Chapter 6: IV Mechanics

Peter Hull

Applied Econometrics II
Brown University
Spring 2024

Outline

1. Just-Identified IV
2. Overidentification
3. Weak- and Many-Instrument Bias
4. Application: Abdulkadiroglu et al. (2016)

Motivation: Constant Effects

Our IV story starts with a causal model: $Y_i = \beta X_i + \varepsilon_i$, w/ $Cov(X_i, \varepsilon_i) \neq 0$

- E.g. $X_i \in \{0,1\}$ indicates enrollment in this class, Y_i measures later job market success / innate happiness
- "Endogeneity": students who take this class have systematically different untreated potential outcomes ε_i (i.e. selection bias)

Imagine the course was "oversubscribed," so I chose the roster by lottery

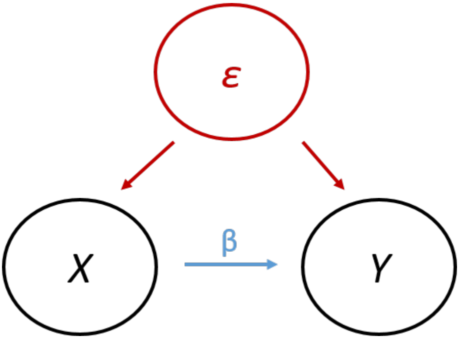
- $Z_i \in \{0,1\}$ indicates random admission to the course
- Randomness + no direct effects of Z_i on Y_i implies $Cov(Z_i, \varepsilon_i) = 0$

Plugging in the model for $\varepsilon_i = Y_i - \beta X_i$, we have IV identification:

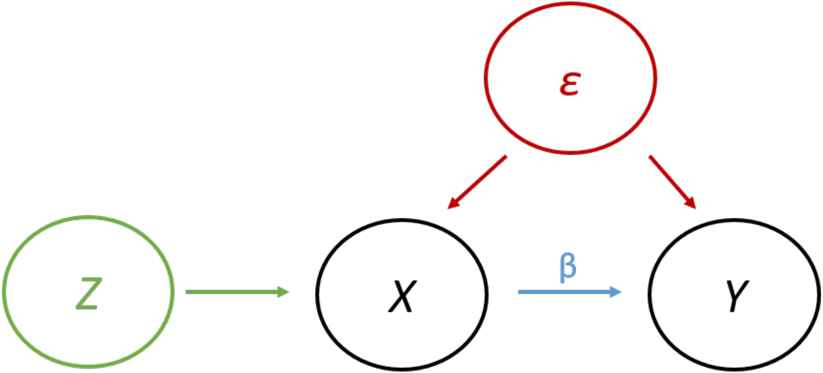
$$Cov(Z_i, Y_i - \beta X_i) = 0 \implies \frac{Cov(Z_i, Y_i)}{Cov(Z_i, X_i)} = \beta$$

so long as $Cov(Z_i, X_i) \neq 0$ ("relevance")

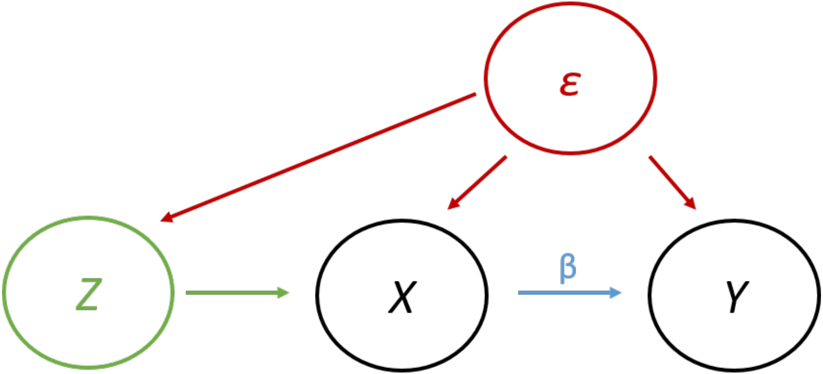
Regression "Endogeneity"



Instrument “Exogeneity” / “Validity”

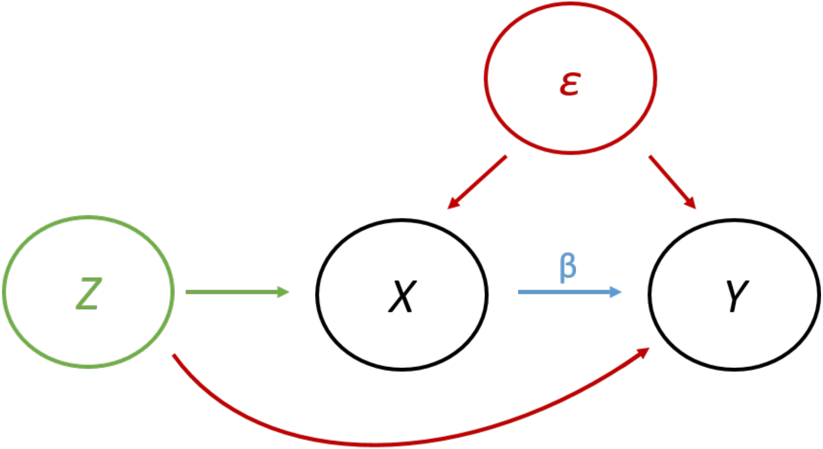


Threats to Validity: Instrument Assignment



We will later formalize this as a failure of instrument "independence"

Threats to Validity: Direct Effects



We will later formalize this as a failure of instrument "exclusion"

Reduced Form and First Stage

To link the IV estimand to regression, divide top and bottom by $Var(Z_i)$:

$$\beta^{IV} \equiv \frac{Cov(Z_i, Y_i)}{Cov(Z_i, X_i)} = \frac{Cov(Z_i, Y_i)/Var(Z_i)}{Cov(Z_i, X_i)/Var(Z_i)} = \frac{\rho}{\pi}$$

where we define:

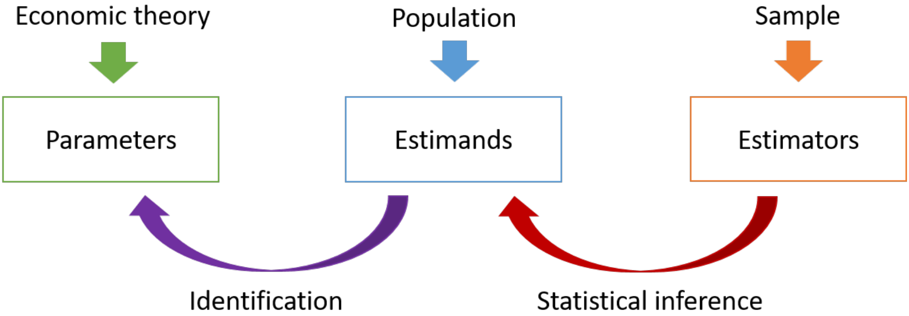
$$Y_i = \kappa + \rho Z_i + \nu_i \text{ , the "reduced form" regression}$$
$$X_i = \mu + \pi Z_i + \eta_i \text{ , the "first stage" regression}$$

If Z_i is binary, this further reduces to the Wald estimand:

$$\frac{\rho}{\pi} = \frac{E[Y_i \mid Z_i = 1] - E[Y_i \mid Z_i = 0]}{E[X_i \mid Z_i = 1] - E[X_i \mid Z_i = 0]}$$

Second stage: $Y_i = \alpha + \beta^{IV} X_i + U_i$ where $Cov(Z_i, U_i) = 0$ by *definition*

Remember! Estimand $\beta^{IV} \neq$ Parameter β



Adding Controls

We can generalize our IV estimand by adding controls to RF + FS:

$$Y_i = \rho Z_i + W_i' \kappa + v_i$$
$$X_i = \pi Z_i + W_i' \mu + \eta_i$$

By the FWL, $\beta^{IV} = \rho / \pi = Cov(\tilde{Z}_i, Y_i) / Cov(\tilde{Z}_i, X_i)$ for residuals $\tilde{Z}_i$

- 2nd stage: $Y_i = \beta^{IV} X_i + W_i' \gamma + U_i$ w/ $Cov(Z_i, U_i) = Cov(W_i, U_i) = 0$

Mechanically, then, IV is just a fancy way to divide two regression coefs!

- W_i may be such that ρ and π are both interpretable as causal effects (e.g. some selection-on-observables or parallel trends arguments hold)
- ρ / π may then be of interest/causal – we'll do this more formally soon

In the sample, of course, things follow analogously: i.e. $\hat{\beta}^{IV} = \frac{\widehat{Cov}(\tilde{Z}_i, Y_i)}{\widehat{Cov}(\tilde{Z}_i, X_i)}$

Illustration: The Oregon Health Insurance Experiment

The 2008 OHIE randomized Medicaid offers to interested individuals with earnings between 100% and 138% of the FPL

- Within lotteries $\ell(i)$, offers Z_i are as-good-as-random
- Plausible that Z_i only affects outcomes thru Medicaid takeup D_i

Finkelstein et al. (2012) estimate:

$$Y_i = \beta D_i + \gamma_{\ell(i)} + W_i' \phi + \varepsilon_i$$
$$D_i = \pi Z_i + \mu_{\ell(i)} + W_i' \psi + \eta_i$$

where W_i are baseline (pre-randomization) characteristics

Pop Quiz 1: what's the (precise) interpretation of the first stage π?

OHIE Reduced Form and Second Stage

	Control mean (1)	ITT (2)	LATE (3)	p-values (4)
Panel A: Extensive margin				
All hospital admissions	0.067	0.0054	0.021	[0.004]
	(0.250)	(0.0019)	(0.0074)	
Admissions through ER	0.048	0.0018	0.0070	[0.265]
	(0.214)	(0.0016)	(0.0062)	
Admissions not through ER	0.029	0.0041	0.016	[0.002]
	(0.167)	(0.0013)	(0.0051)	
Panel B: All hospital admissions				
Days	0.498	0.026	0.101	[0.329]
	(3.795)	(0.027)	(0.104)	{0.328}
List charges	2,613	258	1,009	[0.077]
	(19,942)	(146)	(569)	{0.106}
Procedures	0.155	0.018	0.070	[0.031]
	(1.08)	(0.0083)	(0.032)	{0.059}
Standardized treatment effect		0.012	0.047	[0.073]
		(0.0067)	(0.026)	

Pop Quiz 2: what's the (approximate) value of the first stage π?

Multiple Treatments

Now start with a general causal model: $Y_i = X_i'\beta + \varepsilon_i$ for $dim(X_i) = J \geq 1$

- Instrument vector Z_i, with $dim(Z_i) = L \geq J$
- Population residuals $\tilde{Z}_i$, given a vector of controls W_i

Suppose the residualized instrument is valid: $Cov(\tilde{Z}_i, \varepsilon_i) = 0$. Then

$$Cov(\tilde{Z}_i, Y_i - X_i'\beta) = 0$$
$$\implies Cov(\tilde{Z}_i, Y_i) = Cov(\tilde{Z}_i, X_i)\beta$$

If $L = J$, then $Cov(\tilde{Z}_i, X_i)$ is square. If its furthermore invertible:

$$\begin{aligned}\beta &= Cov(\tilde{Z}_i, X_i)^{-1} Cov(\tilde{Z}_i, Y_i) \\ &= (\underbrace{Var(\tilde{Z}_i)^{-1} Cov(\tilde{Z}_i, X_i)}_{\text{First Stage}})^{-1} \underbrace{Var(\tilde{Z}_i)^{-1} Cov(\tilde{Z}_i, Y_i)}_{\text{Reduced Form}} \equiv \beta^{IV}\end{aligned}$$

Second stage: $Y_i = X_i'\beta^{IV} + W_i'\gamma + U_i$ with $Cov(Z_i, U_i) = Cov(W_i, U_i) = 0$

Outline

1. Just-Identified IV ✓
2. Overidentification
3. Weak- and Many-Instrument Bias
4. Application: Abdulkadiroglu et al. (2016)

Sometimes, You Get More Than You Need

With more valid instruments than treatments, β is *overidentified*:

$$\underbrace{Cov(\tilde{Z}_i, Y_i)}_{L\times 1} = \underbrace{Cov(\tilde{Z}_i, X_i)}_{L\times J} \underbrace{\beta}_{J\times 1}$$

We can drop any $L - J$ instruments to recover β from estimable moments

More generally, we can take any full-rank linear combination $\tilde{Z}_i^* = M\tilde{Z}_i$ for $J \times L$ matrix M such that $Cov(\tilde{Z}_i^*, X_i)$ is invertible:

$$\underbrace{M \cdot Cov(\tilde{Z}_i, Y_i)}_{Cov(\tilde{Z}_i^*, Y_i)} = \underbrace{M \cdot Cov(\tilde{Z}_i, X_i)}_{Cov(\tilde{Z}_i^*, X_i)} \beta$$

$$\text{so } \beta = \left(M \cdot Cov(\tilde{Z}_i, X_i)\right)^{-1} M \cdot Cov(\tilde{Z}_i, Y_i)$$

This defines a class of IV estimands/estimators, parameterized by M

- If $L = J$ (the *just-identified* case), the choice of M is irrelevant

Two-Stage Least Squares

A natural choice for M is the matrix of first-stage coefficients:

$$\beta^{2SLS} = \Bigg(\underbrace{Cov(\tilde{Z}_i, X_i)' Var(\tilde{Z}_i)^{-1}}_{M} Cov(\tilde{Z}_i, X_i)\Bigg)^{-1} \underbrace{Cov(\tilde{Z}_i, X_i)' Var(\tilde{Z}_i)^{-1}}_{M} Cov(\tilde{Z}_i, Y_i)$$

This makes the augmented instrument $\tilde{Z}_i^* = Cov(\tilde{Z}_i, X_i)' Var(\tilde{Z}_i)^{-1} \tilde{Z}_i$ the (residualized) first-stage fitted values

The 2SLS estimator takes this to the sample:

$$\begin{aligned}\hat{\beta}^{2SLS} &= \left(\widehat{Cov}(\tilde{Z}_i, X_i)' \widehat{Var}(\tilde{Z}_i)^{-1} \widehat{Cov}(\tilde{Z}_i, X_i)\right)^{-1} \widehat{Cov}(\tilde{Z}_i, X_i)' \widehat{Var}(\tilde{Z}_i)^{-1} \widehat{Cov}(\tilde{Z}_i, Y_i) \\ &= (X' P_{\tilde{Z}} X)^{-1} X' P_{\tilde{Z}} Y\end{aligned}$$

where X and Y stack observations of X_i' and Y_i, and where $P_{\tilde{Z}}$ is the IV *projection matrix*: the fitted values from OLS on (sample) residualized $\tilde{Z}_i$

- Specifically, $P_{\tilde{Z}} = \hat{\tilde{Z}}(\hat{\tilde{Z}}'\hat{\tilde{Z}})^{-1}\hat{\tilde{Z}}'$ where $\hat{\tilde{Z}}$ stacks OLS residuals

Putting the "Two Stage" in 2SLS

Since $P_{\tilde{Z}}$ is an *idempotent* and *symmetric* matrix, we can rewrite

$$\begin{aligned}\hat{\beta}^{2SLS} &= (X'P_{\tilde{Z}}X)^{-1}X'P_{\tilde{Z}}Y \\ &= (X'P_{\tilde{Z}}P_{\tilde{Z}}X)^{-1}X'P_{\tilde{Z}}Y \\ &= (X'P'_{\tilde{Z}}P_{\tilde{Z}}X)^{-1}X'P'_{\tilde{Z}}Y \\ &= ((P_{\tilde{Z}}X)'P_{\tilde{Z}}X)^{-1}(P_{\tilde{Z}}X)'Y\end{aligned}$$

This is the formula for an OLS regression of Y_i on $\hat{X}_i$, the first-stage fitted values (partialling out controls). Hence *two-stage least squares!*

You may encounter other estimators derived from different M (e.g. LIML). But because we like regression, we also tend to like 2SLS ...

Avoid Manual 2SLS!

Although easy, you should never literally run 2SLS in two stages

- Your point estimates will be right, but your SEs generally won't be

Such "manual 2SLS" also opens you up to making a variety of mistakes:

1. Omitting some controls in the first or second stage
2. "Forbidden regressions": e.g. regressing Y_i on probit fits for X_i
3. Regressing on $\hat{X}_i$ and $\hat{X}_i^2$, instead of instrumenting X_i^2 directly
4. See MHE Section 4.6.1 for more

Just don't do it! Always run 2SLS in one step with (e.g.) *ivreg2 , r*

2SLS in Equation Form

In practice, as with regressions, we tend to specify 2SLS equations. E.g.:

$$
\begin{aligned}
Y_i &= \beta_1 X_{i1} + \beta_2 X_{i2} + W_i' \gamma + U_i \\
X_{i1} &= \pi_{11} Z_{i1} + \pi_{12} Z_{i2} + \pi_{13} Z_{i3} + W_i' \mu_1 + \eta_{i1} \\
X_{i2} &= \pi_{21} Z_{i1} + \pi_{22} Z_{i2} + \pi_{23} Z_{i3} + W_i' \mu_2 + \eta_{i12}
\end{aligned}
$$

... since this is a whole lot easier than writing out the matrix formulas!

- Understand: absent further assumptions this specifies some *estimands*
- Note: all the instruments enter all the first stage regressions (another mistake sometimes made with "manual 2SLS")

Illustration: Angrist and Krueger (1991)

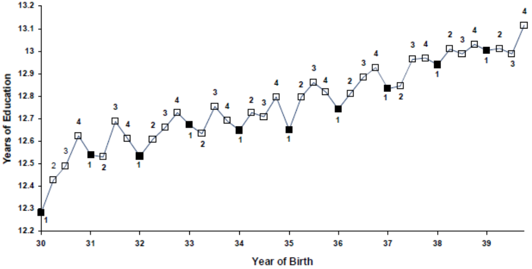
AK '91 famously estimate labor market returns to schooling with a creative instrument: student quarter-of-birth

- Compulsory schooling requirements prevent students from dropping out 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 dates

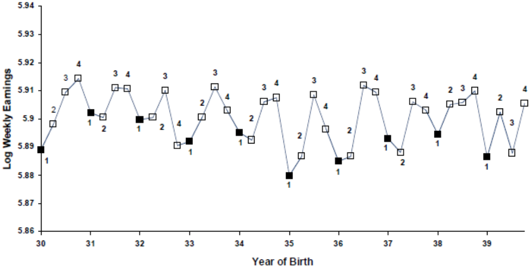
Much debate over whether QOB is a valid instrument for schooling (e.g. Buckles and Hungerman 2013), but the empirics are at least compelling...

AK '91 Reduced Form and First Stage

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



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



The Power of Overidentification

	OLS		2SLS				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Years of education	.071	.067	.102	.13	.104	.108	.087
	(.0004)	(.0004)	(.024)	(.020)	(.026)	(.020)	(.016)
Covariates							
9 year-of-birth dummies		✓			✓	✓	✓
50 state-of-birth dummies		✓			✓	✓	✓
Instruments							
dummy for QOB = 1			✓	✓	✓	✓	✓
dummy for QOB = 2				✓		✓	✓
dummy for QOB = 3				✓		✓	✓
QOB dummies interacted with year-of-birth dummies (30 instruments total)							✓

Notes: The table reports OLS and 2SLS estimates of the returns to schooling using the Angrist and Krueger (1991) 1980 census sample. This sample includes native-born men, born 1930–39, with positive earnings and nonallocated values for key variables. The sample size is 329,509. Robust standard errors are reported in parentheses. QOB denotes quarter of birth.

2SLS is a Many-Splendored Thing

"OLS on first-stage fits" is one way to understand 2SLS. Another is as a weighted average of just-identified IV estimates:

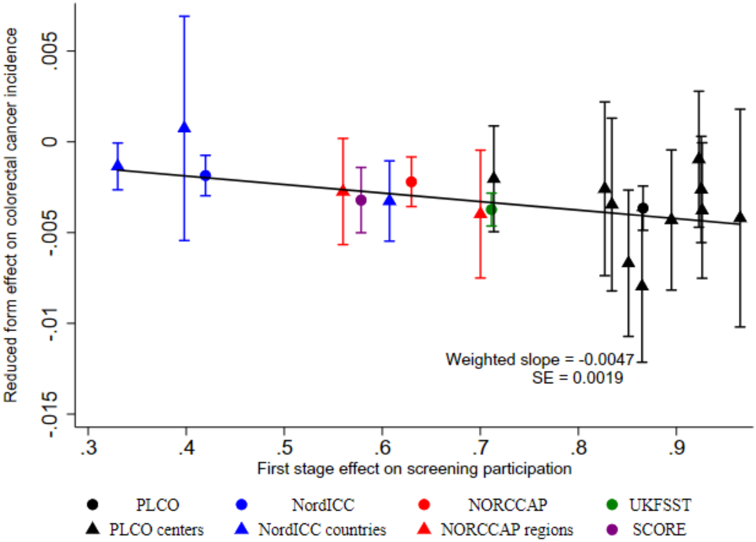
$$\hat{\beta}^{2SLS} = \left(\widehat{Cov}(\tilde{Z}_i, X_i)'\widehat{Var}(\tilde{Z}_i)^{-1}\widehat{Cov}(\tilde{Z}_i, X_i)\right)^{-1}\widehat{Cov}(\tilde{Z}_i, X_i)'\widehat{Var}(\tilde{Z}_i)^{-1}\widehat{Cov}(\tilde{Z}_i, Y_i)$$
$$= (\hat{\pi}'\widehat{Var}(\tilde{Z}_i)\hat{\pi})^{-1}\hat{\pi}'\widehat{Var}(\tilde{Z}_i)\hat{\rho}$$

where $\hat{\pi} = \widehat{Var}(\tilde{Z}_i)^{-1}\widehat{Cov}(\tilde{Z}_i, X_i)$ are the first-stage coefficient estimates and $\hat{\rho} = \widehat{Var}(\tilde{Z}_i)^{-1}\widehat{Cov}(\tilde{Z}_i, Y_i)$ are the reduced-form coefficient estimates

$\hat{\beta}^{2SLS}$ thus comes from a weighted OLS of $\hat{\rho}$ on $\hat{\pi}$ (with no constant)

- When $J = 1$ (one treatment), $\hat{\beta}^{2SLS} = \sum_\ell \hat{\omega}_\ell \hat{\beta}_\ell^{IV}$ where $\hat{\beta}^{IV} = \hat{\rho}_\ell / \hat{\pi}_\ell$ is the ℓth instrument's IV estimate, $\hat{\omega}_\ell = (\hat{\pi}'\widehat{Var}(\tilde{Z}_i)\hat{\pi})^{-1}\hat{\pi}'\widehat{Var}(\tilde{Z}_i)'_\ell\hat{\pi}_\ell$

Angrist & Hull '23: "Visual IV" for CRC Screening Trials



Note: each point gives an estimate from a colorectal cancer (CRC) screening trial which randomized offers for colonoscopy/sigmoidoscopy

Overidentification Testing

If we take our causal model of $Y_i = \beta X_i + \varepsilon_i$ seriously, overidentification gives us a way to test instrument validity

- All just-id. estimands should coincide: $\beta_\ell^{IV} = \frac{Cov(\tilde{Z}_\ell, Y_i)}{Cov(\tilde{Z}_\ell, X_i)} = \beta$, for all ℓ
- Graphically: the R^2 from our visual IV should $= 1$ in the population

Stata will automatically give you the p-value for this test when $L > J$

- Special case of quite general GMM overid. test stat. (Hansen's J)
- If $p > 0.05$, good news: your $\hat{\beta}^{IV}$ are sufficiently similar to each other

Don't place too much stock in overid. tests, however

- They tend to have low power (b/c individual $\hat{\beta}^{IV}$ tend to be noisy)
- If they reject, it need not mean your instruments are invalid (b/c of treatment effect heterogeneity — much more on this soon!)
- Even with constant effects, rejection doesn't tell you *which* β^{IV} is correct (indeed, none of them may be)

Outline

1. Just-Identified IV ✓
2. Overidentification ✓
3. Weak- and Many-Instrument Bias
4. Application: Abdulkadiroglu et al. (2016)

Weak Instruments

When running just-id. IV, always check the "strength" of your instrument

- Specifically the first-stage F-statistic, which tests $\pi = 0$

If π is small relative to its standard error, the IV is "weak"

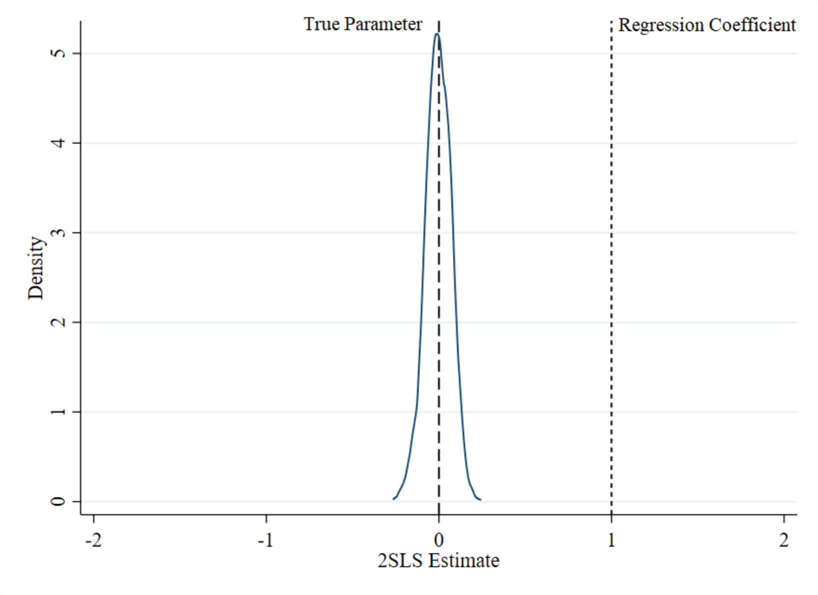
- Typically we use the rule-of-thumb of $F < 10$ (Staiger & Stock 1997)
- In this case the second-stage SEs will be large, and the 2SLS estimate will tend to be biased towards the corresponding OLS

Much has been made of this over the years, but Angrist and Kolesar (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"

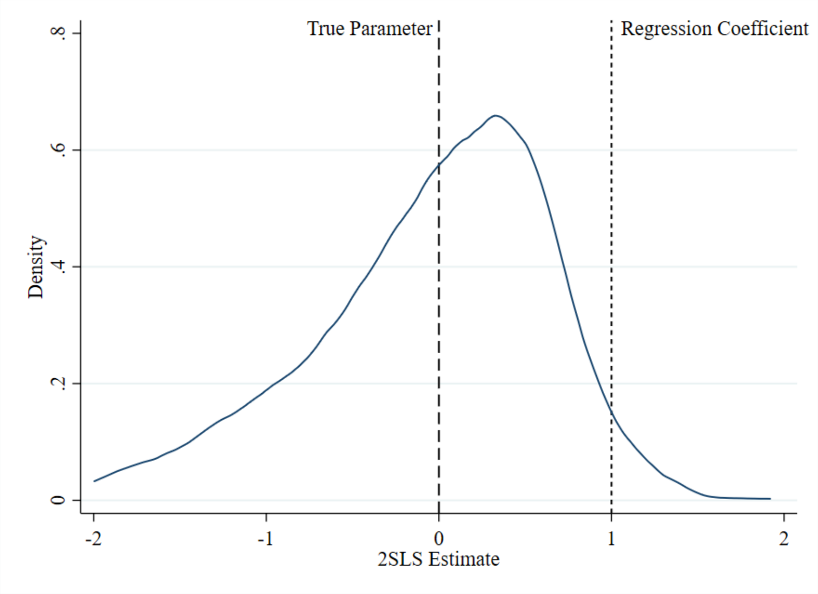
Weak-IV Bias: Visualized

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



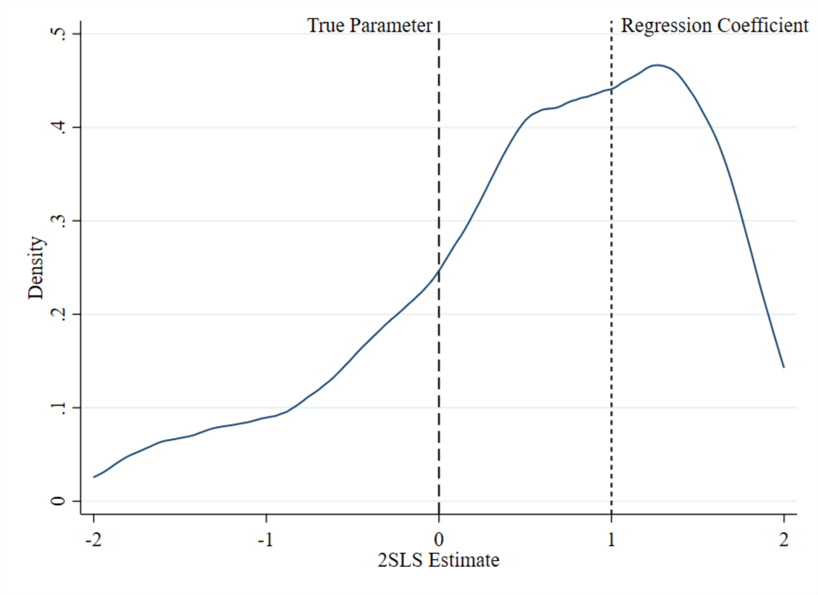
Weak-IV Bias: Visualized

Monte Carlo: $Y_i = \varepsilon_i$, $D_i = \pi Z_i + \eta_i$, $\pi = 0.1 < Var(\eta_i) = 1$



Weak-IV Bias: Visualized

Monte Carlo: $Y_i = \varepsilon_i$, $D_i = \pi Z_i + \eta_i$, $\pi = 0.01 \ll Var(\eta_i) = 1$



Many IV

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

- This also tends to manifest in low first-stage F's, and causes 2SLS to be biased towards OLS

Unlike when just-id., with many-IV bias the SE's go *down*

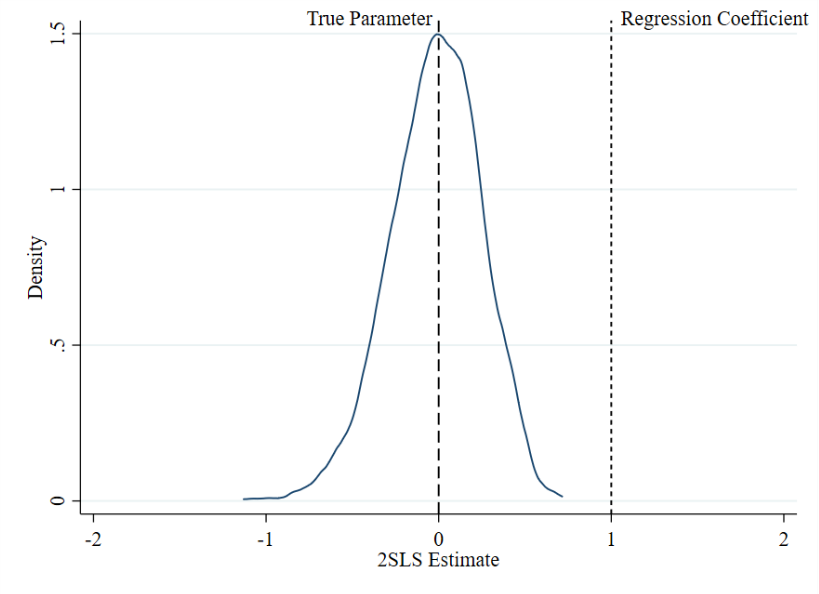
- Intuitively, a more flexible first stage tends to fit D_i better, giving more power in the second stage
- But we can have *overfitting* with lots of instruments, which essentially recreates the (endogenous) variation in D_i

The original cause for concern was Angrist-Krueger '91, where the QOB instrument was interacted with many state/year FE to boost power

- These days folks don't make this mistake ... but many-IV bias can be lurking in other settings with constructed instruments (e.g. judge IV)

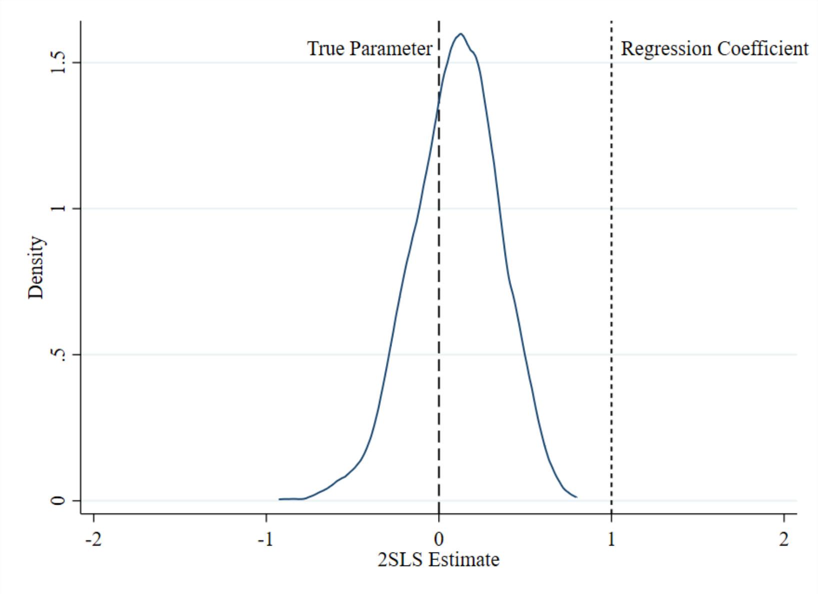
Many-IV Bias: Visualized

Monte Carlo: $Y_i = \varepsilon_i$, $D_i = \pi Z_{i1} + \eta_i$, IV with just Z_{i1}



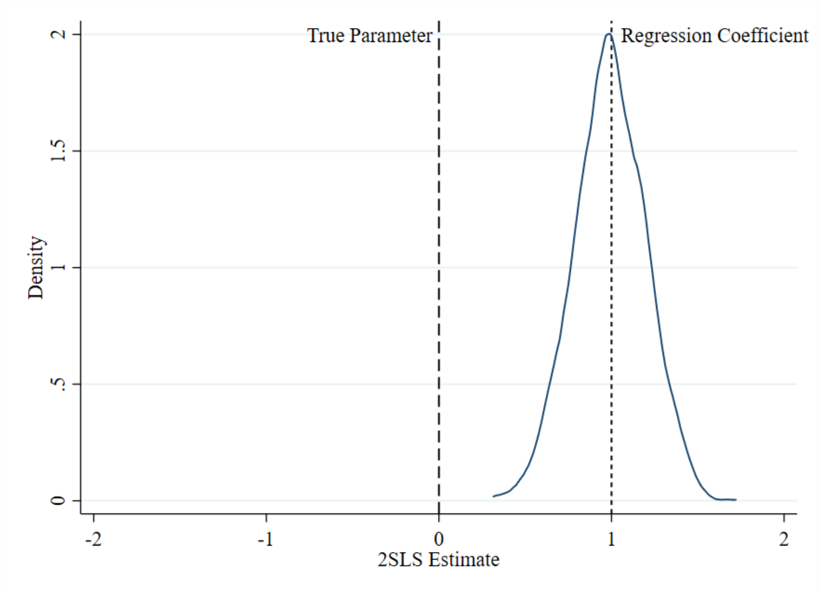
Many-IV Bias: Visualized

Monte Carlo: $Y_i = \varepsilon_i$, $D_i = \pi Z_{i1} + \eta_i$, IV with ten $Z_{i\ell}$



Many-IV Bias: Visualized

Monte Carlo: $Y_i = \varepsilon_i$, $D_i = \pi Z_{i1} + \eta_i$, IV with 100 $Z_{i\ell}$



What to Do?

Aim for few instruments, and check your F's after every *ivreg*

- State of the art: Montiel-Olea and Pflueger '15; *weakivtest* in Stata
- Staiger-Stock rule-of-thumb ($F > 10$) still seems widely held
- See Lee et al. (2022) and Keane and Neal (2022) for some discussions of additional subtleties

If your F is small, some things to consider:

- Is there a better functional form for your instrument?
- Do interactions with covariates help? (note: beware many-weak!)
- Does changing the covariate set help? (note: beware invalidity!)

Check results with a more robust approach (e.g. Anderson-Rubin, JIVE...)

Outline

1. Just-Identified IV✓
2. Overidentification✓
3. Weak- and Many-Instrument Bias✓
4. Application: Abdulkadiroglu et al. (2016)

Playing the Charter Lottery

Abdulkadiroglu et al. (2016) are interested in whether going to a "charter" middle school increases standardized test score performance

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

We leverage a unique 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 $X_i \in \{0, 1\}$, so are plausibly valid instruments
- We just need to control for lottery FE ("risk sets") to make Z_i as-good-as-randomly assigned

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

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

UP Academy: Lottery Balance Checks

	Sample means					Balance coefficients	
	Boston students	Grandfathering-eligible students	Lottery applicants		Lottery compliers	Immediate offer	Waitlist offer
			6th grade	7th grade			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				A. Balance			
Hispanic	0.374	0.241	0.327	0.294	0.326	-0.023	0.035
						(0.033)	(0.037)
Black	0.375	0.469	0.483	0.373	0.510	0.004	-0.026
						(0.036)	(0.040)
Female	0.487	0.483	0.504	0.471	0.553	0.003	-0.017
						(0.036)	(0.040)
Special education	0.222	0.317	0.231	0.275	0.216	0.032	-0.001
						(0.031)	(0.034)
Free/reduced price lunch	0.794	0.928	0.802	0.843	0.815	0.019	-0.024
						(0.029)	(0.032)
N	6,744	290	911	51	434	962	962
Baseline math test score	0.003	-0.253	-0.054	-0.081	-0.069	0.003	-0.064
						(0.066)	(0.073)
N	6,501	258	897	48	426	945	945
Baseline ELA test score	0.006	-0.235	-0.030	-0.169	-0.081	-0.060	0.018
						(0.066)	(0.074)
N	6,387	254	890	47	422	937	937
				B. Attrition			
Has first exposure year outcomes	0.917	0.855	0.924	0.843	0.964	-0.016	0.033
						(0.020)	(0.021)
N	6,744	290	911	51	434	962	962
Has second exposure year outcomes	0.878	0.817	0.872	0.784	0.883	-0.034	0.014
						(0.025)	(0.028)
N	6,744	164	911	51	434	962	962

UP Academy: Lottery IV

		Comparison group mean (1)	OLS (2)	2SLS: First stage: Immediate offer (3)	2SLS: First stage: Waitlist offer (4)	2SLS: 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)
Panel B. By potential exposure						
First exposure year (sixth and seventh grades)	Math (N = 881)	0.056	0.347	0.519	0.397	0.365
			(0.044)	(0.034)	(0.038)	(0.086)
	ELA (N = 882)	0.058	0.239	0.521	0.394	0.220
			(0.044)	(0.034)	(0.038)	(0.088)
Second and third exposure year (seventh and eighth grades)	Math (N = 1,321)	0.061	0.294	0.921	0.665	0.242
			(0.021)	(0.088)	(0.091)	(0.054)
	ELA (N = 1,323)	0.129	0.131	0.918	0.668	0.083
			(0.020)	(0.088)	(0.091)	(0.047)

$$Y_{it} = \beta D_{it} + \sum_j \kappa_j d_{ij} + X'_{it}\gamma + \varepsilon_{it}$$

$$D_{it} = \pi_1 Z_{i1} + \pi_2 Z_{i2} + \sum_j \mu_j d_{ij} + X'_{it}\phi + \nu_{it}$$

Charter Effects Without Lotteries

The UP effects are consistent with a large literature showing large test score effects from urban charters, via lottery IV

- But most students don't apply to charters; are effects generalizable?
- Lottery IV may have high *internal validity* but limited *external validity*

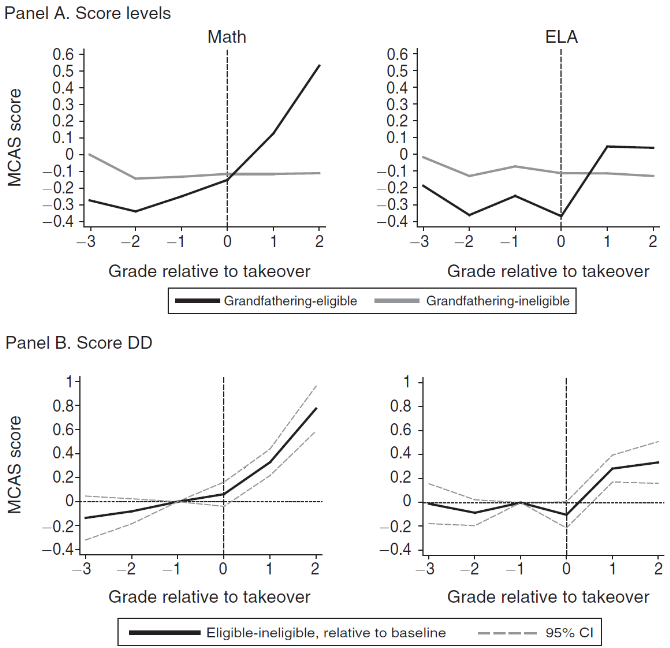
AAHP (2016) thus devise an alternative "grandfathering" IV strategy

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

This IV is less clean: takeovers may have direct effects on students

- Our baseline analysis differences off the takeover year outcomes, using an additive model for any such "disruption" effects (as in DiD)

UP Academy: Grandfathering Trends



UP Academy: Grandfathering IV

				2SLS	
		Comparison group mean (1)	OLS (2)	First stage (3)	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)
Panel B. By potential exposure					
First exposure year (seventh and eighth grades)	Math (N = 1,028)	−0.214	0.365	0.822	0.325
			(0.047)	(0.025)	(0.048)
	ELA (N = 1,025)	−0.195	0.475	0.809	0.495
			(0.055)	(0.026)	(0.060)
Second exposure year (eighth grade)	Math (N = 515)	−0.272	0.408	1.541	0.324
			(0.038)	(0.087)	(0.044)
	ELA (N = 514)	−0.252	0.221	1.543	0.271
			(0.042)	(0.087)	(0.049)

$$Y_{it}^{c} - Y_{it}^{\ell} = \beta D_{it} + \sum_j \kappa_j d_{ij} + X_{it}'\gamma + \varepsilon_{it}$$

$$D_{it} = \pi Z_i + \sum_j \mu_j d_{ij} + X_{it}'\phi + \nu_{it}$$

Shipping Down From Boston

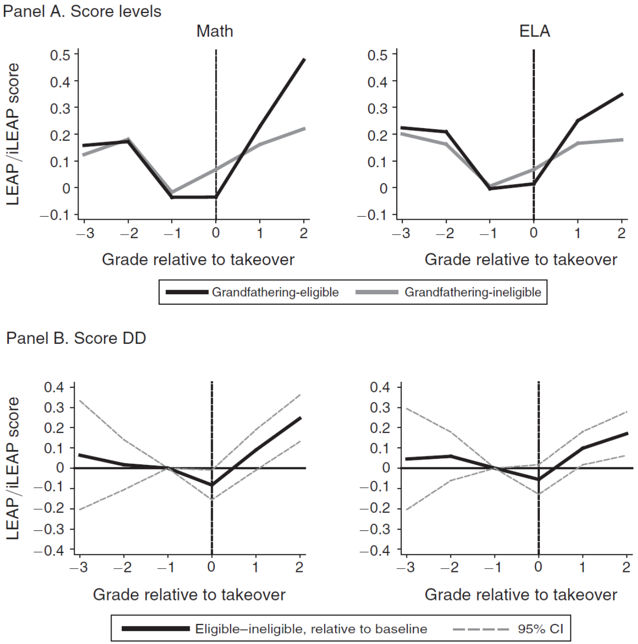
AAHP (2016) actually first develop the grandfathering IV tech in NOLA

- In the aftermath of Hurricane Katrina, charters gradually took over the entire NOLA Recovery School District (mostly without lotteries)
- As RSD changed, the counterfactual enrollment for a given takeover "experiment" became increasingly populated with other charters

To account for the changing counterfactual, we use multiple-treatment IV

- Jointly estimate the effects of attending a takeover charter and other RSD charters, using interactions of the grandfathering instrument
- Intuition: the GF-IV has different first stage effects over time/etc, and this separates multiple-treatment effects (under constant fx...)
- See Hull (2018) for a more formal treatment of this strategy

NOLA RSD: Grandfathering Trends



NOLA RSD: Grandfathering IV

		Comparison group mean (1)	OLS (2)	2SLS First stage (3)	2SLS Enrollment effect (4)
Panel A. All grades					
(Fifth through eighth)	Math (N = 5,625)	−0.089	0.123	1.073	0.212
			(0.020)	(0.052)	(0.038)
	ELA (N = 5,621)	−0.092	0.082	1.075	0.143
			(0.018)	(0.052)	(0.039)
Panel B. By grade					
Fifth and sixth grades	Math (N = 2,579)	−0.091	0.099	0.738	0.165
			(0.035)	(0.041)	(0.068)
	ELA (N = 2,579)	−0.116	0.023	0.745	0.101
			(0.033)	(0.042)	(0.070)
Seventh and eighth grades	Math (N = 3,046)	−0.086	0.133	1.355	0.231
			(0.020)	(0.070)	(0.037)
	ELA (N = 3,042)	−0.071	0.104	1.352	0.171
			(0.019)	(0.070)	(0.036)
Panel C. By potential exposure					
First exposure year (fifth through eighth grades)	Math (N = 2,553)	−0.105	0.200	0.659	0.230
			(0.044)	(0.023)	(0.069)
	ELA (N = 2,553)	−0.103	0.099	0.659	0.197
			(0.043)	(0.023)	(0.068)
Second exposure year (sixth through eighth grades)	Math (N = 1,664)	−0.151	0.168	1.148	0.332
			(0.031)	(0.061)	(0.058)
	ELA (N = 1,664)	−0.124	0.101	1.158	0.158
			(0.028)	(0.061)	(0.051)
Third and fourth exposure year (seventh and eighth grades)	Math (N = 1,408)	0.015	0.097	1.698	0.117
			(0.022)	(0.131)	(0.042)
	ELA (N = 1,404)	−0.033	0.077	1.698	0.094
			(0.020)	(0.132)	(0.043)

NOLA RSD: Multiple-Treatment IV

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Math						
Takeover charter	0.210	0.358	0.344			
	(0.037)	(0.079)	(0.138)			
	[72.3]	[33.4]	[194.9]			
Other RSD charter		0.342	0.308			
		(0.164)	(0.376)			
		[12.6]	[27.4]			
Any RSD charter				0.366	0.359	0.385
				(0.060)	(0.095)	(0.071)
				[41.7]	[237.9]	[223.4]
No. of instruments	23	23	2	23	2	1
N	5,625	5,625	2,553	5,625	2,553	5,625

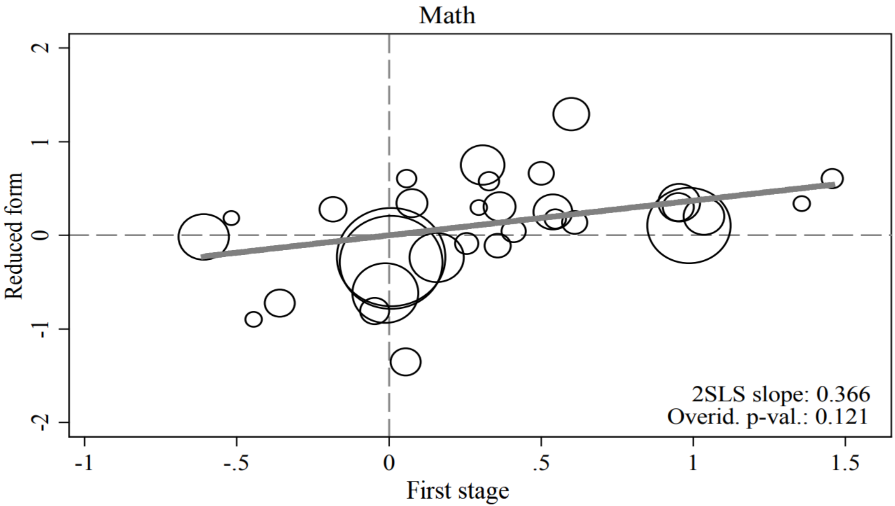
Note: Angrist-Pischke first-stage F statistics reported in brackets

$$Y_{it}^{c} - Y_{it}^{\ell} = \beta_D D_{it} + \beta_C C_{it} + \sum_j \kappa_j d_{ij} + X_{it}'\gamma + \varepsilon_{it}$$

$$D_{it} = \pi_D Z_i + \psi_D X_{it} Z_i + \sum_j \mu_j d_{Dij} + X_{it}'\phi_D + \nu_{Dit}$$

$$C_{it} = \pi_C Z_i + \psi_D X_{it} Z_i + \sum_j \mu_j d_{Cij} + X_{it}'\phi_C + \nu_{Cit}$$

NOLA RSD: Visual IV of Any Charter Effect



Note: from RF and FS from column 4 of the previous table

Summary

2SLS is a many-splendored thing; its mechanics are worth knowing well

- "Reduced form over first stage" for just-identified specifications
- "Weighted average of RF/FS" for over-identified specifications
- OLS gives the key building blocks, so remember those mechanics too!

Different IVs may score differently on internal & external validity

- But we don't yet have a great framework for thinking through either!
- The LATE theorem of Imbens and Angrist, and various extensions, help clarify the two core ingredients of (internal) IV "validty"
- By relaxing constant effects, LATE also helps with external validity / generalizability: whose treatment effects are we estimating?