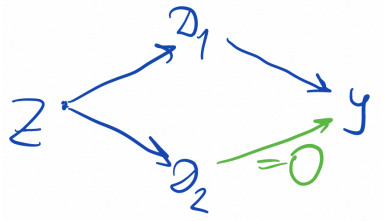


From Mellon (2023); orange = D, purple = Y; circle size = # of studies

Often (but not always) a problem



Problem



OK

OK (see Mellon 2023)

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 villages in India with no first-stage — because of dams
  - Yet, finds a significant reduced-form

Example 5: 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

IV estimand and estimator

- 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**"):

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

- 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 controls

- Expand first-stage $D_i = \pi Z_i + \kappa_{FS}' X_i + u_i$ and structural eqn $Y_i = \tau D_i + \kappa_{SE}' X_i + \varepsilon_i$
- Assume exogeneity $\mathbb{E}[Z_i \varepsilon_i] = 0$; without loss, $\mathbb{E}[X_i \varepsilon_i] = \mathbb{E}[X_i u_i] = \mathbb{E}[Z_i u_i] = 0$
- Moment conditions: for $\boldsymbol{D}_i = \begin{pmatrix} D_i \\ X_i \end{pmatrix}$, $\boldsymbol{\tau} = \begin{pmatrix} \tau \\ \kappa_{SE} \end{pmatrix}$, and $\boldsymbol{Z}_i = \begin{pmatrix} Z_i \\ X_i \end{pmatrix}$,

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

  if $\mathbb{E}[\boldsymbol{Z}_i \boldsymbol{D}_i']$ is full-rank (relevance)
- By Frisch–Waugh–Lovell: for $\tilde{Z}_i$ = residual from regressing Z_i on X_i and if $\pi \neq 0$,

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

- 2SLS also works

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 of IVs $\Pi' Z_i$ (for $K \times J$ matrix Π with full rank J) would produce an 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 $\Pi' Z_i$
- Instrumenting D_i with $\Pi' Z_i$ = regressing Y on $\Pi' Z_i \Longrightarrow$ 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

Outline

Weak IV: Overview

- What is the weak IV situation?
  - $K = J = 1$: low (partial) correlation of D_i and Z_i
  - $K > J = 1$: low (partial) correlation of D_i with all or most Z_{ki}
  - $K = J > 1$: the 1st stage matrix Π is close to rank-deficient $\Longrightarrow$ 2nd stage of 2SLS is close to perfect multicollinearity
- Problems:
  - Large sensitivity to violations of IV exogeneity
  - Large variance of IV estimates
  - Failure of the asymptotic approximation (and bootstrap):
    - 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 (2)

- 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 $\mathrm{Cov}\left[Z_i, u_i\right] = 0$

- When $\mathrm{Cov}\left[Z_i, \varepsilon_i\right] \neq 0$, it is not guaranteed that IV is less biased than OLS, *especially* when the IV is weak: for a scalar IV,

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

- *Note:* $\mathrm{Var}\left[\pi Z_i\right] / \mathrm{Var}\left[\pi Z_i + u_i\right] = R^2_{FS}$ from the first-stage

Large asymptotic variance of IV

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

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

Finite-sample bias towards OLS

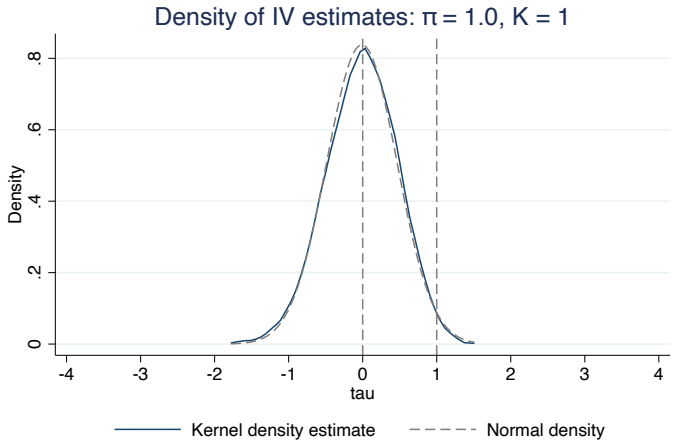
- Bound, Jaeger, Baker (1995): with scalar D_i, IV is biased towards OLS in finite samples when the first-stage F-stat on all IVs (but not including covariates X_i) is small:

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

  where the first factor equals OLS bias when $\pi = 0$ (and $D_i = u_i$)
- 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
- Don't treat this result literally:
  - In just-identified cases, $\mathbb{E}\left[\hat{\tau}_{IV}\right]$ doesn't exist (but think about median bias)
  - Homoskedasticity is required (see MHE Ch. 4.6.4)

Monte Carlo: Single strong IV

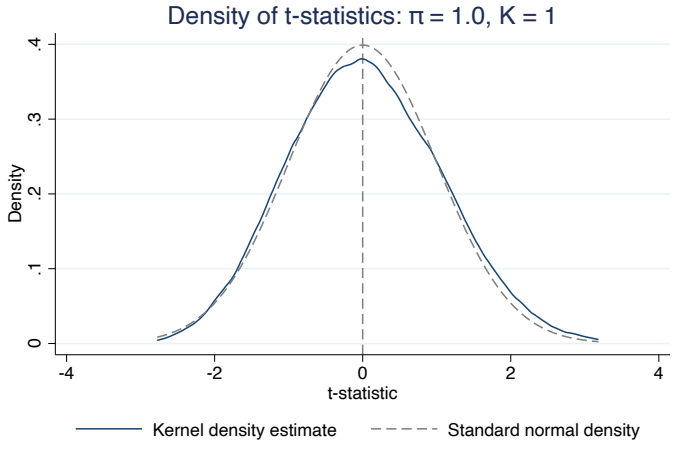
Set $N = 300$, true effect $\tau = 0$, OLS $= 1$. Strong IV: $\pi = 1$



Median bias = 0.006. Avg F-stat = 78.198

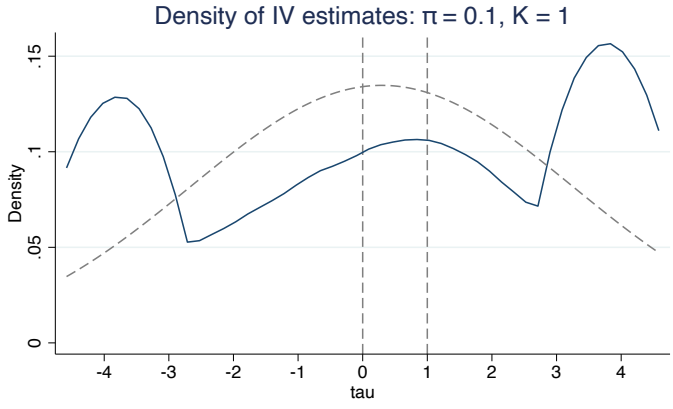
Tests and coverage: Single strong IV

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

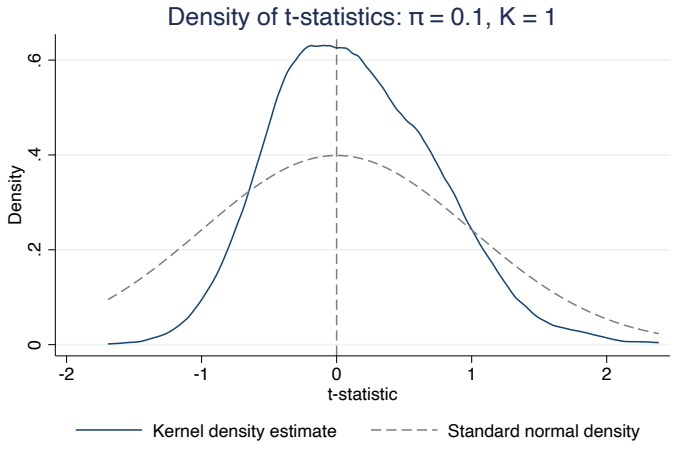


Median bias = 0.546. Avg F-stat = 1.782

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

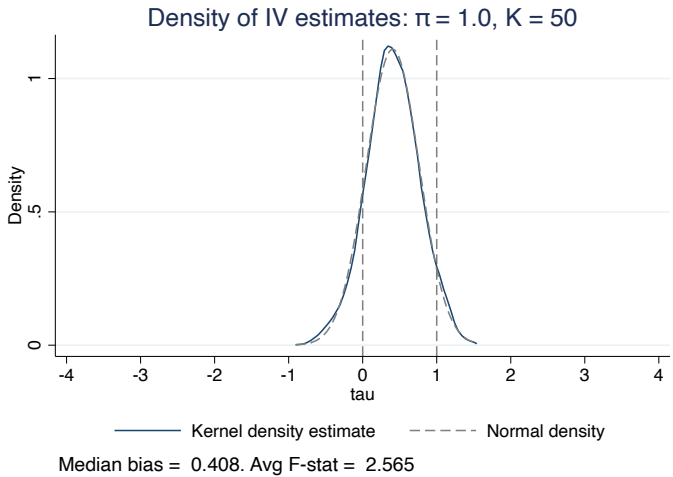
Tests and coverage: Weak IV

Angrist and Kolesar (2023): with $K = J = 1$, unless endogeneity is super strong, SE will be very high and coverage is conservative



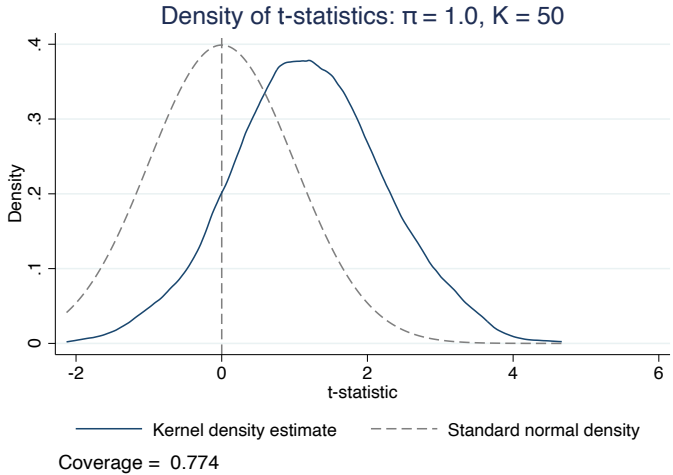
Monte Carlo: Many IV

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 and visualize first-stage estimates
- Report the first-stage F-statistic on the instruments (excluding controls)
- 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)
- With single discrete multi-valued/categorical Z_i, there is "visual IV" graph (see MHE Figure 4.1.2)
- 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 Kolesar (2023): 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 (WP 2022) 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, $\mathbb{R} \setminus [a_1, a_2]$)
  - 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

Limited information maximum likelihood (LIML):

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

Alternative estimators (2)

Jackknife IV estimator (JIVE; Angrist, Imbens, 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