Instrumental Variables

Understanding IV

MIXTAPE SESSIONS

Roadmap

- Where do (Good) Instruments Come From?
  - True Lotteries
  - Natural Experiments
  - Panel Data
- 2SLS Mechanics
  - Just-Identified IV
  - Overidentification
- Weak and Many Instruments
  - Weak IV
  - Many IVs

Subtlties of the Validity Condition

To apply IV, we need to make a good case for instrument validity (note we can always check relevance!)

Subtlties of the Validity Condition

To apply IV, we need to make a good case for instrument validity (note we can always check relevance!)

Consider our simple causal model, $Y_i = \alpha + \beta D_i + \varepsilon_i$. Validity, $Cov(Z_i, \varepsilon_i) = 0$, intuitively requires two distinct assumptions:

Subtlties of the Validity Condition

To apply IV, we need to make a good case for instrument validity (note we can always check relevance!)

Consider our simple causal model, $Y_i = \alpha + \beta D_i + \varepsilon_i$. Validity, $Cov(Z_i, \varepsilon_i) = 0$, intuitively requires two distinct assumptions:

- As-good-as-random assignment : individuals with higher/lower potential earnings face the same distribution of Z_i
- Exclusion : the "assignment" of Z_i only affects Y_i through D_i

Subtlties of the Validity Condition

To apply IV, we need to make a good case for instrument validity (note we can always check relevance!)

Consider our simple causal model, $Y_i = \alpha + \beta D_i + \varepsilon_i$. Validity, $Cov(Z_i, \varepsilon_i) = 0$, intuitively requires two distinct assumptions:

- As-good-as-random assignment: individuals with higher/lower potential earnings face the same distribution of Z_i
- Exclusion: the "assignment" of Z_i only affects Y_i through D_i

Confusingly, old-school econometrics texts sometimes refer to $Cov(Z_i, \varepsilon_i) = 0$ as the "exclusion restriction"

Subtlties of the Validity Condition

To apply IV, we need to make a good case for instrument validity (note we can always check relevance!)

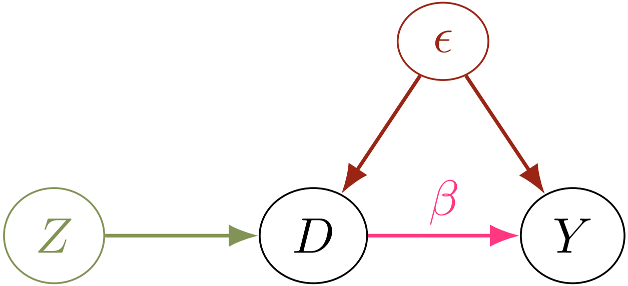
Consider our simple causal model, $Y_i = \alpha + \beta D_i + \varepsilon_i$. Validity, $Cov(Z_i, \varepsilon_i) = 0$, intuitively requires two distinct assumptions:

- As-good-as-random assignment: individuals with higher/lower potential earnings face the same distribution of Z_i
- Exclusion: the "assignment" of Z_i only affects Y_i through D_i

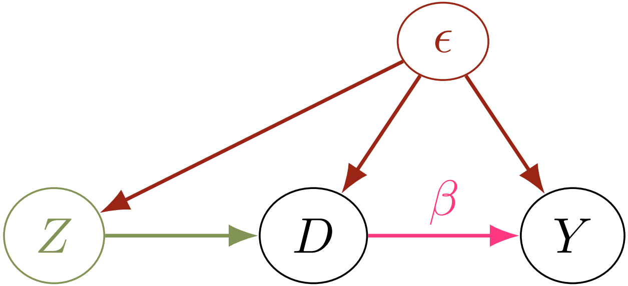
Confusingly, old-school econometrics texts sometimes refer to $Cov(Z_i, \varepsilon_i) = 0$ as the "exclusion restriction"

- More modern IV texts take care to distinguish between these two conceptually distinct requirements...

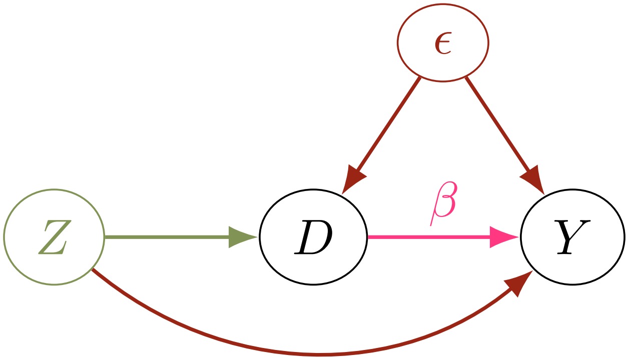
A Valid Instrument



A Violation of As-Good-As-Random Assignment



A Violation of Exclusion



Where do IVs Come From?

1. True Lotteries

One sure-fire way to ensure that a Z_i is as-good-as-randomly assigned is...

Where do IVs Come From?

1. True Lotteries

One sure-fire way to ensure that a Z_i is as-good-as-randomly assigned is... to randomly assign it!

Where do IVs Come From?

1. True Lotteries

One sure-fire way to ensure that a Z_i is as-good-as-randomly assigned is... to randomly assign it!

- Some of the best IVs come from lotteries, either run by the researcher (e.g. an RCT) or so-called "natural experiments"
- We still need to worry about violations of the exclusion restriction
- Relevance holds when Z_i has some effect on X_i

Where do IVs Come From?

1. True Lotteries

One sure-fire way to ensure that a Z_i is as-good-as-randomly assigned is... to randomly assign it!

- Some of the best IVs come from lotteries, either run by the researcher (e.g. an RCT) or so-called "natural experiments"
- We still need to worry about violations of the exclusion restriction
- Relevance holds when Z_i has some effect on X_i

"Gold standard" IV: a randomized offer to participate in a program, with X_i recording program participation

- Exclusion restriction likely to hold for any Y_i, by construction
- Relevance almost guaranteed (provided people want the program!)

Example

Charter School Lotteries

Abdulkadiroglu et al. (2016) are interested in whether going to a “charter” middle school increases standardized test scores

- Charter students tend to score better, but we worry about selection
- History of doubting educational inputs, since Coleman (1966)

Example

Charter School Lotteries

We leverage an institutional feature of charters: *admission lotteries*

- When more kids want to enroll than there are seats, admission offers $Z_i \in \{0, 1\}$ are effectively drawn from a hat
- Offers plausibly only affect later test scores Y_i by changing charter enrollment $D_i \in \{0, 1\}$, so are plausibly valid instruments
- We need to control for lottery fixed effects ("risk sets") to make Z_i as-good-as-randomly assigned – more on this soon

Example

Charter School Lotteries

We leverage an institutional feature of charters: *admission lotteries*

- When more kids want to enroll than there are seats, admission offers $Z_i \in \{0, 1\}$ are effectively drawn from a hat
- Offers plausibly only affect later test scores Y_i by changing charter enrollment $D_i \in \{0, 1\}$, so are plausibly valid instruments
- We need to control for lottery fixed effects ("risk sets") to make Z_i as-good-as-randomly assigned – more on this soon

We study a particular charter (UP Academy), which is "takeover"

- Two offer IVs: "immediate" (on lottery night) and from a waitlist

Lottery IV Estimates of UP Test Score Effects

TABLE 8—LOTTERY IV ESTIMATES OF UP EFFECTS

				2SLS		
				First stage		
		Comparison group mean (1)	OLS (2)	Immediate offer (3)	Waitlist offer (4)	Enrollment effect (5)
Panel A. All grades (Sixth through eighth)	Math (N = 2,202)	0.059	0.301	0.760	0.562	0.270
			(0.022)	(0.063)	(0.067)	(0.056)
	ELA (N = 2,205)	0.103	0.148	0.759	0.562	0.118
			(0.020)	(0.063)	(0.067)	(0.051)

Where do IVs Come From?

2. Natural Experiments

Without appealing to literal randomization, we may credibly argue Z_i is as-good-as-randomly assigned conditional on some $\mathbf{W}_i$

- Such "natural experiments" rely on a selection-on-observables argument (for Z_i, instead D_i)
- Still worry about exclusion: Z_i cannot affect Y_i except through D_i

Where do IVs Come From?

2. Natural Experiments

Without appealing to literal randomization, we may credibly argue Z_i is as-good-as-randomly assigned conditional on some $\mathbf{W}_i$

- Such "natural experiments" rely on a selection-on-observables argument (for Z_i, instead D_i)
- Still worry about exclusion: Z_i cannot affect Y_i except through D_i

Example

Quarter-of-birth

Angrist and Krueger (1991) famously estimate labor market returns to schooling with a creative IV: student quarter-of-birth

- Compulsory schooling requirements prevent students from dropping before the day they turn 16 (used to be more binding)
- Fixed school start dates mean students who drop out at 16 get more or less schooling depending on their birth date

Example

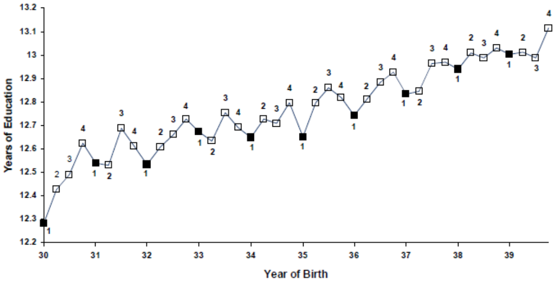
Quarter-of-birth

Angrist and Krueger (1991) famously estimate labor market returns to schooling with a creative IV: student quarter-of-birth

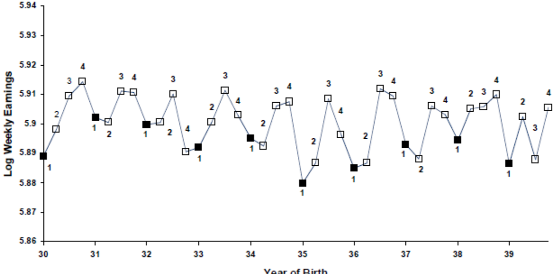
- Compulsory schooling requirements prevent students from dropping before the day they turn 16 (used to be more binding)
- Fixed school start dates mean students who drop out at 16 get more or less schooling depending on their birth date
- Quarter-of-birth seems quasi-randomly assigned — is it excludable? See Buckles and Hungerman (2013)...

The Quarter-of-Birth Natural Experiment: Visualized

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



B. Average Weekly Wage by Quarter of Birth (reduced form)



Quarter-of-Birth IV Estimates of Returns to Schooling

Table 4.1.1: 2SLS estimates of the economic returns to schooling

	OLS		2SLS			
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Years of education	0.075 (0.0004)	0.072 (0.0004)	0.103 (0.024)	0.112 (0.021)	0.106 (0.026)	0.108 (0.019)
Covariates:						
9 year of birth dummies		✓			✓	✓
50 state of birth dummies		✓			✓	✓
Instruments:			dummy for QOB=1	dummy for QOB=1 or QOB=2	dummy for QOB=1	full set of QOB dummies

Where do IVs Come From?

3. Panel Data

We might also combine IV + difference-in-differences identification

- E.g. instrument with $Z_i \times Post_t$, controlling for Z_i and $Post_t$ FEs
- This requires two parallel trends assumptions, for the RF and FS
- Still need to worry about the exclusion restriction, as always

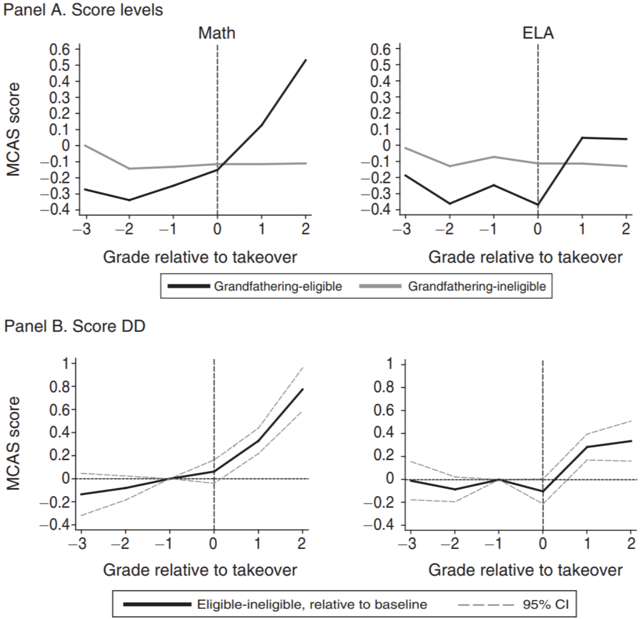
Example

School Lottery

Abdulkadiroglu et al. (2016) complement their lottery analysis of takeover charters with an instrumented diff-in-diff analysis

- Students enrolled in the "legacy" public school were eligible for being "grandfathered" into UP, without having to apply to the charter
- We compare their trends in test scores & enrollment to a matched comparison group of observably-similar students at other schools

Grandfathering IV: Visualized



Grandfathering IV Estimates of UP Test Score Effects

Table 7—Grandfathering IV Estimates of UP Effects

		Comparison group mean (1)	OLS (2)	2SLS First stage (3)	2SLS Enrollment effect (4)
Panel A. All grades (Seventh through eighth)	Math (N = 1,543)	−0.233	0.400	1.051	0.321
			(0.032)	(0.040)	(0.039)
	ELA (N = 1,539)	−0.214	0.296	1.040	0.394
			(0.035)	(0.041)	(0.044)

Roadmap

Where do (Good) Instruments Come From?

- True Lotteries
- Natural Experiments
- Panel Data

2SLS Mechanics

- Just-Identified IV
- Overidentification

Weak and Many Instruments

- Weak IV
- Many IVs

Just-Identified IV

As you likely know, the general `ivregress` command (or its equivalent `fixes::feols` in R) allows for controls and multiple treatments / instruments

- When # treatment = # instruments, we say the IV is "just-identified":

Just-Identified IV

As you likely know, the general `ivregress` command (or its equivalent `fixes::feols` in R) allows for controls and multiple treatments / instruments

- When # treatment = # instruments, we say the IV is "just-identified":

$$Y_i = \beta D_i + \mathbf{W}_i'\boldsymbol{\gamma} + \varepsilon_i \quad \text{(second stage)}$$
$$X_i = \pi Z_i + \mathbf{W}_i'\boldsymbol{\mu} + \eta_i \quad \text{(first stage)}$$

where $\mathbf{W}_i$ includes a constant.

Just-Identified IV

The reduced form is:

$$Y_i = \rho Z_i + \mathbf{W}_i' \boldsymbol{\kappa} + \nu_i$$

Just-Identified IV

The reduced form is:

$$Y_i = \rho Z_i + \mathbf{W}_i' \boldsymbol{\kappa} + \nu_i$$

Same identification logic as before:

- Validity: $Cov(Z_i, \varepsilon_i) = 0$, allowing $Cov(Z_i, \mathbf{W}_i) \neq 0$

Just-Identified IV

The reduced form is:

$$Y_i = \rho Z_i + \mathbf{W}_i' \boldsymbol{\kappa} + \nu_i$$

Same identification logic as before:

- Validity: $Cov(Z_i, \varepsilon_i) = 0$, allowing $Cov(Z_i, \mathbf{W}_i) \neq 0$
- Relevance: $\pi \neq 0$, so Z_i and D_i are correlated controlling for $\mathbf{W}_i$

Just-Identified IV

The reduced form is:

$$Y_i = \rho Z_i + \mathbf{W}_i' \boldsymbol{\kappa} + \nu_i$$

Same identification logic as before:

- Validity: $Cov(Z_i, \varepsilon_i) = 0$, allowing $Cov(Z_i, \mathbf{W}_i) \neq 0$
- Relevance: $\pi \neq 0$, so Z_i and D_i are correlated controlling for $\mathbf{W}_i$

IV is still "reduced form over first stage": $(\beta = \rho/\pi)$

Just-Identified IV

The reduced form is:

$$Y_i = \rho Z_i + \mathbf{W}_i' \boldsymbol{\kappa} + \nu_i$$

Same identification logic as before:

- Validity: $Cov(Z_i, \varepsilon_i) = 0$, allowing $Cov(Z_i, \mathbf{W}_i) \neq 0$
- Relevance: $\pi \neq 0$, so Z_i and D_i are correlated controlling for $\mathbf{W}_i$

IV is still “reduced form over first stage”: ($\beta = \rho / \pi$)

- Can use Frisch-Waugh-Lovell to “partial out” $\mathbf{W}_i$ from Y_i, X_i, D_i, and so get back to an IV regression without controls

Overidentification

Sometimes we have more than one instrument $Z_{i\ell}$, for $\ell = 1, \ldots, L$.

This leads to an "overidentified" IV regression:

$$Y_i = \beta D_i + \mathbf{W}_i' \boldsymbol{\gamma} + \varepsilon_i \text{ (second stage)}$$

$$X_i = \mathbf{Z}_i' \boldsymbol{\pi} + \mathbf{W}_i' \boldsymbol{\mu} + \eta_i \text{ (first stage)}$$

where $\mathbf{Z}_i = [Z_{i1}, \ldots, Z_{iL}]'$. Reduced form: $Y_i = \mathbf{Z}_i' \boldsymbol{\rho} + \mathbf{W}_i' \boldsymbol{\kappa} + \nu_i$

Overidentification

Sometimes we have more than one instrument $Z_{i\ell}$, for $\ell = 1, \dots, L$.
This leads to an "overidentified" IV regression:

$$Y_i = \beta D_i + \mathbf{W}_i'\boldsymbol{\gamma} + \varepsilon_i \text{ (second stage)}$$

$$X_i = \mathbf{Z}_i'\boldsymbol{\pi} + \mathbf{W}_i'\boldsymbol{\mu} + \eta_i \text{ (first stage)}$$

where $\mathbf{Z}_i = [Z_{i1}, \dots, Z_{iL}]'$. Reduced form: $Y_i = \mathbf{Z}_i'\boldsymbol{\rho} + \mathbf{W}_i'\boldsymbol{\kappa} + \nu_i$
Validity: $Cov(Z_{i\ell}, \varepsilon_i) = 0$ for all ℓ

- "Overidentified" because we could use any $Z_{i\ell}$ to estimate $\beta = \boldsymbol{\rho}_\ell / \boldsymbol{\pi}_\ell$

Overidentification

Sometimes we have more than one instrument $Z_{i\ell}$, for $\ell = 1, \ldots, L$.
This leads to an "overidentified" IV regression:

$$Y_i = \beta D_i + \mathbf{W}_i' \boldsymbol{\gamma} + \varepsilon_i \text{ (second stage)}$$

$$X_i = \mathbf{Z}_i' \boldsymbol{\pi} + \mathbf{W}_i' \boldsymbol{\mu} + \eta_i \text{ (first stage)}$$

where $\mathbf{Z}_i = [Z_{i1}, \ldots, Z_{iL}]'$. Reduced form: $Y_i = \mathbf{Z}_i' \boldsymbol{\rho} + \mathbf{W}_i' \boldsymbol{\kappa} + \nu_i$
Validity: $Cov(Z_{i\ell}, \varepsilon_i) = 0$ for all ℓ

- "Overidentified" because we could use any $Z_{i\ell}$ to estimate $\beta = \boldsymbol{\rho}_\ell / \boldsymbol{\pi}_\ell$
- Relevance: $\boldsymbol{\pi}_\ell \neq 0$ for at least some ℓ

Overidentification

Sometimes we have more than one instrument $Z_{i\ell}$, for $\ell = 1, \dots, L$. This leads to an "overidentified" IV regression:

$$Y_i = \beta D_i + \mathbf{W}_i'\boldsymbol{\gamma} + \varepsilon_i \text{ (second stage)}$$

$$X_i = \mathbf{Z}_i'\boldsymbol{\pi} + \mathbf{W}_i'\boldsymbol{\mu} + \eta_i \text{ (first stage)}$$

where $\mathbf{Z}_i = [Z_{i1}, \dots, Z_{iL}]'$. Reduced form: $Y_i = \mathbf{Z}_i'\boldsymbol{\rho} + \mathbf{W}_i'\boldsymbol{\kappa} + \nu_i$

Validity: $Cov(Z_{i\ell}, \varepsilon_i) = 0$ for all ℓ

- "Overidentified" because we could use any $Z_{i\ell}$ to estimate $\beta = \boldsymbol{\rho}_\ell / \boldsymbol{\pi}_\ell$
- Relevance: $\boldsymbol{\pi}_\ell \neq 0$ for at least some ℓ

Overidentification can yield tests of IV validity

- Intuitively, 2SLS checks whether all the $Z_{i\ell}$ yields the same IV estimate, which is sensible in a constant-effects model...

Putting the “2S” in “2SLS”

You’ll notice I haven’t actually defined 2SLS beyond the simple case

- Before we had $\beta^{IV} = \frac{Cov(Z_i, Y_i)}{Cov(Z_i, D_i)}$ leading to $\widehat{\beta}^{IV} = \frac{\widehat{Cov}(Z_i, Y_i)}{\widehat{Cov}(Z_i, D_i)}$
- General form follows similarly (sample analog), but is notation-heavy

Putting the “2S” in “2SLS”

You’ll notice I haven’t actually defined 2SLS beyond the simple case

- Before we had $\beta^{IV} = \frac{Cov(Z_i,Y_i)}{Cov(Z_i,D_i)}$ leading to $\widehat{\beta}^{IV} = \frac{\widehat{Cov}(Z_i,Y_i)}{\widehat{Cov}(Z_i,D_i)}$
- General form follows similarly (sample analog), but is notation-heavy

A more useful way to define 2SLS is by a two-step procedure:

- First regress D_i on all $Z_{i\ell}$ and W_{ik}
- Then regress Y_i on the “fitted values” $\widehat{D}_i$ and controls W_{ik}

Putting the "2S" in "2SLS"

You'll notice I haven't actually defined 2SLS beyond the simple case

- Before we had $\beta^{IV} = \frac{Cov(Z_i,Y_i)}{Cov(Z_i,D_i)}$ leading to $\widehat{\beta}^{IV} = \frac{\widehat{Cov}(Z_i,Y_i)}{\widehat{Cov}(Z_i,D_i)}$
- General form follows similarly (sample analog), but is notation-heavy

A more useful way to define 2SLS is by a two-step procedure:

- First regress D_i on all $Z_{i\ell}$ and W_{ik}
- Then regress Y_i on the "fitted values" $\widehat{D}_i$ and controls W_{ik}

The proof of this follows from some (simple) linear algebra

- Intuitively, regressing Y_i on $\widehat{\pi}^{OLS} Z_i$ gives a scaled RF:

$$\widehat{\beta}^{IV} = \frac{\widehat{\rho}^{OLS}}{\widehat{\pi}^{OLS}}$$

Putting the “2S” in “2SLS”

$$\widehat{\beta}^{IV} = \frac{\widehat{\rho}^{OLS}}{\widehat{\pi}^{OLS}}$$

Although easy, you should never do such “manual 2SLS” yourself!

- Your point estimates will be right, but your SEs won’t be!

Roadmap

Where do (Good) Instruments Come From?

- True Lotteries
- Natural Experiments
- Panel Data

2SLS Mechanics

- Just-Identified IV
- Overidentification

Weak and Many Instruments

- Weak IV
- Many IVs

Weak Instruments

When running just-identified IV, you should always worry about the "strength" of your instrument

- Specifically the first stage F-statistic, which tests $\pi^{OLS} = 0$

Weak Instruments

When running just-identified IV, you should always worry about the "strength" of your instrument

- Specifically the first stage F-statistic, which tests $\pi^{OLS} = 0$

If π^{OLS} is small relative to its standard error, the IV is "weak"

- Typically use the rule-of-thumb of $F < 10$ (Staiger and Stock 1997)
- In this case the second-stage SEs will be large and the 2SLS estimate will tend to be biased towards the corresponding OLS estimate

Weak Instruments

When running just-identified IV, you should always worry about the "strength" of your instrument

- Specifically the first stage F-statistic, which tests $\pi^{OLS} = 0$

If π^{OLS} is small relative to its standard error, the IV is "weak"

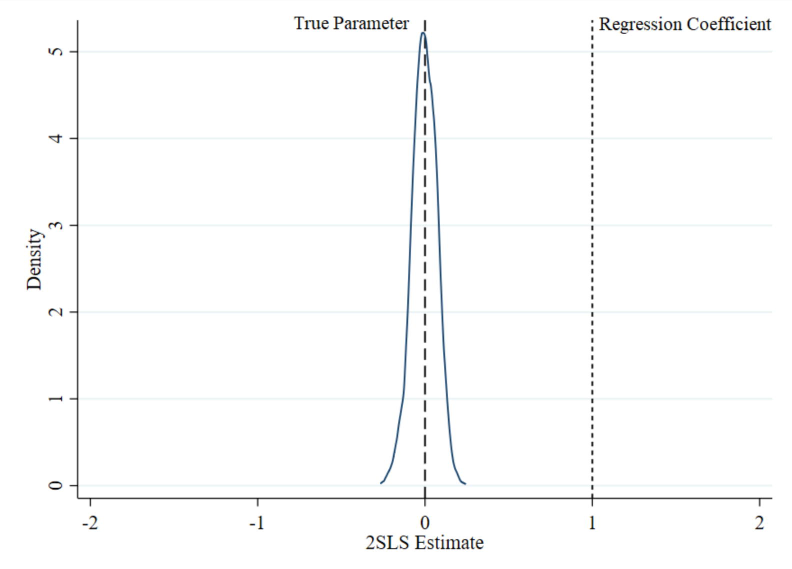
- Typically use the rule-of-thumb of $F < 10$ (Staiger and Stock 1997)
- In this case the second-stage SEs will be large and the 2SLS estimate will tend to be biased towards the corresponding OLS estimate

Much made of this over the years, but Angrist and Kolesár (2022) argue recently that we shouldn't worry too much

- The SE increase tends to be large enough to "cover up" the bias
- Just-id. 2SLS is "approximately median-unbiased" (as it is LIML)

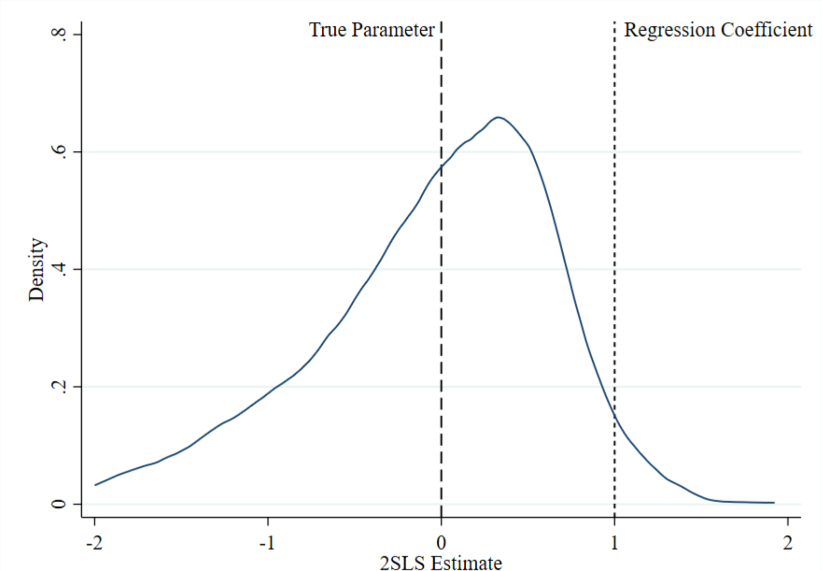
Weak Instruments: Visualized

Monte Carlo: $Y_i = \varepsilon_i$, $D_i = \Pi Z_i + \eta_i$: $\Pi = Var(\varepsilon_i) = Var(\eta_i) = 1$



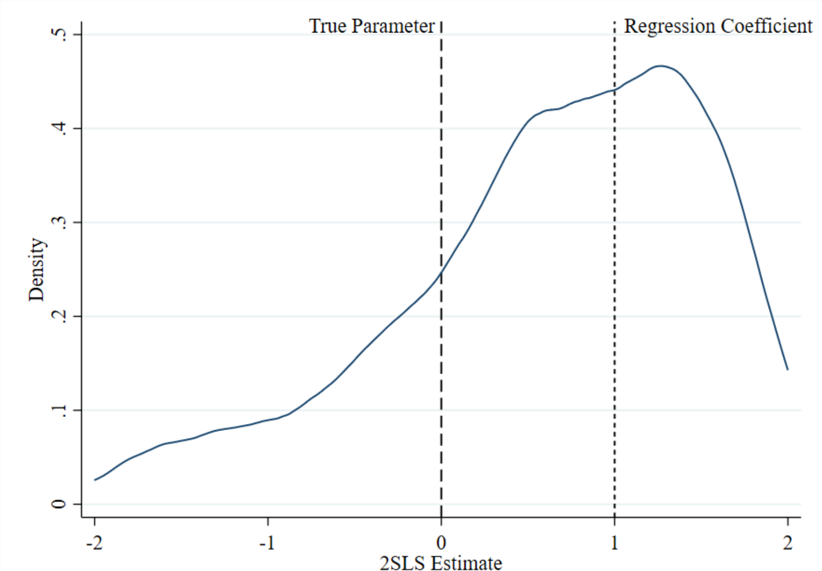
Weak Instruments: Visualized

Monte Carlo: $Y_i = \varepsilon_i$, $D_i = \Pi Z_i + \eta_i$: $\Pi = 0.1$ (Weaker)



Weak Instruments: Visualized

Monte Carlo: $Y_i = \varepsilon_i$, $D_i = \Pi Z_i + \eta_i$: $\Pi = 0.01$ (Very Weak)



Many IVs

A more pernicious problem is many-instrument bias, when overid

- Also tends to manifest in low first-stage F's, so also good to check

Many IVs

A more pernicious problem is many-instrument bias, when overid

- Also tends to manifest in low first-stage F's, so also good to check

Many-IV bias is also towards OLS, but unlike before SEs go *down*

- Intuitively, a more flexible FS tends to fit D_i better $\rightarrow$ more power
- But we can have *overfitting* with lots of Z_i $\rightarrow$ essentially recreate D_i

Many IVs

A more pernicious problem is many-instrument bias, when overid

- Also tends to manifest in low first-stage F's, so also good to check

Many-IV bias is also towards OLS, but unlike before SEs go *down*

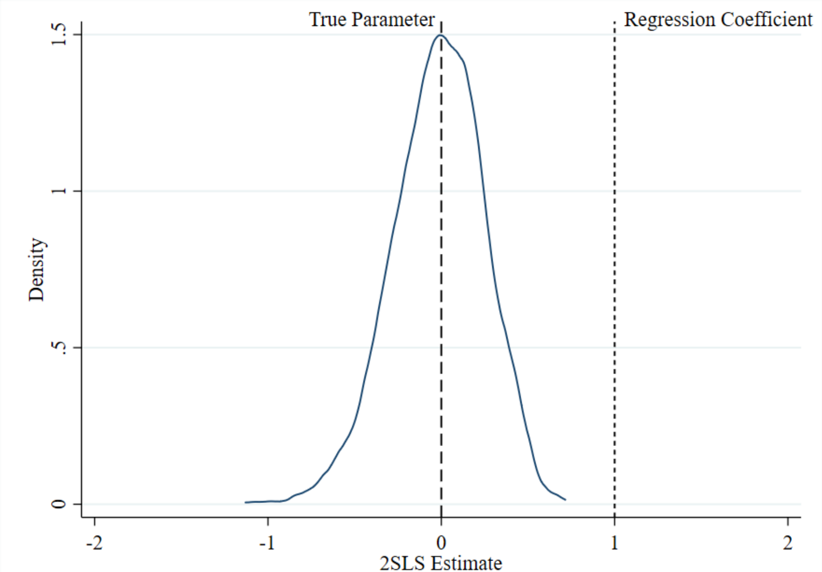
- Intuitively, a more flexible FS tends to fit D_i better $\rightarrow$ more power
- But we can have *overfitting* with lots of Z_i $\rightarrow$ essentially recreate D_i

As we'll see, this bias is especially relevant in judge IV designs

- Potentially many judge assignment indicators as the instrument
- Leave-out corrections (e.g. Angrist et al. 1999) have been adapted to this setting in recent years (e.g. Kolesár 2013)

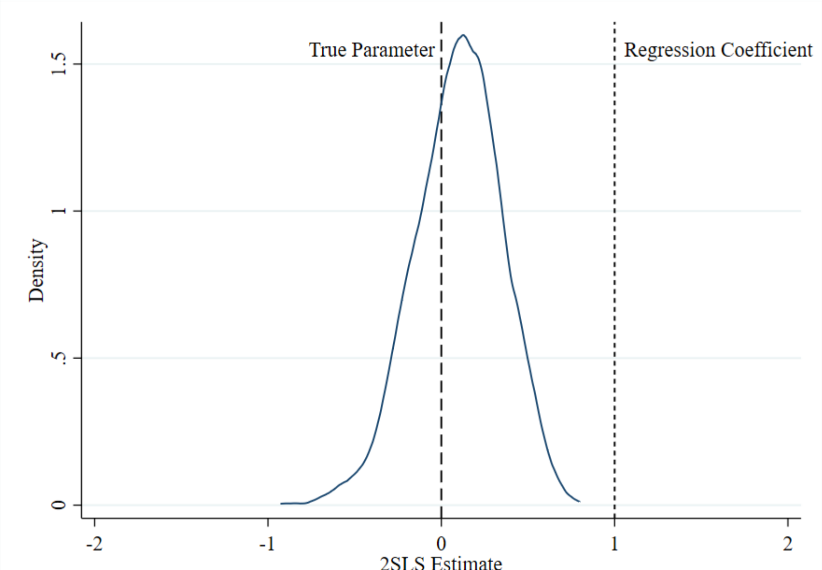
Weak and Many Instruments VIII

Monte Carlo: $Y_i = \varepsilon_i$, $X_i = \Pi Z_{i1} + \eta_i$: IV with one Z_{i1}



Weak and Many Instruments IX

Monte Carlo: $Y_i = \varepsilon_i$, $X_i = \Pi Z_{i1} + \eta_i$: IV with ten Z_{ij}



Weak and Many Instruments X

Monte Carlo: $Y_i = \varepsilon_i$, $X_i = \Pi Z_{i1} + \eta_i$: IV with 100 Z_{ij}