

Part D: Instrumental Variables

D1: IV Idea and Mechanics; Weak IV

Kirill Borusyak

ARE 213 Applied Econometrics

UC Berkeley, Fall 2024

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")

D1 outline

1 Setting and Examples

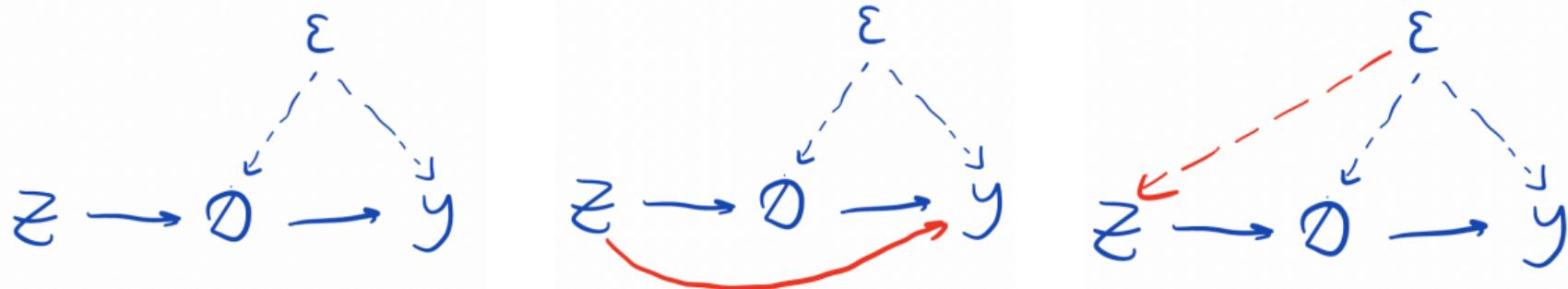
2 IV Mechanics

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)

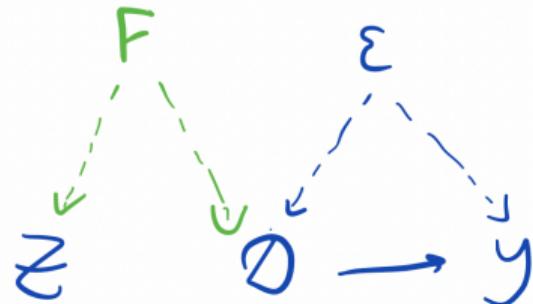
Instrumental variable DAGs



- **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 if there is a causal effect of D on Y
 - ▶ Calling Z an “excluded instrument” is somewhat misleading

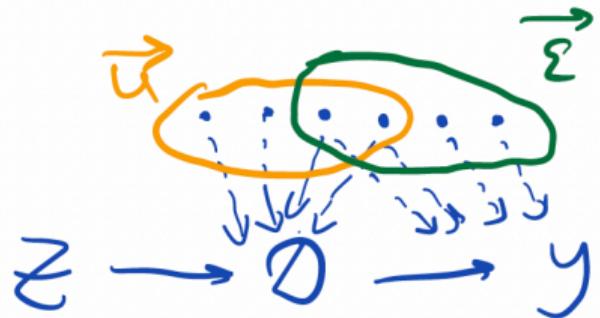
Instrumental variable DAGs (2)

- **Relevance:** $D(z) \neq D(z')$ with positive 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:** $Y_i = \mathcal{Y}(D_i, \vec{\varepsilon}_i) \implies Y_i = \tau D_i + \varepsilon_i$
- **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:** $Y_i = \tau\pi Z_i + (\varepsilon_i + \tau u_i) \equiv \rho Z_i + e_i$

Example 1: Encouragement designs

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

RCT with two-sided noncompliance

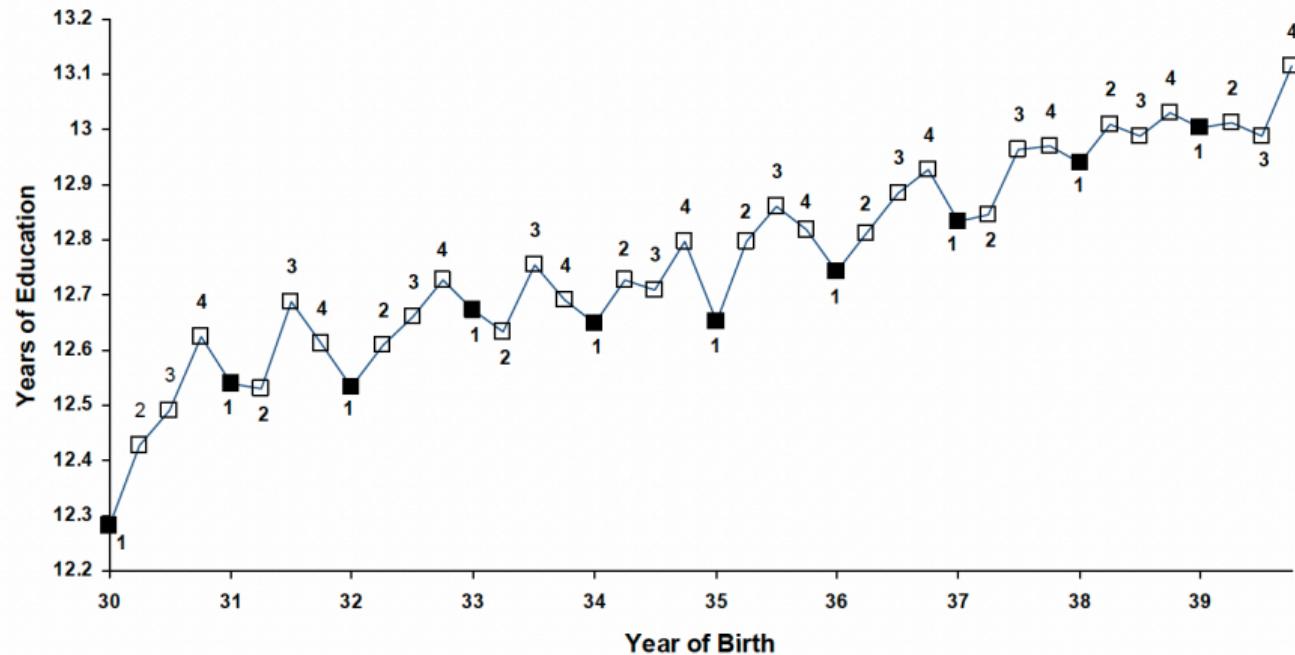
- Angrist (1990): How does serving in the Vietnam war (D_i) affect future earnings (Y_i)?
 - ▶ Draft was based on televised lottery
 - ▶ Men with low lottery numbers were enlisted first
 - ▶ Still, some of them can avoid serving, while others can volunteer
 - ▶ Z_i = bins of lottery numbers

Example 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 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

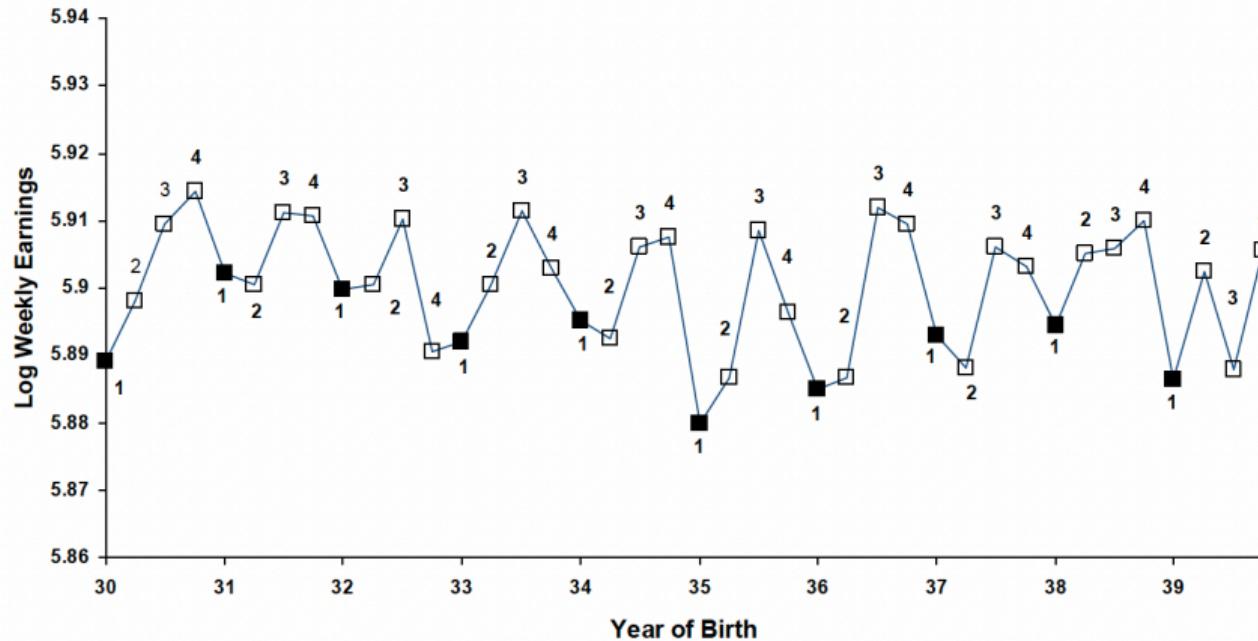
First stage

A. Average Education by Quarter of Birth (first stage)



Reduced form

B. Average Weekly Wage by Quarter 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 disproportionately 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.

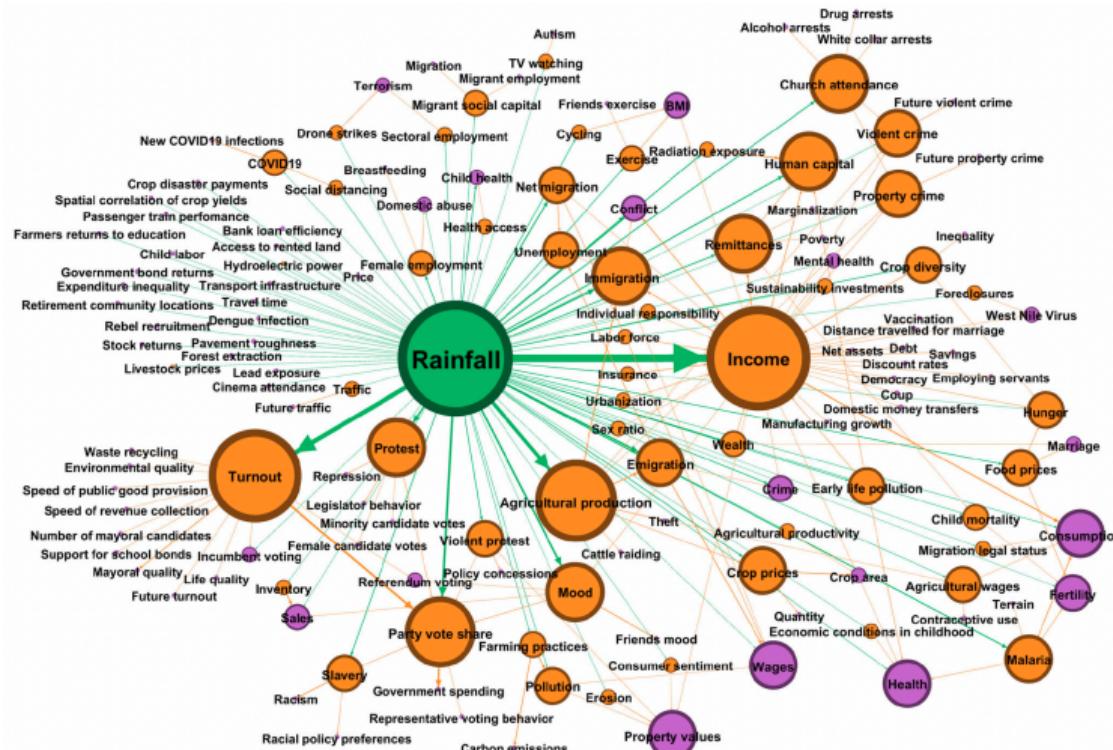
Furthermore, the rail network was constructed by private companies who were looking to make a profit from railroad operations in the not too distant future (Fogel, 1964; Fishlow, 1965). It is difficult to imagine how a rail network built for profit during the civil war and the industrial revolution could 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 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



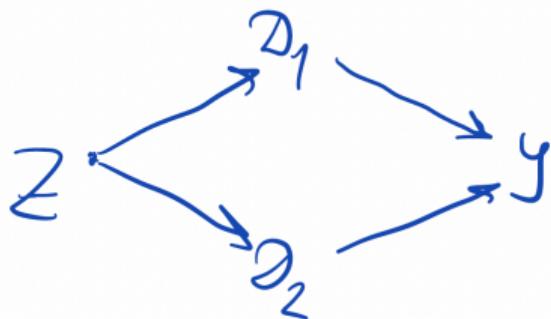
From Mellon (2024); orange = D , purple = Y ; circle size = # of studies

Usage of rainfall IV

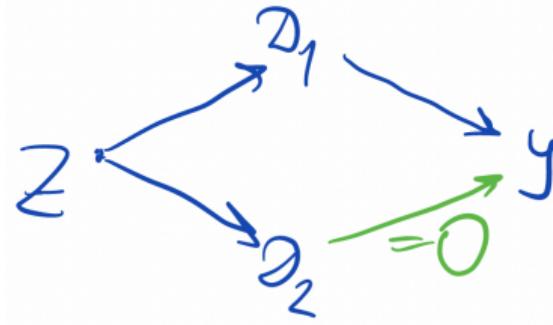


(From Twitter, <https://x.com/ernitedeschi/status/1848143618713997680>)

Often (but not always) a problem



Problem



OK



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 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

$$Q = \tau_d P + \varepsilon_d,$$

$$Q = \tau_s P + \varepsilon_s$$

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

$$Q = \tau_d P + \varepsilon_d \quad \iff \quad P = \frac{1}{\tau_d} Q - \frac{1}{\tau_d} \varepsilon_d$$

Example 6: 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, $D_i = \text{having 3+ kids}$
 - ▶ Z_{1i} : twins at second birth
 - ▶ Z_{2i} : first two kids are both boys or both girls

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) $\not\Rightarrow$ 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 Weak and Many (Weak) IVs

IV estimand and estimator

- With constant effects, exogeneity implies moment condition: $\text{Cov}[Z_i, Y_i - \tau D_i] = 0$
- Thus, if $\text{Cov}[Z_i, D_i] \neq 0$,

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

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

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

- 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"):

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

- 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 $\mathbf{D}_i = (D_i, X_i)'$, $\boldsymbol{\tau} = (\tau, \kappa_{SE})'$, and $\mathbf{Z}_i = (Z_i, X_i)'$,

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

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

- By Frisch–Waugh–Lovell: for $\tilde{Z}_i = \text{residual from regressing } Z_i \text{ on } X_i$ and if $\pi \neq 0$,

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

- 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:

$$\mathbb{E}[Z_i(Y_i - D'_i\tau)] = 0$$

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

$$\mathbb{E}[\Pi' Z_i(Y_i - D'_i\tau)] = 0 \implies \tau = (\Pi' \mathbb{E}[Z_i D'_i])^{-1} \Pi' \mathbb{E}[Z_i Y_i]$$

- 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 :

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

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 $X_i = \text{instrumenting } D_i \text{ with its FWL residual}$

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

1 Setting and Examples

2 IV Mechanics

3 Weak and Many (Weak) IVs

Weak IV: Overview

- What is the weak IV situation?
 - ▶ $K = J = 1$: 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}
 - ▶ $J > 1$: the 1st stage matrix Π is close to rank-deficient \implies 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, 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 (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

Weak IV: Overview (3)

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 $\text{Var} [\hat{\pi}]$

- ▶ With homoskedastic $\hat{\sum}_{\pi\pi}$: “Cragg–Donald F -statistic”
- ▶ With robust $\hat{\sum}_{\pi\pi}$: “Kleibergen–Paap F -statistic”
- ▶ When $K = 1$, $F = t^2$

Sensitivity to exogeneity violations

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

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

- ▶ Here $\text{Var}[\pi Z_i] / \text{Var}[D_i] = R_{FS}^2$ from the first-stage
- ▶ If $R_{FS}^2 = 0.01$, 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 $\text{Var} [\varepsilon_i | Z_i] = \sigma^2$,

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

- Compare with OLS when OLS is unbiased:

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

Finite-sample bias towards OLS

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

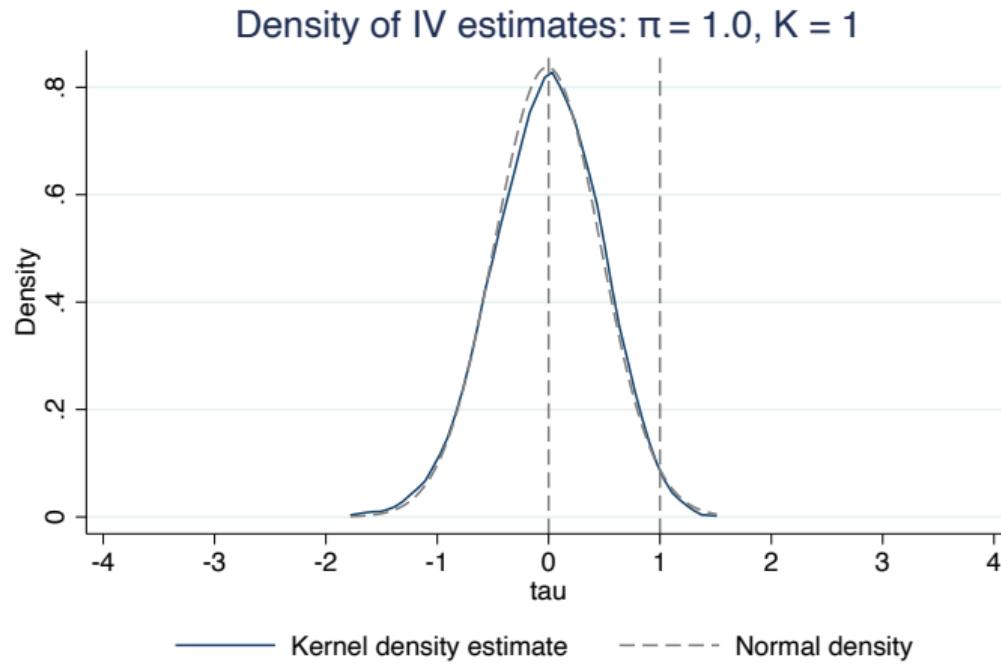
$$\mathbb{E} [\hat{\tau}_{IV} - \tau] \approx \frac{\text{Cov} [\varepsilon_i, u_i]}{\text{Var} [u_i]} \cdot \frac{1}{F+1},$$

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, $\mathbb{E} [\hat{\tau}_{IV}]$ doesn't exist (but think median bias instead)

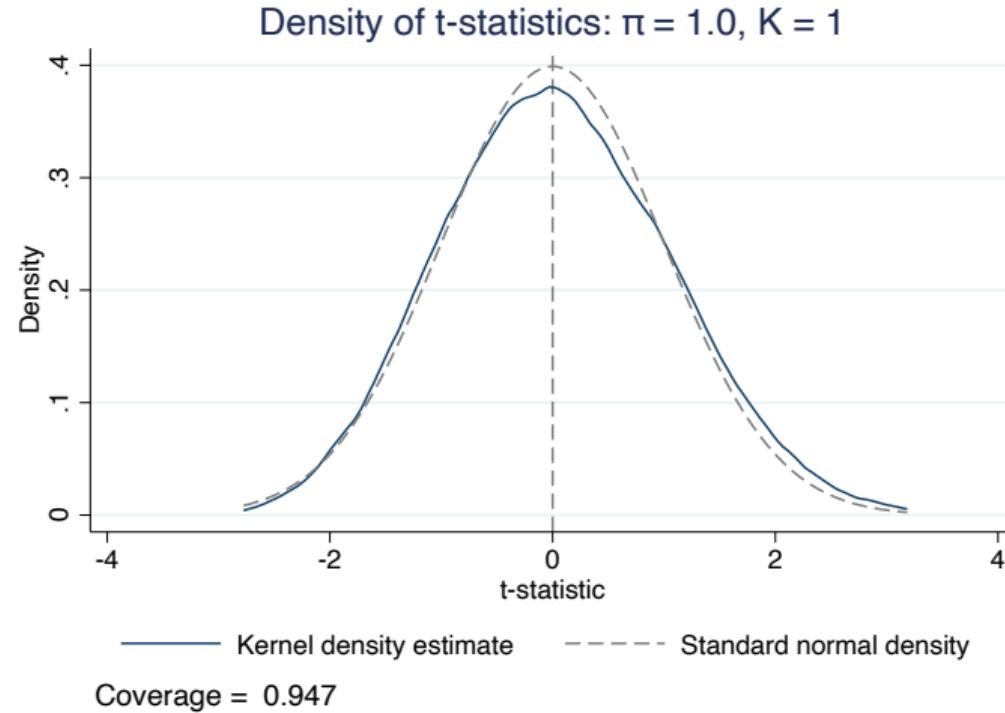
Monte Carlo: Single strong IV

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



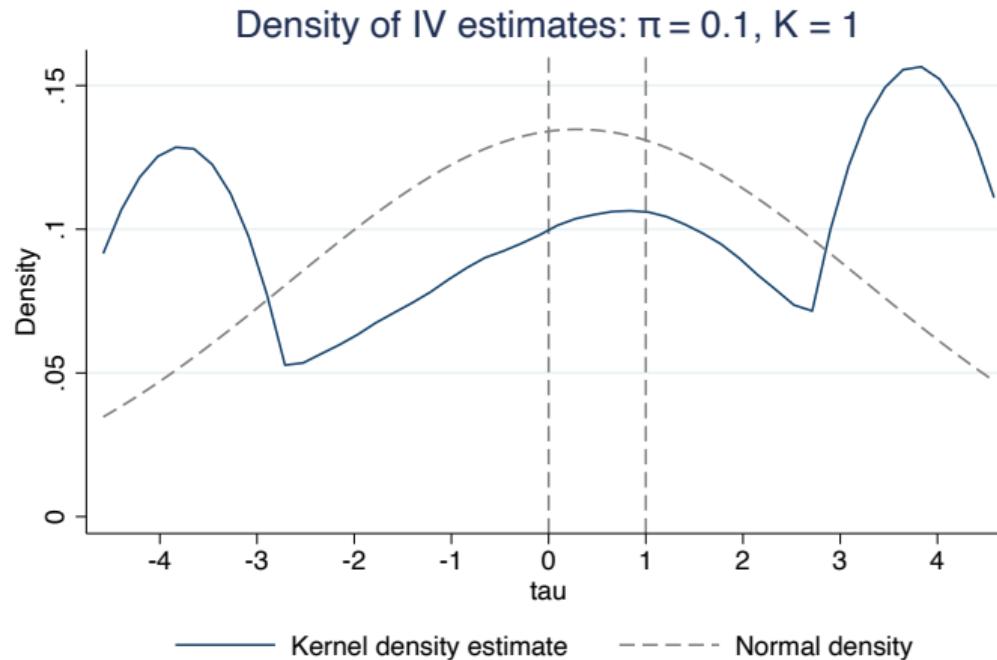
Tests and coverage: Single strong IV

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

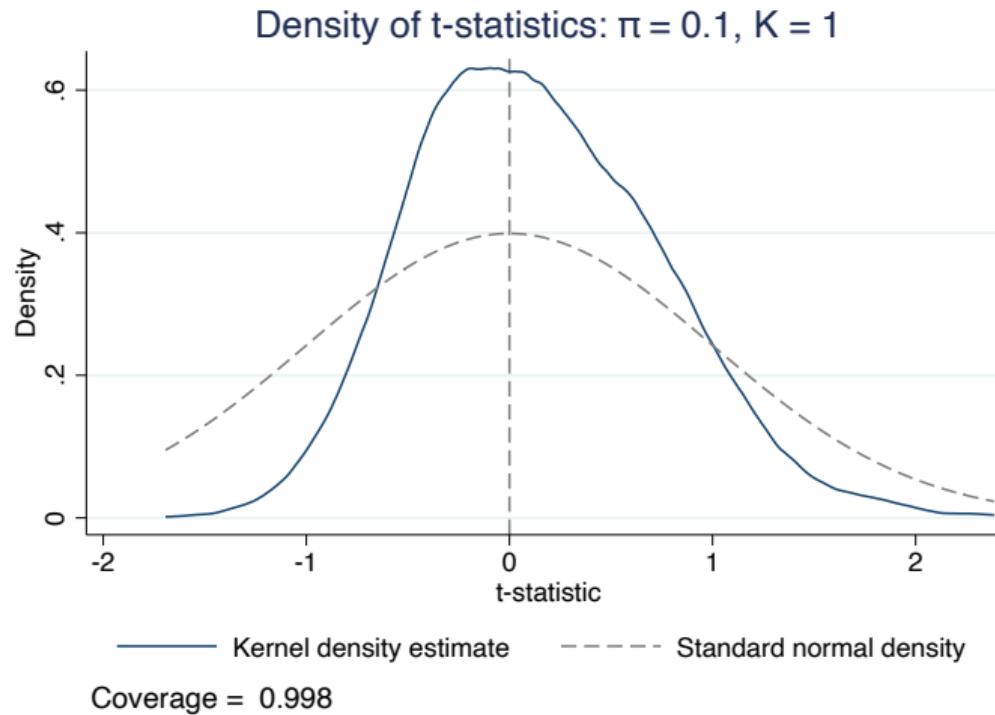


Median bias = 0.546. Avg F-stat = 1.782

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

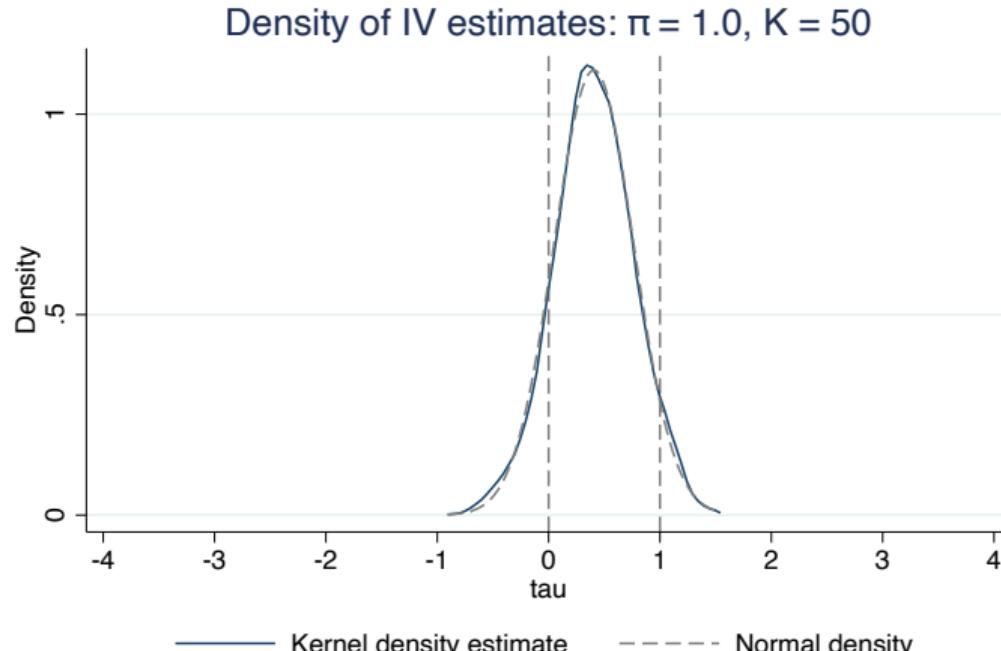
Tests and coverage: Weak IV

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



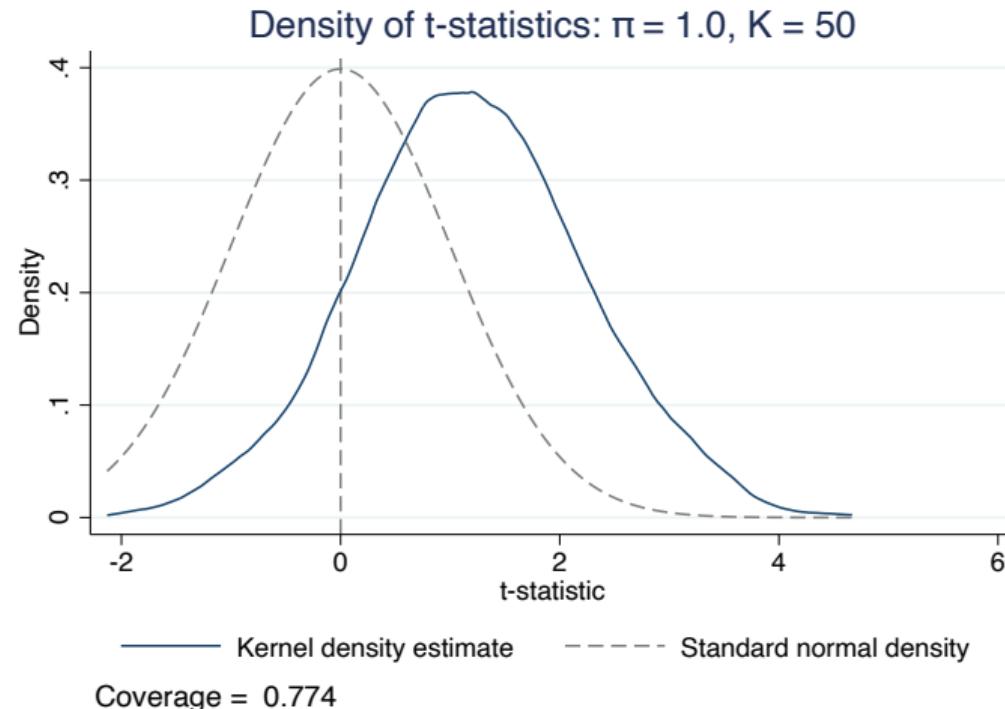
Monte Carlo: Many IV

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 \Rightarrow 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 coeffs centered near OLS, possibly significant

What to do?

- 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 JEP)
- 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 (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 (WP 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

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

- ▶ Under the null, $AR \sim \chi_K^2 \implies$ Reject if $AR > \chi_{K,1-\alpha}^2$
- ▶ 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: LIML

Limited information maximum likelihood (LIML):

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

Alternative estimators: JIVE

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