

Instrumental variables

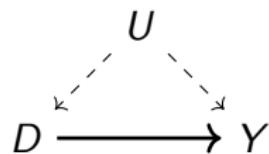
- If treatment is tied to an unobservable, then conditioning strategies, even RDD, are invalid
- Instrumental variables offers some hope at recovering the causal effect of D on Y
- The best instruments come from deep knowledge of institutional details (Angrist and Krueger 1991)
- Certain types of natural experiments can be the source of such opportunities and may be useful

When is IV used?

Instrumental variables methods are typically used to address the following kinds of problems encountered in naive regressions

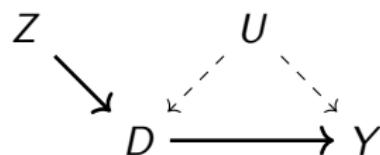
- ① Omitted variable bias
- ② Measurement error
- ③ Simultaneity bias
- ④ Reverse causality
- ⑤ Randomized control trials with noncompliance

Selection on unobservables



Then D is endogenous due to backdoor path $D \leftarrow U \rightarrow Y$ and causal effect $D \rightarrow Y$ is not identified using the backdoor criterion.

Instruments



Notice how the path from $Z \rightarrow D \leftarrow U \rightarrow Y$ is blocked by a collider.

Phillip Wright

- Philip Wright was a renaissance man - published in JASA, QJE, AER, you name it, while on a very intense teaching load.
- Also published poetry, and even personally published Carl Sandburg's first book of poetry!
- Spent a long time at Tufts
- He was very concerned about the negative effects of tariffs and wrote a book about commodity markets

Elasticity of demand is unidentified

- James Stock notes that his publications had a theme regarding identification
- He knew, for instance, that he couldn't simply look at correlations between price and quantity if he wanted the elasticity of demand due to simultaneous shifts in supply and demand
- The pairs of quantity and price weren't demand, or supply – they were demand and supply equilibrium values and therefore didn't reflect the demand or the supply curve, both of which are counterfactuals
- Those points are nothing more than a bunch of numbers – no more, no less – that have no practical use, scientific or otherwise

Exhibit 1

The Graphical Demonstration of the Identification Problem in Appendix B (p. 296)

FIGURE 4. PRICE-OUTPUT DATA FAIL TO REVEAL EITHER SUPPLY OR DEMAND CURVE.

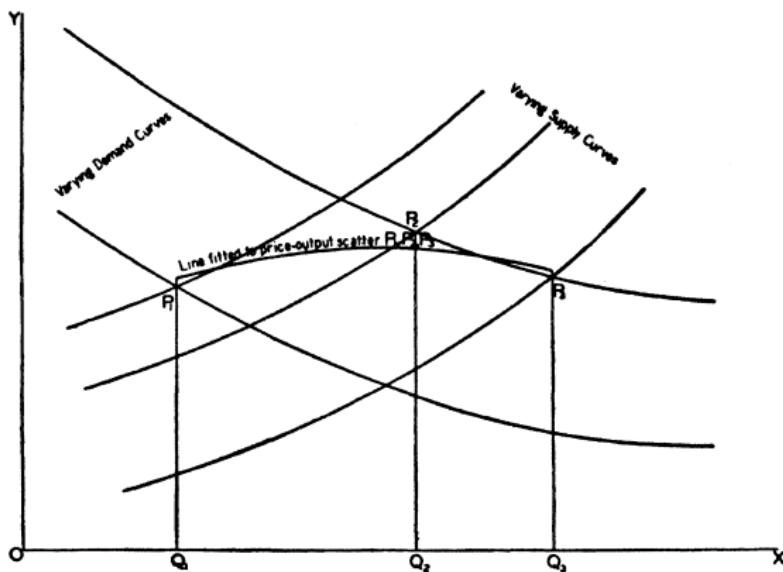


Figure: Wright's graphical demonstration of the identification problem

Sewell Wright

- Sewell was his son, who did *not* go into the family business
- Rather, he decided to become a genius and invent genetics
- Developed path diagrams (which Pearl revived 50 years later for causal inference)
- Father and son engage in letter correspondence as Philip tried to solve the “identification problem”

March 4, 1926.

Dear Sewell:

It may interest you to see a very simple geometric demonstration which I have worked out for you without of estimating supply and demand curves without reference to the theory of path coefficients.

Figure: Wright's letter to Sewell, his son

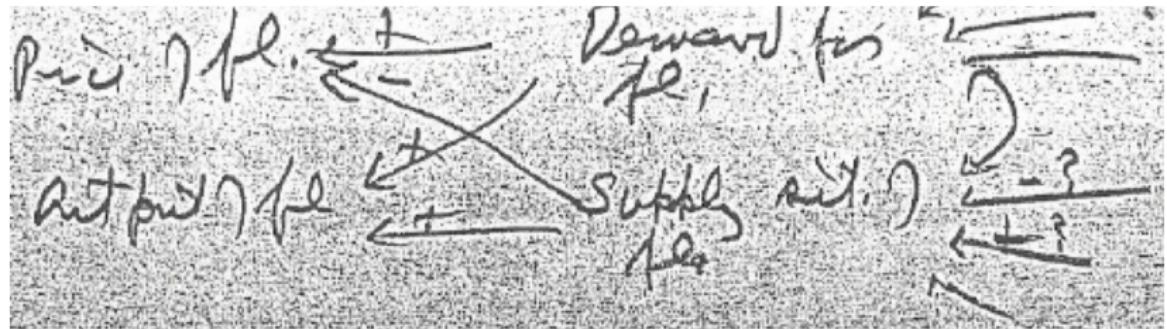


Figure: Recognize these?

QJE Rejects

- QJE misses a chance to make history and rejects his paper proving an IV estimator
- Sticks his proof in Appendix B of 1928 book,
The Tariff on Animal and Vegetable Oils
- His work on IV is ignored, and is then rediscovered 15 years later (e.g., Olav Reiersøl).
- James Stock and others have helped correct the record

Sidebar: stylometric analysis

- Long standing question was who *wrote* Appendix B? Answer according to Stock and Trebbi (2003) using stylometric methods is that Philip *wrote* it.
- But who invented it? It was collaborative, but Sewell acknowledged he didn't know how to handle endogeneity and simultaneity (that was Philip)

Constant treatment effects

- Constant treatment effects (i.e., β is constant across all individual units)
 - Constant treatment effects is the traditional econometric pedagogy when first learning instrumental variables, and doesn't need the potential outcomes model or notation to get the point across
 - Constant treatment effects is identical to assuming that $ATE=ATT=ATU$ because constant treatment effects assumes $\beta_i = \beta_{-i} = \beta$ for all units

Heterogenous treatment effects

- Heterogeneous treatment effects (i.e., β_i varies across individual units)
 - Heterogeneous treatment effects means that the $ATE \neq ATT \neq ATU$ because β_i differs across the population
 - This is equivalent to assuming the coefficient, β_i , is a random variable that varies across the population
 - Heterogenous treatment effects is based on work by Angrist, Imbens and Rubin (1996) and Imbens and Angrist (1994) which introduced the “local average treatment effect” (LATE) concept

Data requirements

- Your data isn't going to come with a codebook saying "instrumental variable". So how do you find it?
- Well, sometimes the researcher just *knows*.
- That is, the researcher knows of a variable (Z) that actually *is* randomly assigned and that affects the endogenous variable but not the outcome (except via the endogenous variable)
- Such a variable is called an "instrument".

Picking a good instrument

- The best instruments you think of first, then you seek the data second (but often students go in the reverse order which is basically guaranteed to be a crappy instrument)
- If you want to use IV, then ask:

What moves around the covariate of interest that might be plausibly random?

- Is there any element in the treatment that could be construed as random?
- If you were to find that random piece, then you have found an instrument
- Once you have identified such a variable, begin to think about what data sets might have information on an outcome of interest, the treatment, and the instrument you have put your finger on.

Does family size reduce labor supply or is it selection?

Angrist and Evans (1998), "Children and their parents' labor supply" *American Economic Review*,

- They want to know the effect of family size on labor supply, but need exogenous changes in family size
- So what if I told you if the first two children born were of the same gender, then you're less likely to work. What?!

Angrist and Evans cont.

- Many parents have a preference for having at least one child of each gender
 - Consider a couple whose first two kids were both boys; they will often have a third, hoping to have a girl
 - Consider a couple whose first two kids were girls; they will often have a third, hoping for a boy
 - Consider a couple with one boy and one girl; they will often not have a third kid
- The gender of your kids is arguably randomly assigned (maybe not exactly, but close enough)

Good instruments must be a bit strange

- On its face, it's puzzling that the first two kids' gender predicts labor market participation
- Instrumental variables strategies formalize *strangeness of the instrument*, which is the inference drawn by an intelligent layperson with no particular knowledge of the phenomena or background in statistics.
- You need more information, in other words, otherwise the layperson can't understand what same gender of your children has to do with working

When a good IV strategy finally makes sense

- But then the researchers point out that women whose first two children are of the same gender are more likely to have additional children than women whose first two children are of different genders
- The layperson then asks himself, “Hm. I wonder if the labor market differences are due *solely* to the differences in the number of kids the woman has...”

Sunday Candy is a good instrument

- Let's listen to a few lines from "Ultralight Beam" by Kanye West. Chance the Rapper sings on it and says

*"I made Sunday Candy, I'm never going to hell
I met Kanye West, I'm never going to fail."
- Chance the Rapper*

- What does making a song have to do with hell? What does meeting Kanye West have to do with success? Let's consider each in order

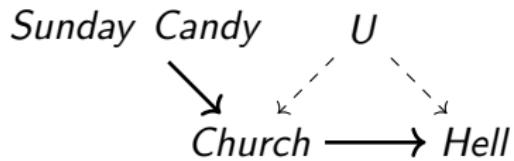
What are we missing?

*"I made Sunday Candy,
I'm never going to hell",*

- There must be more to this story, right?
- So what if it's something like this

*"I made Sunday Candy
this pastor invited me to church on Sunday,
I'm never going to hell"*

Sunday Candy DAG



Kanye West is a bad instrument

- Chance long idolized and was inspired by Kanye West – both Chicago, both very creative hip hop artists
- Kanye West is not a good instrument for Chance's inspiration, though, because Kanye West can singlehandedly make a person's career
- Kanye is not strange enough

Kanye West DAG



Foreshadowing the questions you need to be asking

- ① Is our instrument highly correlated with the treatment? With the outcome? Can you test that?
- ② Are there random elements within the treatment? Why do you think that?
- ③ Is the instrument exogenous? Why do you think that?
- ④ Could the instrument affect outcomes directly? Why do you think that?
- ⑤ Could the instrument be associated with anything that causes the outcome even if it doesn't directly? Why do you think that?

Our causal model: Returns to schooling again

$$Y = \alpha + \delta S + \gamma A + \nu$$

where Y is log earnings, S is years of schooling, A is unobserved ability, and ν is the error term

- Suppose there exists a variable, Z_i , that is correlated with S_i .
- We can estimate δ with this variable, Z :

How can IV be used to obtain consistent estimates?

$$\begin{aligned}\text{Cov}(Y, Z) &= \text{Cov}(\alpha + \delta S + \gamma A + \nu, Z) \\&= E[(\alpha + \delta S + \gamma A + \nu)Z] - E[\alpha + \delta S + \gamma A + \nu]E[Z] \\&= \{\alpha E(Z) - \alpha E(Z)\} + \delta\{E(SZ) - E(S)E(Z)\} \\&\quad + \gamma\{E(AZ) - E(A)E(Z)\} + E(\nu Z) - E(\nu)E(Z) \\ \text{Cov}(Y, Z) &= \delta \text{Cov}(S, Z) + \gamma \text{Cov}(A, Z) + \text{Cov}(\nu, Z)\end{aligned}$$

Divide both sides by $\text{Cov}(S, Z)$ and the first term becomes δ , the LHS becomes the ratio of the reduced form to the first stage, plus two other scaled terms.

Consistency

- What conditions must hold for a valid IV design?
 - $\text{Cov}(S, Z) \neq 0$ – “first stage” exists. S and Z are correlated
 - $\text{Cov}(A, Z) = \text{Cov}(\nu, Z) = 0$ – “exclusion restriction”. This means Z that orthogonal to the factors in ν , such as unobserved ability, A , as well as the structural disturbance term, ν
- Assuming the first stage exists and that the exclusion restriction holds, then we can estimate δ with δ_{IV} :

$$\begin{aligned}\delta_{IV} &= \frac{\text{Cov}(Y, Z)}{\text{Cov}(S, Z)} \\ &= \delta\end{aligned}$$

IV is Consistent if IV Assumptions are Satisfied

- The IV estimator is consistent if the IV assumptions are satisfied. Substitute true model for Y :

$$\begin{aligned}\delta_{IV} &= \frac{\text{Cov}([\alpha + \rho S + \gamma A + \nu], Z)}{\text{Cov}(S, Z)} \\ &= \delta \frac{\text{Cov}([S], Z)}{\text{Cov}(S, Z)} + \gamma \frac{\text{Cov}([A], Z)}{\text{Cov}(S, Z)} + \frac{\text{Cov}([\nu], Z)}{\text{Cov}(S, Z)} \\ &= \delta + \gamma \frac{\text{Cov}(\eta, Z)}{\text{Cov}(S, Z)}\end{aligned}$$

Identifying assumptions and consistency

- Taking the probability limit which is an asymptotic operation to show consistency:

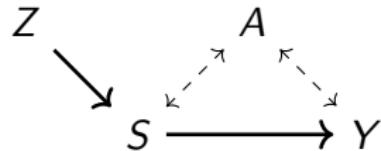
$$\begin{aligned}\text{plim } \widehat{\delta}_{IV} &= \text{plim } \delta + \gamma \frac{\text{Cov}(\eta, Z)}{\text{Cov}(S, Z)} \\ &= \delta\end{aligned}$$

because $\text{Cov}([A], Z) = 0$ and $\text{Cov}([\nu], Z) = 0$ due to the exclusion restriction, and $\text{Cov}(S, Z) \neq 0$ (due to the first stage)

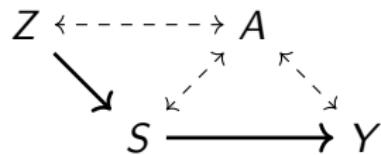
IV Assumptions

- But, if Z is *not* independent of η (either correlated with A or ν), *and* if the correlation between S and Z is “weak”, then the second term blows up.
- We will explore the problems created by weak instruments in just a moment.
- First, let’s look at a DAG summarizing all this information

One of these DAGs is not like the other



(a)



(b)

Notice - the top DAG, *a*, satisfies both exclusion and relevance (i.e., non-zero first stage), but the bottom DAG, *b*, satisfies relevance but not exclusion.

Two-stage least squares

- The two-stage least squares estimator was developed by Theil (1953) and Basman (1957) independently
- Note, while IV is a research design, 2SLS is a specific estimator.
- Others include LIML, the Wald estimator, jackknife IV, two sample IV, and more

Two Sample IV

- In a pinch, you can even get by with two different data sets
 - ❶ Dataset 1 needs information on the outcome and the instrument
 - ❷ Dataset 1 needs information on the treatment and the instrument.
- This is known as “Two sample IV” because there are two *samples* involved, rather than the traditional one sample.
- Once we define what IV is measuring carefully, you will see why this works.

Two-stage least squares concepts

- Causal model. Sometimes called the structural model:

$$Y_i = \alpha + \delta S_i + \eta_i$$

- First-stage regression. Gets the name because of two-stage least squares:

$$S_i = \gamma + \rho Z_i + \zeta_i$$

- Second-stage regression. Notice the fitted values, \hat{S} :

$$Y_i = \beta + \delta \hat{S}_i + \nu_i$$

Reduced form

- Some people like a simpler approach because they don't want to defend IV's assumptions
- Reduced form a regression of Y onto the instrument:

$$Y_i = \psi + \pi Z_i + \varepsilon_i$$

- This would be like regressing hell onto Sunday Candy, as opposed to regressing hell onto church with Sunday Candy instrumenting for church

Two-stage least squares

Suppose you have a sample of data on Y , X , and Z . For each observation i we assume the data are generated according to

$$Y_i = \alpha + \delta S_i + \eta_i$$

$$S_i = \gamma + \rho Z_i + \zeta_i$$

where $\text{Cov}(Z, \eta_i) = 0$ and $\rho \neq 0$.

Two-stage least squares

Plug in covariance and write out the following:

$$\begin{aligned}\widehat{\delta_{2sls}} &= \frac{\text{Cov}(Z, Y)}{\text{Cov}(Z, S)} \\ &= \frac{\frac{1}{n} \sum_{i=1}^n (Z_i - \bar{Z})(Y_i - \bar{Y})}{\frac{1}{n} \sum_{i=1}^n (Z_i - \bar{Z})(S_i - \bar{S})} \\ &= \frac{\frac{1}{n} \sum_{i=1}^n (Z_i - \bar{Z})Y_i}{\frac{1}{n} \sum_{i=1}^n (Z_i - \bar{Z})S_i}\end{aligned}$$

Two-stage least squares

Substitute the causal model definition of Y to get:

$$\begin{aligned}\widehat{\delta_{2sls}} &= \frac{\frac{1}{n} \sum_{i=1}^n (Z_i - \bar{Z}) \{\alpha + \delta S_i + \eta_i\}}{\frac{1}{n} \sum_{i=1}^n (Z_i - \bar{Z}) S_i} \\ &= \delta + \frac{\frac{1}{n} (Z_i - \bar{Z}) \eta_i}{\frac{1}{n} \sum_{i=1}^n (Z_i - \bar{Z}) S_i} \\ &= \delta + \text{"small if } n \text{ is large"}$$

Where did the first term go? Why did the second term become δ ?

Two-stage least squares

- Calculate the ratio of “reduced form” (π) to “first stage” coefficient (ρ):

$$\hat{\delta}_{2sls} = \frac{Cov(Z, Y)}{Cov(Z, S)} = \frac{\frac{Cov(Z, Y)}{Var(Z)}}{\frac{Cov(Z, S)}{Var(Z)}} = \frac{\hat{\pi}}{\hat{\rho}}$$

- Rewrite $\hat{\rho}$ as

$$\begin{aligned}\hat{\rho} &= \frac{Cov(Z, S)}{Var(Z)} \\ \hat{\rho}Var(Z) &= Cov(Z, S)\end{aligned}$$

Two-stage least squares

Then rewrite $\hat{\delta}_{2sls}$

$$\begin{aligned}\hat{\delta}_{2sls} &= \frac{\text{Cov}(Z, Y)}{\text{Cov}(Z, S)} = \frac{\hat{\rho}\text{Cov}(Z, Y)}{\hat{\rho}\text{Cov}(Z, S)} = \frac{\hat{\rho}\text{Cov}(Z, Y)}{\hat{\rho}^2\text{Var}(Z)} \\ &= \frac{\text{Cov}(\hat{\rho}Z, Y)}{\text{Var}(\hat{\rho}Z)}\end{aligned}$$

Two-stage least squares

Recall

$$S_i = \gamma + \rho Z_i + \zeta_i$$

Then

$$\hat{S} = \hat{\gamma} + \hat{\rho} Z$$

Then

$$\hat{\delta}_{2sls} = \frac{Cov(\hat{\rho}Z, Y)}{Var(\hat{\rho}Z)} = \frac{Cov(\hat{S}, Y)}{Var(\hat{S})}$$

Proof.

We will show that $\widehat{\delta} \text{Cov}(Y, Z) = \text{Cov}(\widehat{S}, Y)$. I will leave it to you to show that $\text{Var}(\widehat{\delta}Z) = \text{Var}(\widehat{S})$

$$\begin{aligned}\text{Cov}(\widehat{S}, Y) &= E[\widehat{S}Y] - E[\widehat{S}]E[Y] \\&= E(Y[\widehat{\rho} + \widehat{\delta}Z]) - E(Y)E(\widehat{\rho} + \widehat{\delta}Z) \\&= \widehat{\rho}E(Y) + \widehat{\delta}E(YZ) - \widehat{\rho}E(Y) - \widehat{\delta}E(Y)E(Z) \\&= \widehat{\delta}[E(YZ) - E(Y)E(Z)] \\ \text{Cov}(\widehat{S}, Y) &= \widehat{\delta} \text{Cov}(Y, Z)\end{aligned}$$



Intuition of 2SLS

- Two stage least squares is nice because in addition to being an estimator, there's also great intuition contained in it which you can use as a device for thinking about IV more generally.
- The intuition is that 2SLS estimator replaces S with the fitted values of S (i.e., \widehat{S}) from the first stage regression of S onto Z and all other covariates.
- By using the fitted values of the endogenous regressor from the first stage regression, our regression now uses *only* the exogenous variation in the regressor due to the instrumental variable itself

Intuition of IV in 2SLS

- ... but think about it – that variation was there before, but was just a subset of all the variation in the regressor
- Go back to what we said in the beginning - we need the endogenous variable to have pieces that are random, and IV finds them.
- Instrumental variables therefore reduces the variation in the data, but that variation which is left is *exogenous*
- “With a long enough [instrument], you can [estimate any causal effect]” - Scott Cunningham paraphrasing Archimedes

Estimation with software

- One manual way is just to estimate the reduced form and first stage coefficients and take the ratio of the respective coefficients on Z
- But while it is always a good idea to run these two regressions, don't compute your IV estimate this way

Estimation with software

- It is often the case that a pattern of missing data will differ between Y and S
- In such a case, the usual procedure of “casewise deletion” is to keep the subsample with non-missing data on Y , S , and Z .
- But the reduced form and first stage regressions would be estimated off of different sub-samples if you used the two step method before
- The standard errors from the second stage regression are also wrong

Estimation with software

- Estimate this in Stata using -ivregress 2sls-.
- Estimate this in R -ivreg()- which is in the AER package

Weak instruments

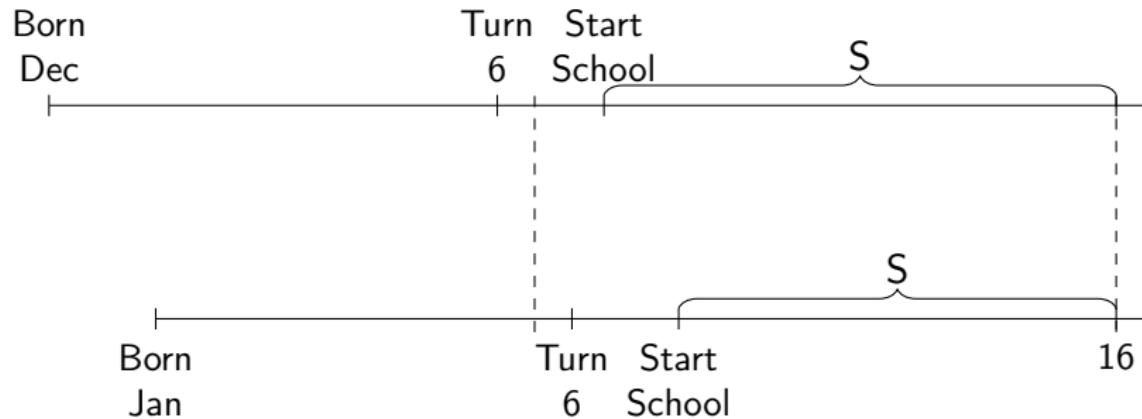
- A weak instrument is one that is not strongly correlated with the endogenous variable in the first stage
- This can happen if the two variables are independent or the sample is small
- If you have a weak instrument, then the bias of 2SLS is centered on the bias of OLS and the cure ends up being worse than the disease
- We knew this was a problem, but it was brought into sharp focus with Angrist and Krueger (1991) and some papers that followed

Angrist and Krueger (1991)

- In practice, it is often difficult to find convincing instruments – usually because potential instruments don't satisfy the exclusion restriction
- But in an early paper in the causal inference movement, Angrist and Krueger (1991) wrote a very interesting and influential study instrumental variable
- They were interested in schooling's effect on earnings and instrumented for it with *which quarter of the year you were born*
- Remember Chance quote - what the heck would birth quarter have to do with earnings such that it was an excludable instrument?

Compulsory schooling

- In the US, you could drop out of school once you turned 16
- "School districts typically require a student to have turned age six by January 1 of the year in which he or she enters school"
(Angrist and Krueger 1991, p. 980)
- Children have different ages when they start school, though, and this creates different lengths of schooling at the time they turn 16 (potential drop out age):



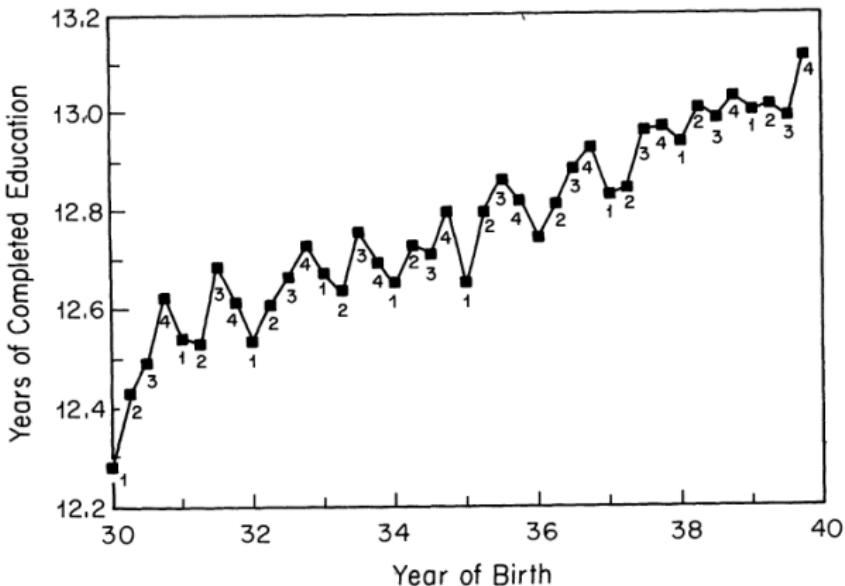
If you're born in the fourth quarter, you hit 16 with more schooling than those born in the first quarter

Visuals

- You need good data visualization for IV partly because of the scrutiny around the design
- The two pieces you should be ready to build pictures for are the first stage and the reduced form
- Angrist and Krueger (1991) provide simple, classic and compelling pictures of both

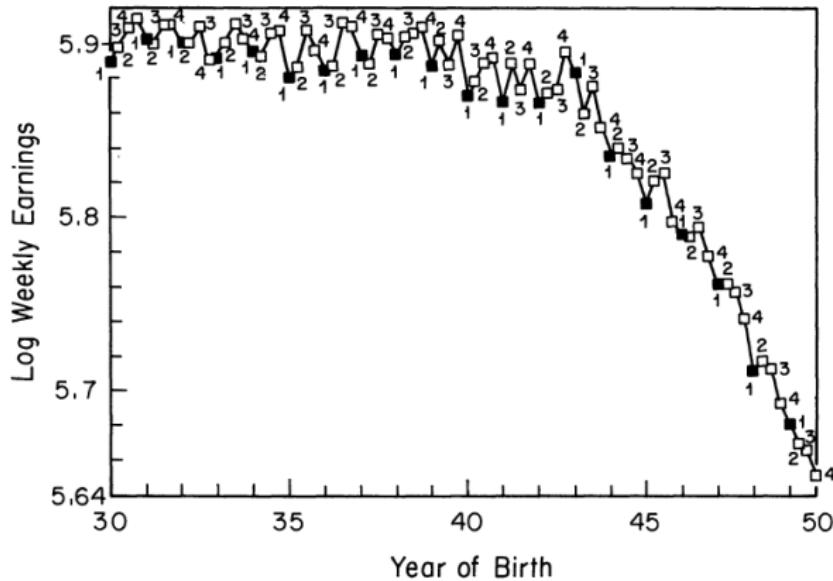
First Stage

Men born earlier in the year have lower schooling. This indicates that there is a first stage. Notice all the 3s and 4s at the top. But then notice how it attenuates over time . . .



Reduced Form

Do differences in schooling due to different quarter of birth translate into different earnings?



Two Stage Least Squares model

- The causal model is

$$Y_i = \delta S_i + \varepsilon$$

- The first stage regression is:

$$S_i = X\pi_{10} + \pi_{11}Z_i + \eta_{1i}$$

- The reduced form regression is:

$$Y_i = X\pi_{20} + \pi_{21}Z_i + \eta_{2i}$$

- The covariate adjusted IV estimator is the sample analog of the ratio, $\frac{\pi_{21}}{\pi_{11}}$

Two Stage Least Squares

- Angrist and Krueger instrument for schooling using three quarter of birth dummies: a dummies for 2nd, 3rd and 4th qob
- Their estimated first-stage regression is:

$$S_i = X\pi_{10} + Z_{1i}\pi_{11} + Z_{2i}\pi_{12} + Z_{3i}\pi_{13} + \eta_1$$

- The second stage is the same as before, but the fitted values are from the new first stage

First stage regression results

Quarter of birth is a strong predictor of total years of education

Outcome variable	Birth cohort	Mean	Quarter-of-birth effect ^a			F-test ^b [P-value]
			I	II	III	
Total years of education	1930–1939	12.79	-0.124 (0.017)	-0.086 (0.017)	-0.015 (0.016)	24.9 [0.0001]
	1940–1949	13.56	-0.085 (0.012)	-0.035 (0.012)	-0.017 (0.011)	18.6 [0.0001]
High school graduate	1930–1939	0.77	-0.019 (0.002)	-0.020 (0.002)	-0.004 (0.002)	46.4 [0.0001]
	1940–1949	0.86	-0.015 (0.001)	-0.012 (0.001)	-0.002 (0.001)	54.4 [0.0001]
Years of educ. for high school graduates	1930–1939	13.99	-0.004 (0.014)	0.051 (0.014)	0.012 (0.014)	5.9 [0.0006]
	1940–1949	14.28	0.005 (0.011)	0.043 (0.011)	-0.003 (0.010)	7.8 [0.0017]
College graduate	1930–1939	0.24	-0.005 (0.002)	0.003 (0.002)	0.002 (0.002)	5.0 [0.0021]
	1940–1949	0.30	-0.003 (0.002)	0.004 (0.002)	0.000 (0.002)	5.0 [0.0018]

First stage regression results: Placebos

Completed master's degree	1930–1939	0.09	−0.001	0.002	−0.001	1.7	
			(0.001)	(0.001)	(0.001)	[0.1599]	
Completed doctoral degree	1940–1949	0.11	0.000	0.004	0.001	3.9	
			(0.001)	(0.001)	(0.001)	[0.0091]	
	1930–1939	0.03	0.002	0.003	0.000	2.9	
			(0.001)	(0.001)	(0.001)	[0.0332]	
	1940–1949	0.04	−0.002	0.001	−0.001	4.3	
			(0.001)	(0.001)	(0.001)	[0.0050]	

a. Standard errors are in parentheses. An $MA(+2, -2)$ trend term was subtracted from each dependent variable. The data set contains men from the 1980 Census, 5 percent Public Use Sample. Sample size is 312,718 for 1930–1939 cohort and is 457,181 for 1940–1949 cohort.

b. F-statistic is for a test of the hypothesis that the quarter-of-birth dummies jointly have no effect.

IV Estimates Birth Cohorts 20-29, 1980 Census

Independent variable	(1) OLS	(2) TSLS
Years of education	0.0711 (0.0003)	0.0891 (0.0161)
Race (1 = black)	—	—
SMSA (1 = center city)	—	—
Married (1 = married)	—	—
9 Year-of-birth dummies	Yes	Yes
8 Region-of-residence dummies	No	No
Age	—	—
Age-squared	—	—
χ^2 [dof]	—	25.4 [29]

Sidebar: Wald estimator

- Recall that 2SLS uses the predicted values from a first stage regression – but we showed that the 2SLS method was equivalent to $\frac{\text{Cov}(Y, Z)}{\text{Cov}(X, Z)}$
- The Wald estimator simply calculates the return to education as the ratio of the difference in earnings by quarter of birth to the difference in years of education by quarter of birth – it's a version of the above
- Formally, $IV_{Wald} = \frac{E(Y|Z=1) - E(Y|Z=0)}{E(D|Z=1) - E(D|Z=0)}$

Mechanism

- In addition to log weekly wage, they examined the impact of compulsory schooling on log annual salary and weeks worked
- The main impact of compulsory schooling is on the log weekly wage – not on weeks worked

More instruments

To incorporate the cross-state seasonal variation in education, we computed TSLS estimates that use as instruments for education a set of three quarter-of-birth dummies interacted with fifty state-of-birth dummies, in addition to three quarter-of-birth dummies interacted with nine year-of-birth dummies.¹⁸ The estimates also include fifty state-of-birth dummies in the wage equation, so the variability in education used to identify the return to education in the TSLS estimates is solely due to differences by season of birth. Unlike the previous TSLS estimates, the seasonal differences are now allowed to vary by state as well as by birth year.

Problem enters with many quarter of birth interactions

- They want to increase the precision of their 2SLS estimates, so they load up their first stage with more instruments
- Specifications with 30 (quarter of birth \times year) dummy variables and 150 (quarter of birth \times state) instruments
 - What's the intuition here? The effect of quarter of birth may vary by birth year or by state
- It reduced the standard errors, but that comes at a cost of potentially having a weak instruments problem

More instruments

Table VII presents the TSLS and OLS estimates of the new specification for the sample of 40–49 year-old men in the 1980 Census. This is the same sample used in the estimates in Table V. Freeing up the instruments by state of birth and including 50 state-of-birth dummies in the wage equation results in approximately a 40 percent reduction in the standard errors of the TSLS estimates. Furthermore, in the specifications in each of the columns in Table VII, the estimated return to education in the TSLS model is slightly greater than the corresponding TSLS estimate in Table V, whereas in each of the OLS models the return is slightly smaller in Table VII than in Table V. As a consequence, the difference between the TSLS and OLS estimates is of greater significance. For example, the TSLS estimate in column (6) of Table VII is 0.083 with a standard error of 0.010, and the OLS estimate is 0.063 with a standard error of 0.0003: the TSLS estimate is nearly 30 percent greater than the OLS estimate.

More instruments

TABLE VII
OLS AND TSLS ESTIMATES OF THE RETURN TO EDUCATION FOR MEN BORN 1930–1939: 1980 CENSUS^a

Independent variable	(1) OLS	(2) TSLS	(3) OLS	(4) TSLS	(5) OLS	(6) TSLS	(7) OLS	(8) TSLS
Years of education	0.0673 (0.0003)	0.0928 (0.0093)	0.0673 (0.0003)	0.0907 (0.0107)	0.0628 (0.0003)	0.0831 (0.0095)	0.0628 (0.0003)	0.0811 (0.0109)
Race (1 = black)	—	—	—	—	-0.2547 (0.0043)	-0.2333 (0.0109)	-0.2547 (0.0043)	-0.2354 (0.0122)
SMSA (1 = center city)	—	—	—	—	0.1705 (0.0029)	0.1511 (0.0095)	0.1705 (0.0029)	0.1531 (0.0107)
Married (1 = married)	—	—	—	—	0.2487 (0.0032)	0.2435 (0.0040)	0.2487 (0.0032)	0.2441 (0.0042)
9 Year-of-birth dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
8 Region-of-residence dummies	No	No	No	No	Yes	Yes	Yes	Yes
50 State-of-birth dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	—	—	-0.0757 (0.0617)	-0.0880 (0.0624)	—	—	-0.0778 (0.0603)	-0.0876 (0.0609)
Age-squared	—	—	0.0008 (0.0007)	0.0009 (0.0007)	—	—	0.0008 (0.0007)	0.0009 (0.0007)
χ^2 [dof]	—	163 [179]	—	161 [177]	—	164 [179]	—	162 [177]

a. Standard errors are in parentheses. Excluded instruments are 30 quarter-of-birth times year-of-birth dummies and 150 quarter-of-birth times state-of-birth interactions. Age and age-squared are measured in quarters of years. Each equation also includes an intercept term. The sample is the same as in Table VI. Sample size is 329,509.

Weak Instruments

- For a long time, applied empiricists were not attentive to the small sample bias of IV
- But in the early 1990s, a number of papers highlighted that IV can be *severely* biased – in particular, when instruments have only a weak correlation with the endogenous variable of interest and when many instruments are used to instrument for one endogenous variable (i.e., there are many overidentifying restrictions).
- In the worst case, if the instruments are so weak that there is no first stage, then the 2SLS sampling distribution is centered on the probability limit of OLS

Causal model

- Let's consider a model with a single endogenous regressor and a simple constant treatment effect (i.e., "just identified")
- The causal model of interest is:

$$Y = \beta X + \nu$$

Matrices and instruments

- We'll sadly need some matrix notation, but I'll try to make it painless.
- The matrix of instrumental variables is Z with the first stage equation:

$$X = Z'\pi + \eta$$

- And let P_z be the project matrix producing residuals from population regression of X on Z

$$P_z = Z(Z'Z)^{-1}Z'$$

Weak instruments and bias towards OLS

- If ν_i and η_i are correlated, estimating the first equation by OLS would lead to biased results, wherein the OLS bias is:

$$E[\beta_{OLS} - \beta] = \frac{Cov(\nu, X)}{Var(X)}$$

- If ν_i and η_i are correlated the OLS bias is therefore: $\frac{\sigma_{\nu\eta}}{\sigma_\eta^2}$

Deriving the bias of 2SLS

$$\begin{aligned}\widehat{\beta}_{2sls} &= (X'P_zX)^{-1}X'P_ZY \\ &= \beta + (X'P_zX)^{-1}X'PZ_z\nu\end{aligned}$$

substitution of $Y = \beta X + \nu$

2SLS bias

$$\begin{aligned}\hat{\beta}_{2SLS} - \beta &= (X'P_zX)^{-1}X'P_z\nu \\&= aX'P_z\nu \\&= a[\pi'Z' + \eta']P_z\nu \\&= a\pi'Z'\nu + a\eta'P_z\nu \\&= (X'P_ZX)^{-1}\pi'Z'\nu + (X'P_zX)^{-1}\eta'P_z\nu\end{aligned}$$

The bias of 2SLS comes from the non-zero expectation of terms on the right-hand-side even though Z and ν are not correlated.

Taking expectations

- Angrist and Pischke (ch. 4) note that taking expectations of that prior expression is hard because the expectation operator won't pass through $(X'P_zX)^{-1}$.
- However, the expectation of the ratios in the second term can be closely approximated

$$\begin{aligned}\widehat{\beta}_{2sls} - \beta &= (X'P_Z X)^{-1} \pi' Z' \nu + (X'P_z X)^{-1} \eta' P_z \nu \\ E[\widehat{\beta}_{2sls} - \beta] &\approx \left(E[X'P_Z X] \right)^{-1} E[\pi' Z' \nu] + \left(E[X'P_z X] \right)^{-1} E[\eta' P_z \nu]\end{aligned}$$

Approximate bias of 2SLS

We know $E[\pi' Z' \nu] = 0$ and $E[\pi' Z' \eta] = 0$. So letting $E[\eta' P_z \nu] = b$ bc this is hard for me otherwise

$$\begin{aligned} E[\widehat{\beta}_{2SLS} - \beta] &\approx E[X' P_z X)^{-1} b \\ &\approx E[9X' Z(Z' Z)^{-1} Z' X)^{-1} b \\ &\approx E[(\pi Z + \eta)' P_z (\pi Z + \eta)]^{-1} b \\ &\approx \left(E(\pi' Z' Z \pi) + E(\eta' P_z \eta)^{-1} \right) b \\ &\approx \left(E(\pi' Z' Z \pi) + E(\eta' P_z \eta)^{-1} \right) E[\eta' P_z \nu] \end{aligned}$$

That last term is what creates the bias so long as η and ν are correlated – which it's because they are that you picked up 2SLS to begin with

First stage F

With some algebra and manipulation, Angrist and Pischke show that the bias of 2SLS is equal to

$$E[\hat{\beta}_{2SLS} - \beta] \approx \frac{\sigma_{\nu\eta}}{\sigma_{\eta}^2} \left[\frac{E(\pi' Z' Z \pi)/Q}{\sigma_{\eta}^2} + 1 \right]^{-1}$$

where the interior term is the population F-statistic for the joint significance of all regressions in the first stage

Weak instruments and bias towards OLS

- Substituting F for that big term, we can derive the approximate bias of 2SLS as:

$$E[\widehat{\beta}_{2SLS} - \beta] \approx \frac{\sigma_{\nu\eta}}{\sigma_{\eta}^2} \frac{1}{F + 1}$$

- Consider the intuition all that work bought us now: if the first stage is weak (i.e., $F \rightarrow 0$), then the bias of 2SLS approaches $\frac{\sigma_{\nu\eta}}{\sigma_{\eta}^2}$

Weak instruments and bias towards OLS

- This is the same as the OLS bias as for $\pi = 0$ in the second equation on the earlier slide (i.e., there is no first stage relationship) $\sigma_x^2 = \sigma_\eta^2$ and therefore the OLS bias $\frac{\sigma_{\nu\eta}}{\sigma_\eta^2}$ becomes $\frac{\sigma_{\nu\eta}}{\sigma_\eta^2}$.
- But if the first stage is very strong ($F \rightarrow \infty$) then the 2SLS bias is approaching 0.
- Cool thing is – you can test this with an F test on the joint significance of Z in the first stage
- It's absolutely critical therefore that you choose instruments that are strongly correlated with the endogenous regressor, otherwise the cure is worse than the disease

Weak Instruments - Adding More Instruments

- Adding more weak instruments will increase the bias of 2SLS
 - By adding further instruments without predictive power, the first stage F -statistic goes toward zero and the bias increases
 - We will see this more closely when we cover judge fixed effects
- If the model is “just identified” – mean the same number of instrumental variables as there are endogenous covariates – weak instrument bias is less of a problem

Weak instrument problem

- After Angrist and Krueger study, there were new papers highlighting issues related to weak instruments and finite sample bias
- Key papers are Nelson and Startz (1990), Buse (1992), Bekker (1994) and especially Bound, Jaeger and Baker (1995)
- Bound, Jaeger and Baker (1995) highlighted this problem for the Angrist and Krueger study.

Bound, Jaeger and Baker (1995)

Remember, AK present findings from expanding their instruments to include many interactions

- ① Quarter of birth dummies → 3 instruments
- ② Quarter of birth dummies + (quarter of birth) × (year of birth)
+ (quarter of birth) × (state of birth) → 180 instruments

So if any of these are weak, then the approximate bias of 2SLS gets worse

Adding instruments in Angrist and Krueger

	(1) OLS	(2) IV	(3) OLS	(4) IV
Coefficient	.063 (.000)	.142 (.033)	.063 (.000)	.081 (.016)
<i>F</i> (excluded instruments)		13.486		4.747
Partial <i>R</i> ² (excluded instruments, ×100)		.012		.043
<i>F</i> (overidentification)		.932		.775
<i>Age Control Variables</i>				
Age, Age ²	x	x		
9 Year of birth dummies			x	x
<i>Excluded Instruments</i>				
Quarter of birth		x		x
Quarter of birth × year of birth			x	
Number of excluded instruments	3		30	

Adding more weak instruments reduced the first stage *F*-statistic and increases the bias of 2SLS. Notice its also moved closer to OLS.

Adding instruments in Angrist and Krueger

	(1) OLS	(2) IV
Coefficient	.063 (.000)	.083 (.009)
<i>F</i> (excluded instruments)		2.428
Partial <i>R</i> ² (excluded instruments, ×100)		.133
<i>F</i> (overidentification)		.919
<i>Age Control Variables</i>		
Age, Age ²		
9 Year of birth dummies	x	x
<i>Excluded Instruments</i>		
Quarter of birth	x	
Quarter of birth × year of birth	x	
Quarter of birth × state of birth	x	
Number of excluded instruments	180	

More instruments increase precision, but drive down *F*, therefore we know the problem has gotten worse

Guidance on working around weak instruments

- Use a just identified model with your strongest IV
- Use a limited information maximum likelihood estimator (LIML) as it is approximately median unbiased for over identified constant effects models and provides the same asymptotic distribution as 2SLS (under constant effects) with a finite-sample bias reduction.
- Find stronger instruments – easier said than done

Hidden curriculum
Foundational causality stuff
Regression discontinuity designs
Instrumental variables
Two way fixed effects estimator
Differences-in-differences
Comparative case studies
Matching and weighting
Concluding remarks

Intuition
Two stage least squares
Weak instruments
Practical IV Tips
Heterogeneity and the LATE
Lottery designs
Judge fixed effects
Bartik

Look at the reduced form

① Look at the reduced form

- The reduced form is estimated with OLS and is therefore unbiased
- If you can't see the causal relationship of interest in the reduced form, it is probably not there

Report the first stage

- ② Report the first stage (preferably in the same table as your main results)
 - Does it make sense?
 - Do the coefficients have the right magnitude and sign?
 - Please make beautiful IV tables – you'll be celebrated across the land if you do

Report F statistic and OLS

- ③ Report the F -statistic on the excluded instrument(s).
 - Stock, Wright and Yogo (2002) suggest that F -statistics > 10 indicate that you do not have a weak instrument problem – this is not a proof, but more like a rule of thumb
 - If you have more than one endogenous regressor for which you want to instrument, reporting the first stage F -statistic is not enough (because 1 instrument could affect both endogenous variables and the other could have no effect – the model would be under identified). In that case, you want to report the Cragg-Donald EV statistic.
- ④ Report OLS – you said it was biased, but we want to still see it

Table: OLS and 2SLS regressions of Log Earnings on Schooling

Dependent variable	Log wage	
	OLS	2SLS
educ	0.071*** (0.003)	0.124** (0.050)
exper	0.034*** (0.002)	0.056*** (0.020)
black	-0.166*** (0.018)	-0.116** (0.051)
south	-0.132*** (0.015)	-0.113*** (0.023)
married	-0.036*** (0.003)	-0.032*** (0.005)
smsa	0.176*** (0.015)	0.148*** (0.031)
<hr/>		
First Stage Instrument		
College in the county		0.327***
Robust standard error		0.082
F statistic for IV in first stage		15.767
N	3,003	3,003
Mean Dependent Variable	6.262	6.262
Std. Dev. Dependent Variable	0.444	0.444

Standard errors in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Practical Tips for IV Papers

- ⑤ If you have many IVs, pick your best instrument and report the just identified model (weak instrument problem is much less problematic)
- ⑥ Check over identified 2SLS models with LIML

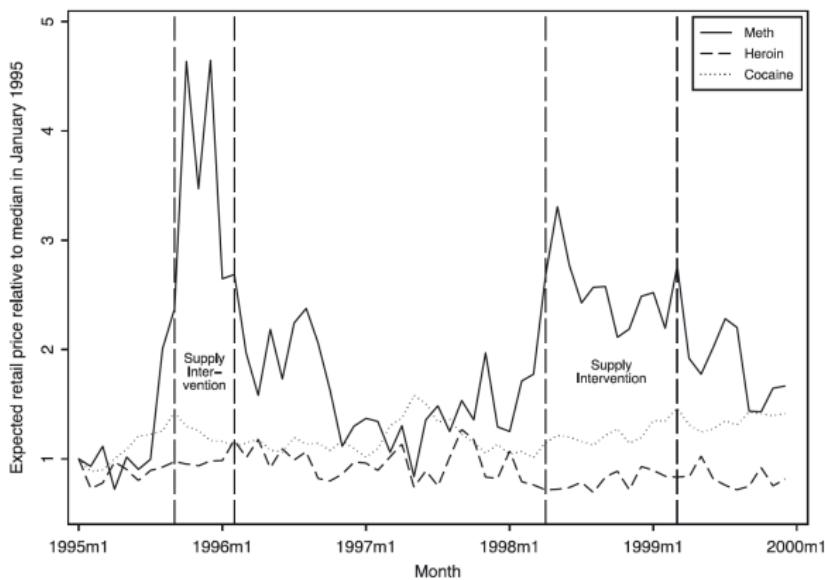
Make beautiful pictures of first stage and reduced form

- ⑦ This cannot be overstated: you must present your main results in beautiful pictures
 - Show pictures of the first stage. Convince the reader something is there. The eyeball is underrated
 - You can't show a second stage with raw data, so instead show pictures of the reduced form.

Visualizing the instrument: supply shocks on meth prices

FIGURE 3

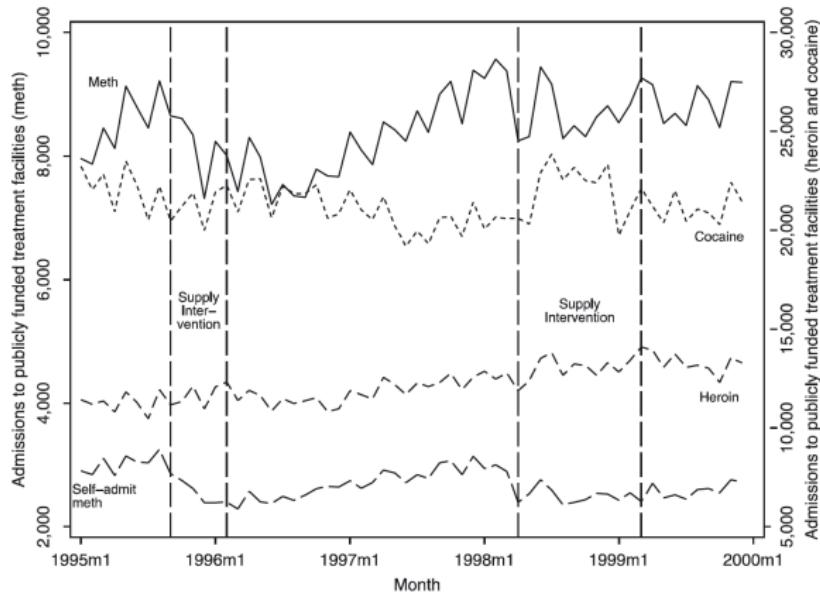
Ratio of Median Monthly Expected Retail Prices of Meth, Heroin, and Cocaine Relative to Their Respective Values in January 1995, STRIDE, 1995–1999



Visualizing the first stage

FIGURE 5

Total Admissions to Publicly Funded Treatment Facilities by Drug and Month, Selected States,
Whites, TEDS, Seasonally Adjusted, 1995–1999

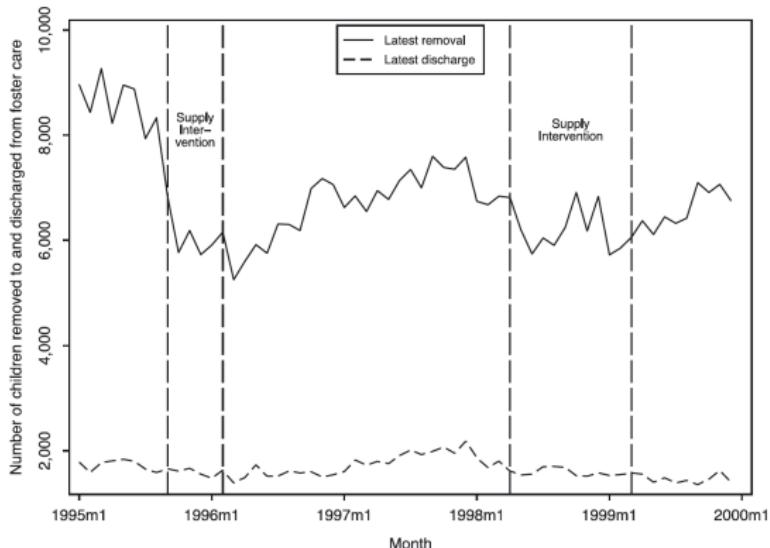


Notes: Authors' calculations from TEDS. Arizona, the District of Columbia, Kentucky, Mississippi, West Virginia, and Wyoming are excluded because of poor data quality. Patients can report the use of more than one drug.

Visualizing the reduced form

FIGURE 4

Number of Children Removed to and Discharged from Foster Care in a Set of Five States by Month, AFCARS, Seasonally Adjusted, 1995–1999



Sources: Authors' calculations from AFCARS. This figure contains AFCARS data only from California, Illinois, Massachusetts, New Jersey, and Vermont. These states form a balanced panel through the entire sample period.

Heterogenous Treatment Effects

- Up to this point, we only considered models where the causal effect was the same for all individuals
 - Constant treatment effects where $Y_i^1 - Y_i^0 = \delta$ for all i units)
- Let's now try to understand what instrumental variables estimation is measuring if treatment effects are *heterogenous*
 - $Y_i^1 - Y_i^0 = \delta_i$ which varies across the population

Why do we care about heterogeneity?

- Heterogeneity, it turns out, makes life interesting and challenging
- There are two issues here:
 - ① We care about internal validity: Does the design successfully uncover causal effects for the population that we are studying?
 - ② We care about external validity: Does the study's results inform us about different populations?
- What parameter did we even estimate using IV when there were heterogenous treatment effects?

Potential outcome notation

“Potential treatment status” (D^j) versus “observed” treatment status (D)

- $D_i^1 = i$ ’s treatment status when $Z_i = 1$
- $D_i^0 = i$ ’s treatment status when $Z_i = 0$

We’ll represent outcomes as a function of both treatment status and instrument status. In other words, $Y_i(D_i = 0, Z_i = 1)$ is represented as $Y_i(0, 1)$

Switching equation

Move from potential treatment status to observed treatment status

$$\begin{aligned} D_i &= D_i^0 + (D_i^1 - D_i^0)Z_i \\ &= \pi_{0i} + \pi_{1i}Z_i + \zeta_i \end{aligned}$$

$$\pi_{0i} = E[D_i^0]$$

$\pi_{1i} = (D_i^1 - D_i^0)$ is the heterogenous causal effect of the IV
on D_i .

$E[\pi_{1i}]$ = The average causal effect of Z_i on D_i

Identifying assumptions under heterogenous treatment effects

- ① Stable Unit Treatment Value Assumption (SUTVA)
- ② Random Assignment
- ③ Exclusion Restriction
- ④ Nonzero First Stage
- ⑤ Monotonicity

Stable Unit Treatment Value Assumption (SUTVA)

Stable Unit Treatment Value Assumption (SUTVA)

If $Z_i = Z'_i$, then $D_i(\mathbf{Z}) = D_i(\mathbf{Z}')$

If $Z_i = Z'_i$ and $D_i = D'_i$, then $Y_i(\mathbf{D}, \mathbf{Z}) = Y_i(\mathbf{D}', \mathbf{Z}')$

- Potential outcomes for each person i are unrelated to the treatment status of other individuals.
- Example: Your instrument is the death of a CEO for hirings. But if a CEO dies, then perhaps other companies lose a CEO as they are hired in the vacant spots.
- In which case, the instrument is related to treatment status of other individuals.

Independence assumption

Independence assumption (e.g., “as good as random assignment”)

$$\{Y_i(D_i^1, 1), Y_i(D_i^0, 0), D_i^1, D_i^0\} \perp\!\!\!\perp Z_i$$

- The IV is independent of the vector of potential outcomes and potential treatment assignments (i.e. “as good as randomly assigned”)
- First two children of the same gender are assigned to families randomly. That is, they are assigned to those with higher likelihood of working just as often as it is assigned to those less likely to work.
- It’s all about the *randomness* of the instrument, in other words, not the instrument’s effect.

Independence

Independence means that the first stage measures the causal effect of Z_i on D_i :

$$\begin{aligned} E[D_i|Z_i = 1] - E[D_i|Z_i = 0] &= E[D_i^1|Z_i = 1] - E[D_i^0|Z_i = 0] \\ &= E[D_i^1 - D_i^0] \end{aligned}$$

Independence

The independence assumption is sufficient for a causal interpretation of the reduced form:

$$\begin{aligned} E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] &= E[Y_i(D_i^1, 1)|Z_i = 1] \\ &\quad - E[Y_i(D_i^0, 0)|Z_i = 0] \\ &= E[Y_i(D_i^1, 1)] - E[Y_i(D_i^0, 0)] \end{aligned}$$

Exclusion Restriction

$$Y(D, Z) = Y(D, Z') \text{ for all } Z, Z', \text{ and for all } D$$

- Any effect of Z on Y must be via the effect of Z on D . In other words, $Y_i(D_i, Z_i)$ is a function of D only. Or formally:

$$Y_i(D_i, 0) = Y_i(D_i, 1) \text{ for } D = 0, 1$$

- Sometimes called the “only through” assumption because you’re assuming the effect of Z on Y is “only through” its effect on D .
- Recall the DAG and the *missing arrows* from Z to u and from Z to Y .

Exclusion restriction

- Use the exclusion restriction to define potential outcomes indexed solely against treatment status:

$$\begin{aligned} Y_i^1 &= Y_i(1, 1) = Y_i(1, 0) \\ Y_i^0 &= Y_i(0, 1) = Y_i(0, 0) \end{aligned}$$

- Rewrite the switching equation:

$$\begin{aligned} Y_i &= Y_i(0, Z_i) + [Y_i(1, Z_i) - Y_i(0, Z_i)]D_i \\ Y_i &= Y_i^0 + [Y_i^1 - Y_i^0]D_i \end{aligned}$$

- Random coefficients notation for this is:

$$Y_i = \alpha_0 + \delta_i D_i$$

with $\alpha_0 = E[Y_i^0]$ and $\delta_i = Y_i^1 - Y_i^0$

Spotting violations of exclusion is a sport

Watch the gears turn:

- We are interested in causal effect of military service on earnings, and so use draft number are instrument for military service.
- Draft number is generated by a random number generator. Therefore independence is met as draft number is independent of potential outcomes and potential treatment status.
- But, people with higher draft numbers evade draft by investing in schooling. Earnings change for reasons other than military service. Exclusion is violated
- In other words, random lottery numbers (independence) do not imply that the exclusion restriction is satisfied

Strong first stage

Nonzero Average Causal Effect of Z on D

$$E[D_i^1 - D_i^0] \neq 0$$

- D^1 means instrument is turned on, and D^0 means it is turned off. We need treatment to change when instrument changes.
- Z has to have some statistically significant effect on the average probability of treatment
- First two children of the same gender makes you more likely to have a third.
- Finally – a testable assumption. We have data on Z and D

Monotonicity

Monotonicity

Either $\pi_{1i} \geq 0$ for all i or $\pi_{1i} \leq 0$ for all $i = 1, \dots, N$

- Recall that π_{1i} is the reduced form causal effect of the instrumental variable on an individual i 's treatment status.
- Monotonicity requires that the instrumental variable (weakly) operate in the same direction on all individual units.
- In other words, while the instrument may have no effect on some people, all those who are affected are affected *in the same direction* (i.e., positively or negatively, but not both).

Monotonicity cont.

- We instrument for schooling with birth quarter. Under monotonicity scenarios 1-2:
 - ① they get more schooling or the same schooling if born in the fourth quarter
 - ② they get less schooling or the same schooling if born in the fourth quarter
- Monotonicity says either of these can be true, but they cannot both be true in your data – yet it's not hard to imagine violations where two people respond differently
- Without monotonicity, IV estimators are not guaranteed to estimate a weighted average of the underlying causal effects of the affected group, $Y_i^1 - Y_i^0$.

Force yourself to think of monotonicity violations

- In the quarter of birth example for schooling, this assumption may not be satisfied (see Barua and Lang 2009).
- Being born in the 4th quarter (which typically increases schooling) may have reduced schooling for some because their school enrollment was held back by their parents

Local average treatment effect

If all 1-5 assumptions are satisfied, then IV estimates the **local average treatment effect (LATE)** of D on Y :

$$\delta_{IV,LATE} = \frac{\text{Effect of } Z \text{ on } Y}{\text{Effect of } Z \text{ on } D}$$

Estimand

Instrumental variables (IV) estimand:

$$\begin{aligned}\delta_{IV,LATE} &= \frac{E[Y_i(D_i^1, 1) - Y_i(D_i^0, 0)]}{E[D_i^1 - D_i^0]} \\ &= E[(Y_i^1 - Y_i^0)|D_i^1 - D_i^0 = 1]\end{aligned}$$

Local Average Treatment Effect

- The LATE parameters is the average causal effect of D on Y for those whose treatment status was changed by the instrument, Z
- For example, IV estimates the average effect of military service on earnings for the subpopulation who enrolled in military service because of the draft but would not have served otherwise.
- LATE does not tell us what the causal effect of military service was for patriots (volunteers) or those who were exempted from military service for medical reasons

LATE cont.

- We have reviewed the properties of IV with heterogenous treatment effects using a very simple dummy endogenous variable, dummy IV, and no additional controls example.
- The intuition of LATE generalizes to most cases where we have continuous endogenous variables and instruments, and additional control variables.

LATE and subpopulations

The instrument partitions any population into 4 distinct groups:

- ① Compliers: The subpopulation with $D_i^1 = 1$ and $D_i^0 = 0$. Their treatment status is affected by the instrument in the “correct direction”.
- ② Always takers: The subpopulation with $D_i^1 = D_i^0 = 1$. They always take the treatment independently of Z .
- ③ Never takers: The subpopulation with $D_i^1 = D_i^0 = 0$. They never take the treatment independently of Z .
- ④ Defiers: The subpopulation with $D_i^1 = 0$ and $D_i^0 = 1$. Their treatment status is affected by the instrument in the “wrong direction”.

Subpopulations of soldiers

Examples of subpopulations:

- ① Compliers: I only enrolled in the military because I was drafted otherwise I wouldn't have served
- ② Always takers: My family have always served, so I serve regardless of whether I am drafted
- ③ Never takers: I'm a contentious objector so under no circumstances will I serve, even if drafted
- ④ Defiers: When I was drafted, I dodged. But had I not been drafted, I would have served. I can't make up my mind.

Never-Takers

$$D_i^1 - D_i^0 = 0$$

$$Y_i(0, 1) - Y_i(0, 0) = 0$$

By **Exclusion Restriction**, causal effect of Z on Y is zero.

Complier

$$D_i^1 - D_i^0 = 1$$

$$Y_i(1, 1) - Y_i(0, 0) = Y_i(1) - Y_i(0)$$

Average Treatment Effect among Compliers

Defier

$$D_i^1 - D_i^0 = -1$$

$$Y_i(0, 1) - Y_i(1, 0) = Y_i(0) - Y_i(1)$$

By **Monotonicity**, no one in this group

Always-taker

$$D_i^1 - D_i^0 = 0$$

$$Y_i(1, 1) - Y_i(1, 0) = 0$$

By **Exclusion Restriction**, causal effect of Z on Y is zero.

Monotonicity Ensures that there are no defiers

- Why is it important to not have defiers?
 - If there were defiers, effects on compliers could be (partly) canceled out by opposite effects on defiers
 - One could then observe a reduced form which is close to zero even though treatment effects are positive for everyone (but the compliers are pushed in one direction by the instrument and the defiers in the other direction)
- Monotonicity assumes there are no defiers

What Does IV (Not) Estimate?

- As said, with all 5 assumptions satisfied, IV estimates the average treatment effect for *compliers*, or LATE
- Without further assumptions (e.g., constant causal effects), LATE is not informative about effects on never-takers or always-takers because the instrument does not affect their treatment status
- So what? Well, it matters because in most applications, we would be mostly interested in estimating the average treatment effect on the whole population:

$$ATE = E[Y_i^1 - Y_i^0]$$

- But that's not possible usually with IV

Sensitivity to assumptions: exclusion restriction

- Someone at risk of draft (low lottery number) changes education plans to retain draft deferments and avoid conscription.
- Increased bias to IV estimand through two channels:
 - Average direct effect of Z on Y for compliers
 - Average direct effect of Z on Y for noncompliers multiplied by odds of being a non-complier
- Severity depends on:
 - Odds of noncompliance (smaller → less bias)
 - “Strength” of instrument (stronger → less bias)
 - Effect of the alternative channel on Y

Sensitivity to assumptions: Monotonicity violations

- Someone who would have volunteered for Army when not at risk of draft (high lottery number) chooses to avoid military service when at risk of being drafted (low lottery number)
- Bias to IV estimand (multiplication of 2 terms):
 - Proportion defiers relative to compliers
 - Difference in average causal effects of D on Y for compliers and defiers
- Severity depends on:
 - Proportion of defiers (small \rightarrow less bias)
 - “Strength” of instrument (stronger \rightarrow less bias)
 - Variation in effect of D on Y (less \rightarrow less bias)

Summarizing

- The potential outcomes framework gives a more subtle interpretation of what IV is measuring
 - In the constant coefficients world, IV measures δ which is “the” causal effect of D_i on Y_i , and assumed to be the same for all i units
 - In the random coefficients world, IV measures instead an average of heterogeneous causal effects across a particular population – $E[\delta_i]$ for some group of i units
 - IV, therefore, measures the *local average treatment effect* or LATE parameter, which is the average of causal effects across the subpopulation of *compliers*, or those units whose covariate of interest, D_i , is influenced by the instrument.

Summarizing

- Under heterogeneous treatment effects, Angrist and Evans (1996) identify the causal effect of the gender composition of the first two kids on labor supply
- This is not the same thing as identifying the causal effect of children on labor supply; the former is a LATE whereas the latter might be better described as an ATE
- *Ex post* this is probably obvious, but like many obvious things, it wasn't obvious until it was worked out. This was a real breakthrough (see Angrist, Imbens and Rubin 1996; Imbens and Angrist 1994)

Hidden curriculum
Foundational causality stuff
Regression discontinuity designs
Instrumental variables
Two way fixed effects estimator
Differences-in-differences
Comparative case studies
Matching and weighting
Concluding remarks

Intuition
Two stage least squares
Weak instruments
Practical IV Tips
Heterogeneity and the LATE
Lottery designs
Judge fixed effects
Bartik

IV in Randomized Trials

- In many randomized trials, participation is nonetheless voluntary among those randomly assigned to treatment
- Consequently, noncompliance is not uncommon and without correcting for it, creates selection biases
- IV designs may even be helpful when evaluating a randomized trial, even though treatment was randomly assigned
- The solution is to instrument for treatment with whether you “won the lottery” and estimate LATE

Lottery designs

- The instrument is your randomized lottery
- Examples might be randomized lottery for attending charter schools to study effect of charter schools on educational outcomes, or a randomized voucher to encourage the collection of health information
- Recall Thornton (2008) instrumented for getting HIV results to estimate causal effect of learning one was HIV+ on condom purchases
- We'll discuss two papers from 2012 and 2014 evaluating a lottery-based expansion of Medicaid health insurance on Oregon on numerous health and financial outcomes

Overarching question

- What are the effects of expanding access to public health insurance for low income adults?
 - Magnitudes, and even the signs, associated with that question were uncertain
- Limited existing evidence
 - Institute of Medicine review of evidence was suggestive, but a lot of uncertainty
 - Observational studies are confounded by selection into health insurance
 - Quasi-experimental work often focuses on elderly and children
 - Only one randomized experiment in a developed country: the RAND health insurance experiment
 - 1970s experiment on a general population
 - Randomized cost-sharing, not coverage itself

The Oregon Health Insurance Experiment

Setting: Oregon Health Plan Standard

- Oregon's Medicaid expansion program for poor adults
- Eligibility
 - Poor (<100% federal poverty line) adults 19-64
 - Not eligible for other programs
 - Uninsured > 6 months
 - Legal residents
- Comprehensive coverage (no dental or vision)
- Minimum cost-sharing
- Similar to other states in payments, management
- Closed to new enrollment in 2004

The Oregon Medicaid Experiment

Oregon held a lottery

- Waiver to operate lottery
- 5-week sign-up period, heavy advertising (January to February 2008)
- Low barriers to sign up, no eligibility pre-screening
- Limited information on list
- Randomly drew 30,000 out of 85,000 on list (March–October 2008)
- Those selected given chance to apply
 - Treatment at household level
 - Had to return application within 45 days
 - 60% applied; 50% of those deemed eligible → 10,000 enrollees

Oregon Health Insurance Experiment

- Evaluate effects of Medicaid using lottery as randomized controlled trial (RCT)
 - Intent-to-treat: Reduced form comparison of outcomes between treatment group (lottery selected individuals) and controls (not selected)
 - LATE: IV using lottery as instrument for insurance coverage
 - First stage: about a 25 percentage point increase in insurance coverage
 - Archived analysis plan
 - Massive data collect effort – primary and secondary
- Similar to ACA expansion but limits to generalizability
 - Partial equilibrium vs. General equilibrium
 - Mandate and external validity
 - Oregon vs. other states
 - Short vs. Long-run

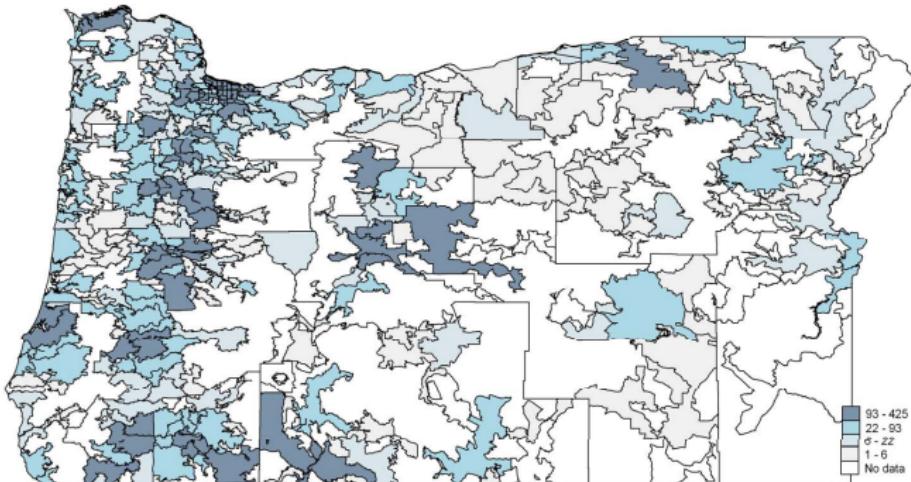
Examine Broad Range of Outcomes

- Costs: Health care utilization
 - Insurance increases resources (income) and lowers price, increasing utilization
 - But improved efficiency (and improved health), decreasing utilization ("offset")
 - Additional uncertainty when comparing Medicaid to no insurance
- Benefits I: Financial risk exposure
 - Insurance supposed to smooth consumption
 - But for very low income, is most care *de jure* or *de facto* free?
- Benefits II: Health
 - Expected to improve (via increased quantity / quality of care)
 - But could discourage health investments ("ex ante moral hazard")

Data

- Pre-randomization demographic information
 - From lottery sign-up
- State administrative records on Medicaid enrollment
 - Primary measure of first stage (i.e., insurance coverage)
- Outcomes
 - Administrative data (~16 months post-notification): Hospital discharge data, mortality, credit reports
 - Mail surveys (~15 months): some questions ask 6-month look-back; some ask current
 - In-person survey and measurements (~25 months): Detailed questionnaires, blood samples, blood pressure, body mass index

Lottery List Distribution Across Zip Codes



Empirical Framework

- They present reduced form estimates of the causal effect of lottery selection

$$Y_{ihj} = \beta_0 + \beta_1 LOTTERY_h + X_{ih}\beta_2 + V_{ih}\beta_3 + \varepsilon_{ihj}$$

- Validity of experimental design: randomization; balance on treatment and control. This is what readers expect

Empirical framework

- They also present IV results because they want to isolate the causal effect of insurance coverage

$$\begin{aligned} INSURANCE_{ihj} &= \delta_0 + \delta_1 LOTTERY_{ih} + X_{ih}\delta_2 + V_{ih}\delta_3 + \mu_{ihj} \\ y_{ihj} &= \pi_0 + \pi_1 \widehat{INSURANCE}_{ih} + X_{ih}\pi_2 + V_{ih}\pi_3 + v_{ihj} \end{aligned}$$

- Effect of lottery on coverage: about 25 percentage points
- We have independence guaranteed; now we need exclusion: the primary pathway of the lottery must be via being on Medicaid
 - Could affect participation in other programs, but actually small
 - "Warm glow" of winning – especially early
- Analysis plan, multiple inference adjustment

Effect of lottery on coverage (first stage)

	Full sample		Credit subsample		Survey respondents	
	Control mean	Estimated FS	Control mean	Estimated FS	Control mean	Estimated FS
Ever on Medicaid	0.141 (0.004)	0.256 (0.004)	0.135 (0.004)	0.255 (0.004)	0.135 (0.007)	0.290 (0.007)
Ever on OHP Standard	0.027 (0.003)	0.264 (0.003)	0.028 (0.004)	0.264 (0.004)	0.026 (0.005)	0.302 (0.005)
# of Months on Medicaid	1.408 (0.045)	3.355 (0.045)	1.352 (0.055)	3.366 (0.055)	1.509 (0.055)	3.943 -0.09
On Medicaid, end of study period	0.106 (0.003)	0.148 (0.003)	0.101 (0.004)	0.151 (0.004)	0.105 (0.006)	0.189 (0.006)
Currently have any insurance (self report)					0.325 (0.008)	0.179 (0.008)
Currently have private ins. (self report)					0.128 (0.005)	-0.008 (0.005)
Currently on Medicaid (self report)					0.117 (0.006)	0.197 (0.006)
Currently on Medicaid					0.093 (0.006)	0.177 (0.006)

Amy Finkelstein, et al. (2012). "The Oregon Health Insurance Experiment: Evidence from the First Year", *Quarterly Journal of Economics*, vol. 127, issue 3, August.

Effects of Medicaid

Use primary and secondary data to gauge 1-year effects

- Mail surveys: 70,000 surveys at baseline, 12 months
- Administrative data
 - Medicaid enrollment records
 - Statewide Hospital discharge data, 2007-2010
 - Credit report data, 2007-2010
 - Mortality data, 2007-2010

Mail survey data

- **Fielding protocol**
 - ~70,000 people, surveyed at baseline and 12 months later
 - Basic protocol: three-stage male survey protocol, English/Spanish
 - Intensive protocol on a 30% subsample included additional tracking, mailings, phone attempts (done to adjust for non-response bias)
- **Response rate**
 - Effective response rate = 50%
 - Non-response bias always possible, but response rate and pre-randomization measures in administrative data were balanced between treatment and control

Administrative data

- Medicaid records
 - Pre-randomization demographics from list
 - Enrollment records to assess “first stage” (how many of the selected got insurance coverage)
- Hospital discharge data
 - Probabilistically matched to list, de-identified at Oregon Health Plan
 - Includes dates and source of admissions, diagnoses, procedures, length of stay, hospital identifier
 - Includes years before and after randomization
- Other data
 - Mortality data from Oregon death records
 - Credit report data, probabilistically matched, de-identified

Sample

- 89,824 unique individuals on the waiting list
- Sample exclusions (based on pre-randomization data only)
 - Ineligible for OHP Standard (out of state address, age, etc.)
 - Individuals with institutional addresses on list
- Final sample: 79,922 individuals out of 66,385 households
 - 29,834 treated individuals (surveyed 29,589)
 - 40,088 control individuals (surveyed 28,816)

Sample characteristics

Variable	Mean	Variable	Mean
Panel A: Full sample			
% Female	0.56	Average Age	41
Panel B: Survey responders only			
<i>Demographics:</i>		<i>Health Status: Ever diagnosed with:</i>	
% White	0.82	Diabetes	0.18
% Black	0.04	Asthma	0.28
% Spanish/Hispanic/Latino	0.12	High Blood Pressure	0.40
% High school or less	0.67	Emphysema or Chronic Bronchitis	0.13
% don't currently work	0.55	Depression	0.56
<i>Determinants of eligibility:</i>			
Average hh income (2008)	13,050	% with any insurance	0.33
% below Federal poverty line	0.68	% with private insurance	0.13

Outcomes

- **Access and use of care**
 - Is access to care improved? Do the insured use more care? Is there a shift in the types of care being used?
 - Mail surveys and hospital discharge data
- **Financial strain**
 - How much does insurance protect against financial strain?
 - What are the out-of-pocket implications?
 - Mail surveys and credit reports
- **Health**
 - What are the short-term impacts on self-reported physical and mental health?
 - Mail surveys and vital statistics (mortality)

Effect of lottery on coverage

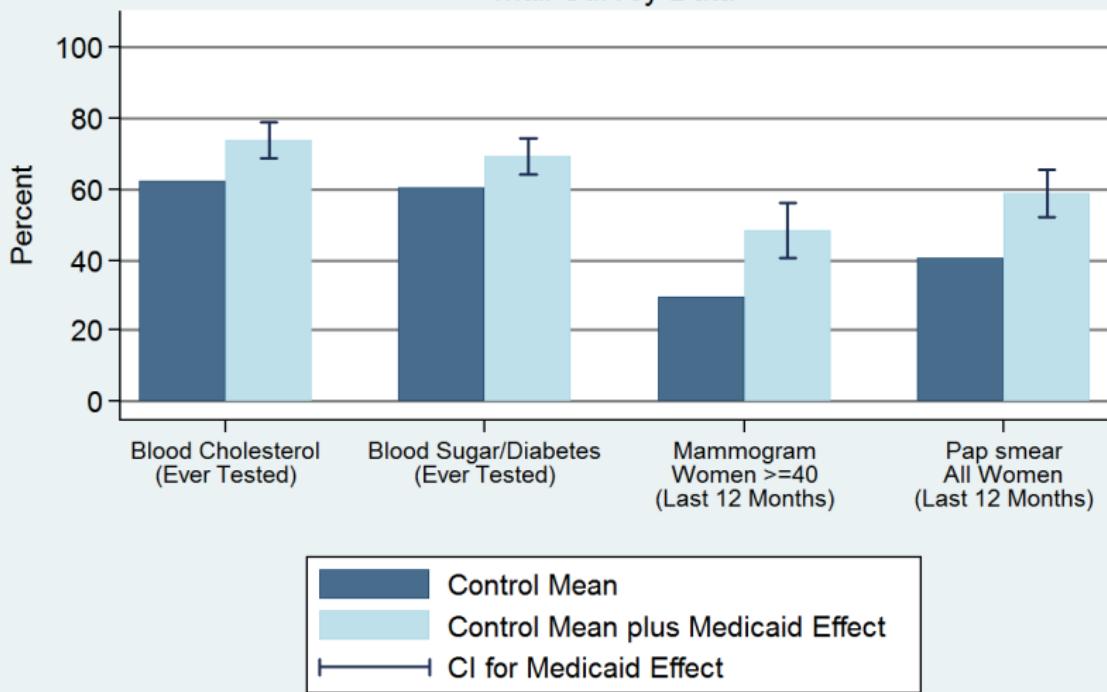
Gaining insurance resulted in better access to care and higher satisfaction with care (conditional on actually getting care)

	CONTROL	RF Model (ITT)	IV Model (LATE)	P-Value
Have a usual place of care	49.9%	+9.9%	+33.9%	.0001
Have a personal doctor	49.0%	+8.1%	+28.0%	.0001
Got all needed health care	68.4%	+6.9%	+23.9%	.0001
Got all needed prescriptions	76.5%	+5.6%	+19.5%	.0001
Satisfied with quality of care	70.8%	+4.3%	+14.2%	.001

SOURCE: Survey data

Preventive Care

Mail Survey Data



Effect of lottery on coverage

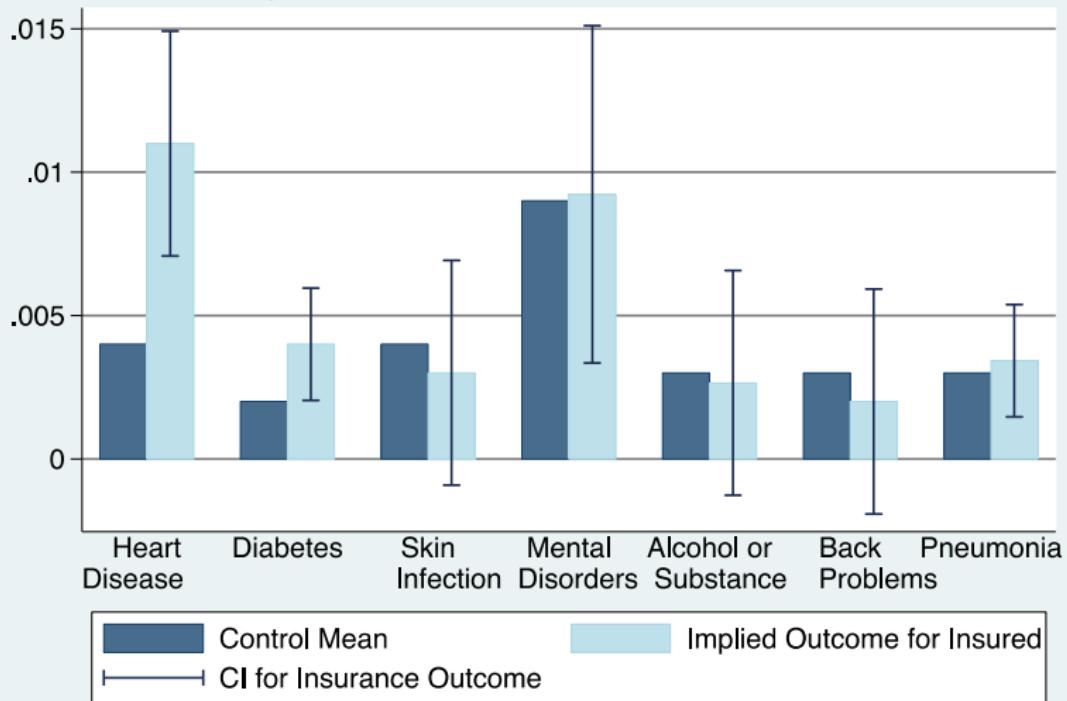
Gaining insurance resulted in increased probability of hospital admissions, primarily driven by non-emergency department admissions

	CONTROL	RF Model (ITT)	IV Model (LATE)	P-Value
Any hospital admission	6.7%	+.50%	+2.1%	.004
--Admits through ED	4.8%	+.2%	+.7%	.265
--Admits NOT through ED	2.9%	+.4%	+1.6%	.002

SOURCE: Hospital Discharge Data

Overall, this represents a 30% higher probability of admission, although admissions are still rare events

Hospital Utilization for Selected Conditions



Summary: Access and use of care

- Overall, utilization and costs went up relative to controls
 - 30% increase in probability of an inpatient admission
 - 35% increase in probability of an outpatient visit
 - 15% increase in probability of taking prescription medications
 - Total \$777 increase in average spending (a 25% increase)
- With this increased spending, those who gained insurance were
 - 35% more likely to get all needed care
 - 25% more likely to get all needed medications
 - Far more likely to follow preventive care guidelines, such as mammograms (60%) and PAP tests (45%)

Results: Financial Strain

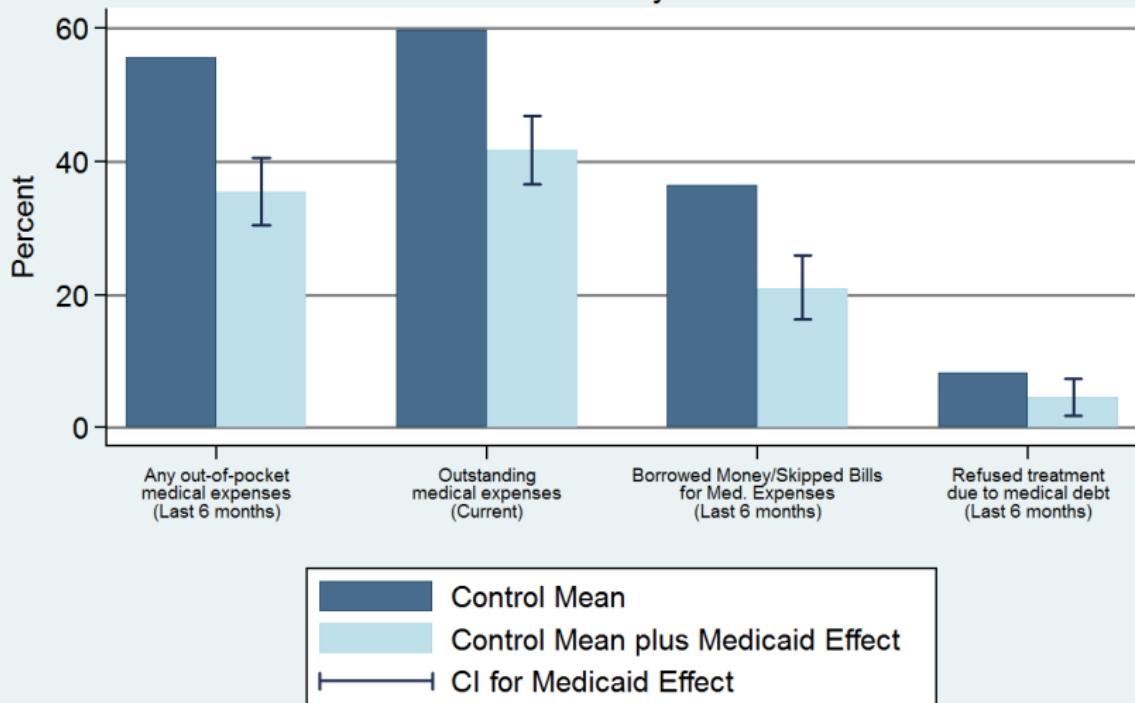
Gaining insurance resulted in a reduced probability of having medical collections in credit reports, and in lower amounts owed

	CONTROL	RF Model (ITT)	IV Model (LATE)	P-Value
Had a bankruptcy	1.4%	+0.2%	+0.9%	.358
Had a collection	50.0%	-1.2%	-4.8%	.013
--Medical collections	28.1%	-1.6%	-6.4%	.0001
--Non-medical collections	39.2%	-0.5	-1.8%	.455
\$ owed medical collections	\$1,999	-\$99	-\$390	.025

Source: Credit report data

Self-reported Financial Strain

Mail Survey Data



Summary: Financial Strain

- Overall, reductions in collections on credit reports were evident
 - 25% decrease in probability of a medical collection
 - Those with a collection owed significantly less
- Household financial strain related to medical costs was mitigated
 - Substantial reduction across all financial strain measures
 - Captures “informal channels” people use to make it work
- Implications for both patients and providers
 - Only 2% of bills sent to collections are ever paid

Results: Self-reported health

Self-reported measures showed significant improvements one year after randomization

	CONTROL	RF Model (ITT)	IV Model (LATE)	P-Value
Health good, v good, excellent	54.8%	+3.9%	+13.3%	.0001
Health stable or improving	71.4%	+3.3%	+11.3%	.0001
Depression screen NEGATIVE	67.1%	+2.3%	+7.8%	.003
CDC Healthy Days (physical)	21.86	+.381	+1.31	.018
CDC Healthy Days (mental)	18.73	+.603	+2.08	.003

Source: Survey data

Summary: Self-reported health

- Overall, big improvements in self-reported physical and mental health
 - 25% increase in probability of good, very good or excellent health
 - 10% decrease in probability of screening for depression
- Physical health measures open to several interpretations
 - Improvements consistent with findings of increased utilization, better access, and improved quality
 - BUT in their baseline surveys, results appeared shortly after coverage ($\sim 2/3$ rds magnitude of full result)
 - May suggest increase in *perception* of well-being rather than physical health
- Biomarker data can shed light on this issue

Discussion

- At 1 year, found increases in utilization, reductions in financial strain, and improvements in self-reported health
 - Medicaid expansion had benefits and costs – didn't "pay for itself"
 - Confirmed biases inherent in observational studies – would have estimated bigger increases in use and smaller improvements in outcomes
- Policy-makers may have different views on value of different aspects of improved well-being
 - "I have an incredible amount of fear because I don't know if the cancer has spread or not."
 - "A lot of times I wanted to rob a bank so I could pay for the medicine I was just so scared . . . People with cancer either have a good chance or no chance. In my case it's hard to recover from lung cancer but it's possible. Insurance took so long to kick in that I didn't think I would get it. Now there is a big bright light shining on me." (Anecdotes)
- Important to have broad evidence on multifaceted effects of Medicaid expansions

Baicker, Katherine, et al. (2014). “The Oregon Experiment – Effects of Medicaid on Clinical Outcomes”, *The New England Journal of Medicine*.

In-person data collection

- Questionnaire and health examination including
 - Survey questions
 - Anthropometric and blood pressure measurement
 - Dried blood spot collection
 - Catalog of all medications
- Fielded between September 2009 and December 2010
 - Average response ~25 months after lottery began
- Limited to Portland area: 20,745 person sample
- 12,229 interviews for effective response rate of 73%

Analytic approach

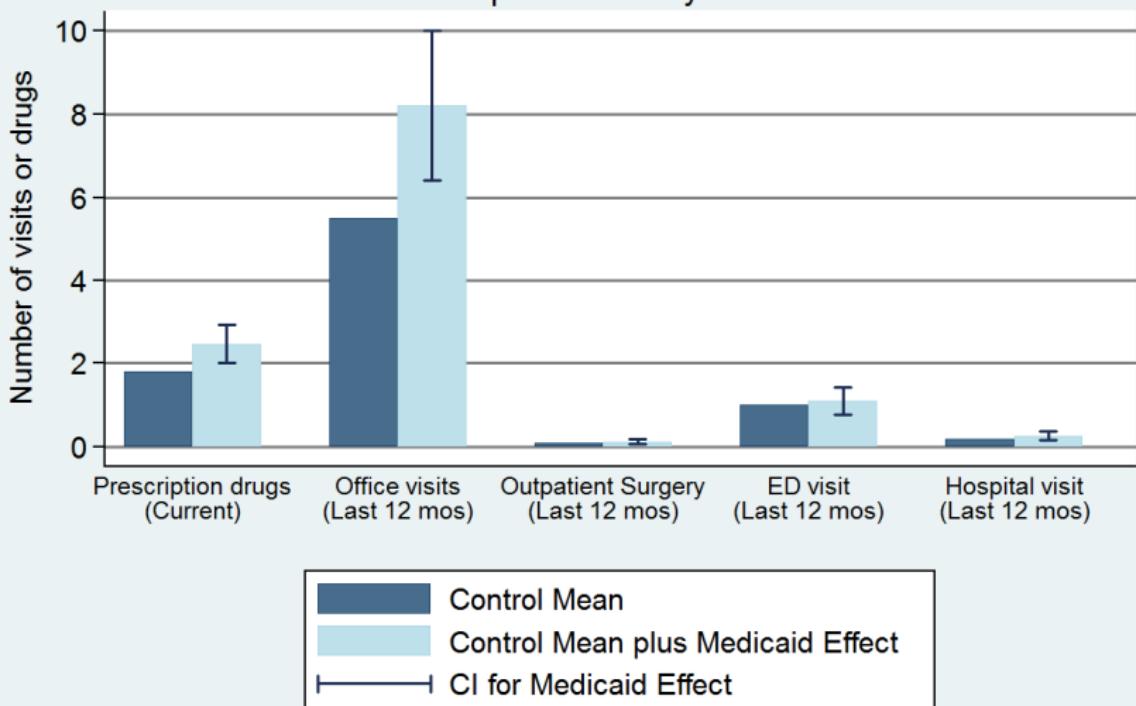
- Intent to treat effect of *lottery selection*
 - Comparing all selected with all not selected
 - Random treatment assignment
 - No differential selection for outcome measurement
- Local average treatment effect on *Medicaid coverage*
 - Using lottery selection as an instrument for coverage
 - ~24 percentage point increase in Medicaid enrollment
 - No change in private insurance (no crowd-out)
 - No effect of lottery except via Medicaid coverage
- Statistical inference is the same for both

Results

- ① *Health care use*
- ② Financial strain
- ③ Clinical health outcomes

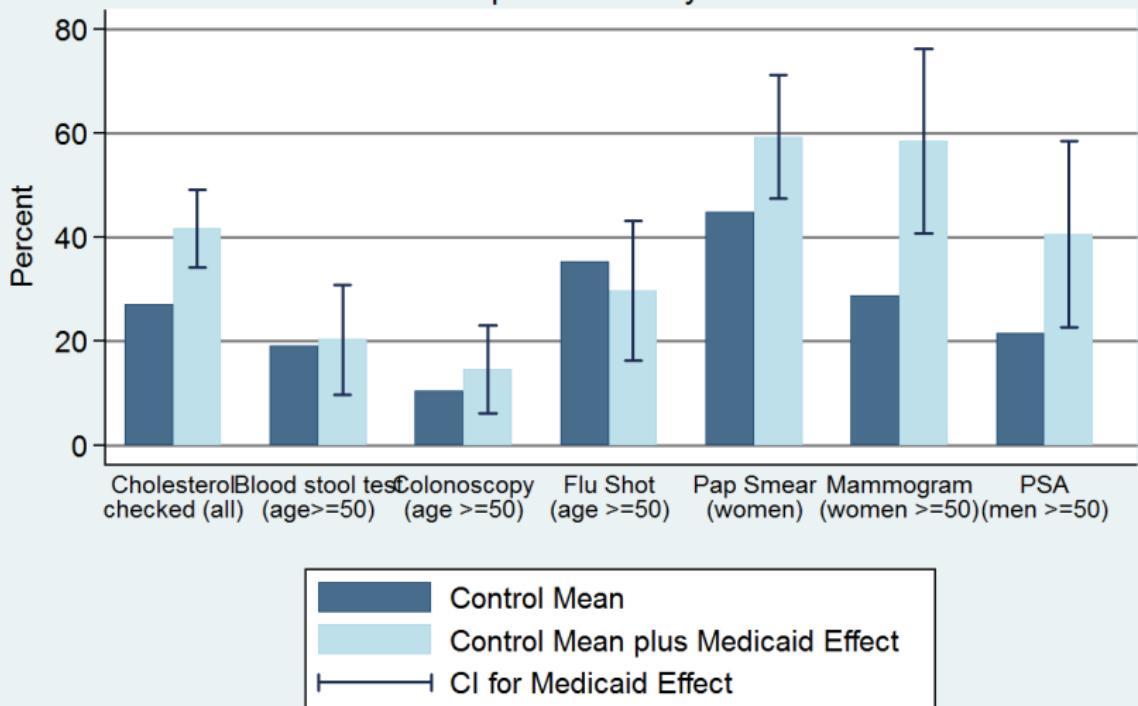
Health Care Utilization

Inperson Survey Data



Preventive Care (Last 12 Months)

Inperson Survey Data



Health care use results

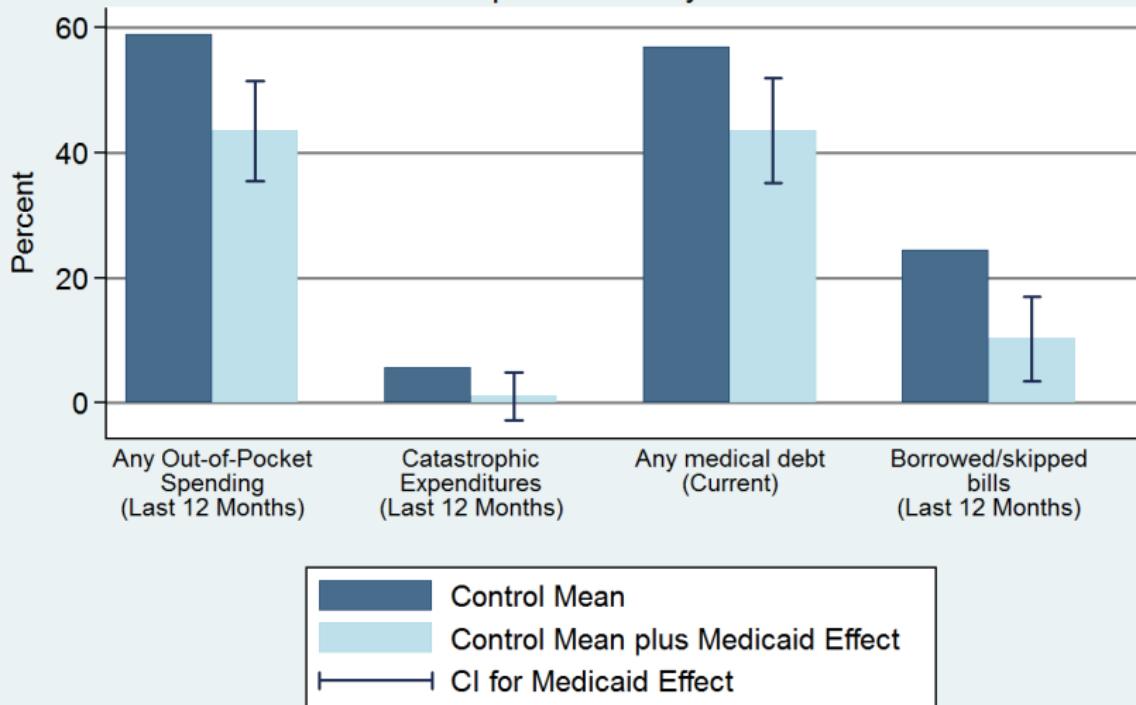
- Increases in use in various settings
 - Increases in probability and number of outpatient visits
 - Increases in probability and number of prescription drugs
 - No discernible change in hospital or ED use (imprecise)
- Increases in preventive care across range of services
- Increases in perceived access and quality
- Implied 35% increase in spending for insured

Results

- ① Health care use
- ② *Financial strain*
- ③ Clinical health outcomes

Financial Hardship

Inperson Survey Data



Financial Hardship Results

- Reduction in strain, out-of-pocket (OOP), money owed
 - Substantial reduction across measures
 - Elimination of catastrophic OOP health spending
- Implications for distribution of burden/benefits
 - Some borne by patients, some by providers
 - Non-financial burden of medical expenses and debt

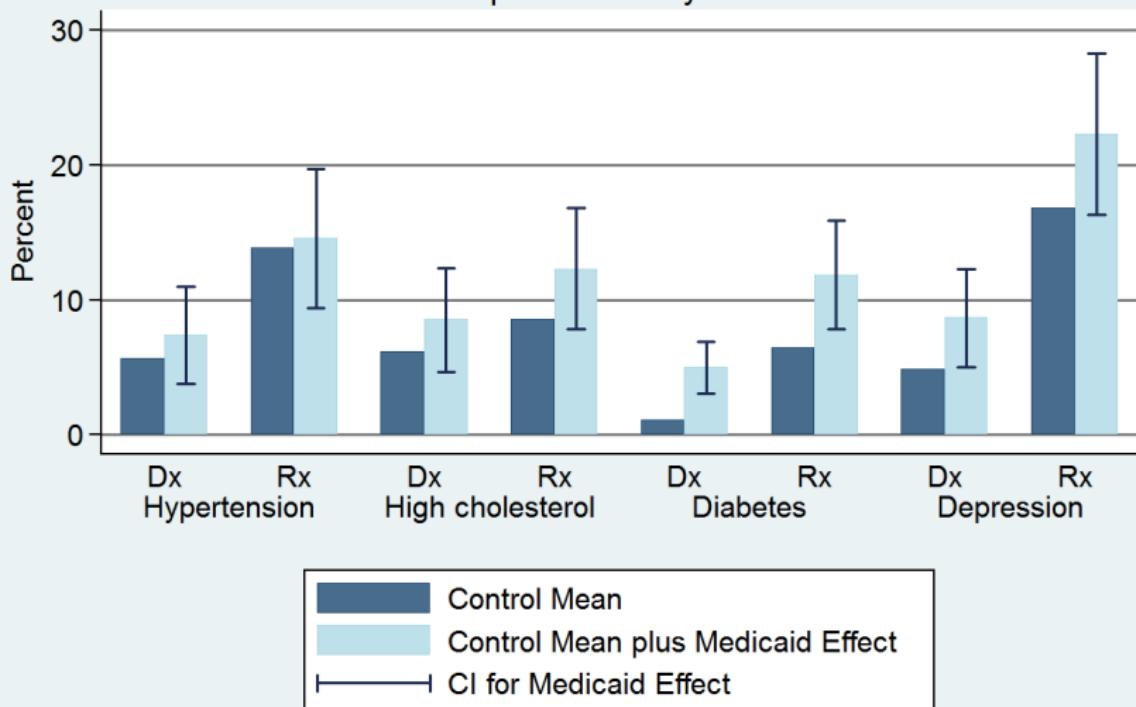
Results

- ① Health care use
- ② Financial strain
- ③ *Clinical health outcomes*

Focusing on specific conditions

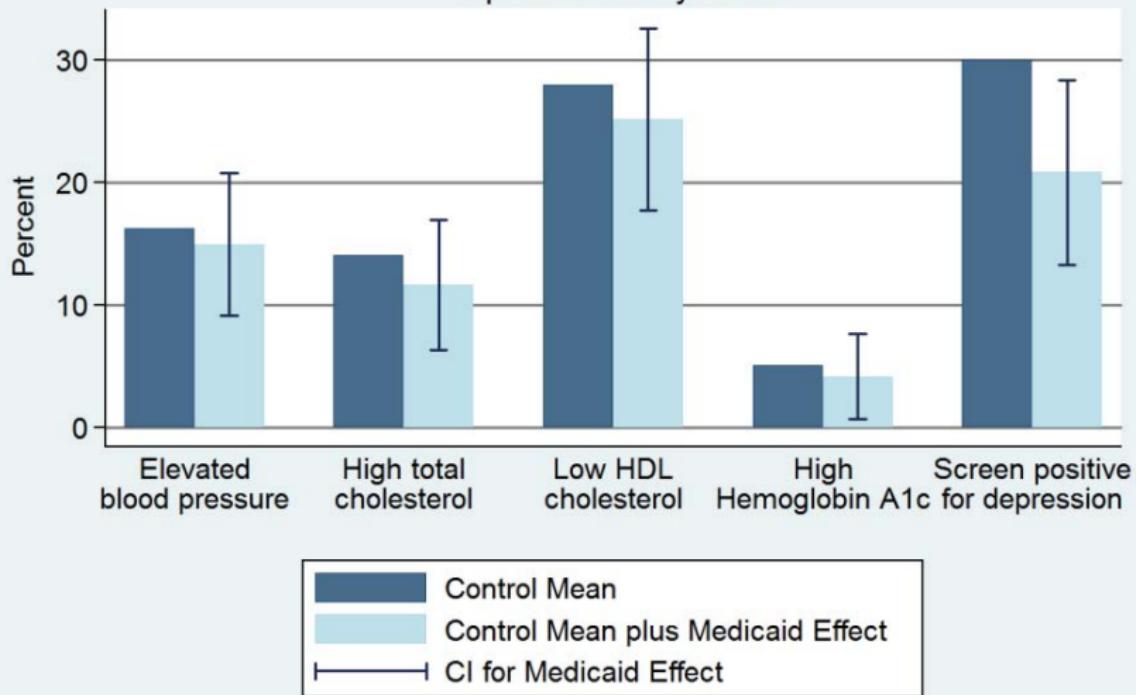
- Measured:
 - Blood pressure
 - Cholesterol levels
 - Glycated hemoglobin
 - Depression
- Reasons for selecting these:
 - Reasonably prevalent conditions
 - Clinically effective medications exist
 - Markers of longer term risk of cardiovascular disease
 - Can be measured by trained interviewers and lab tests
- A limited window into health status

Post-lottery Diagnosis (Dx) and Current Medication (Rx) Inperson Survey Data

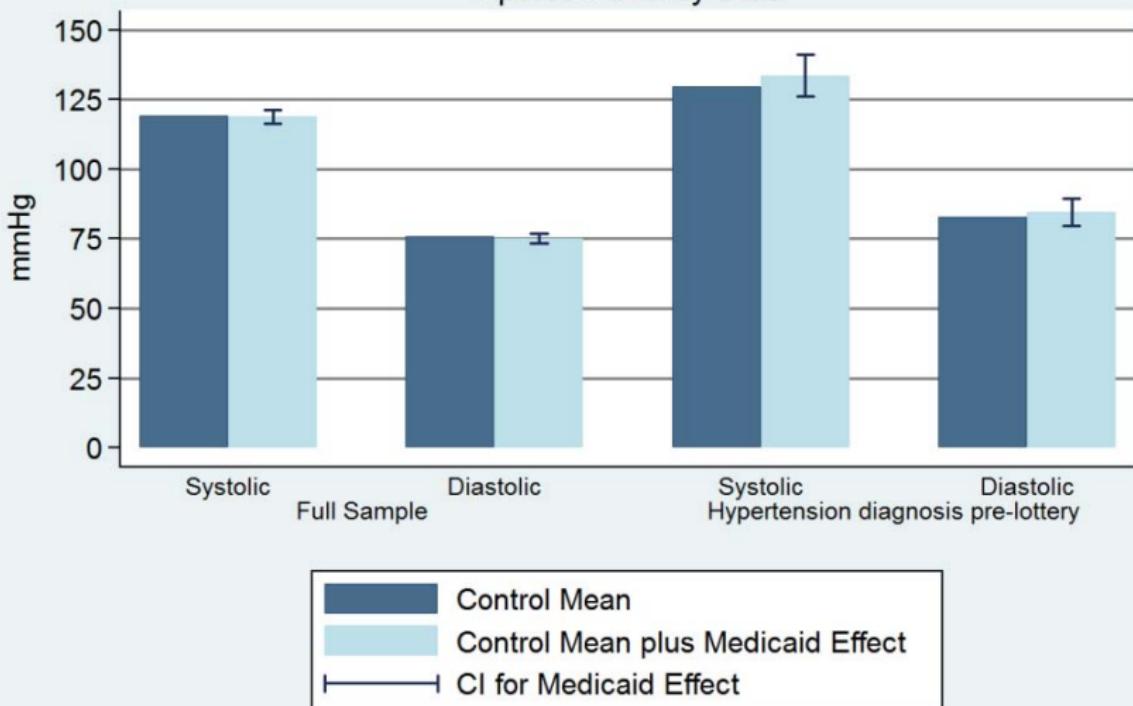


Current Clinical Measures

Inperson Survey Data

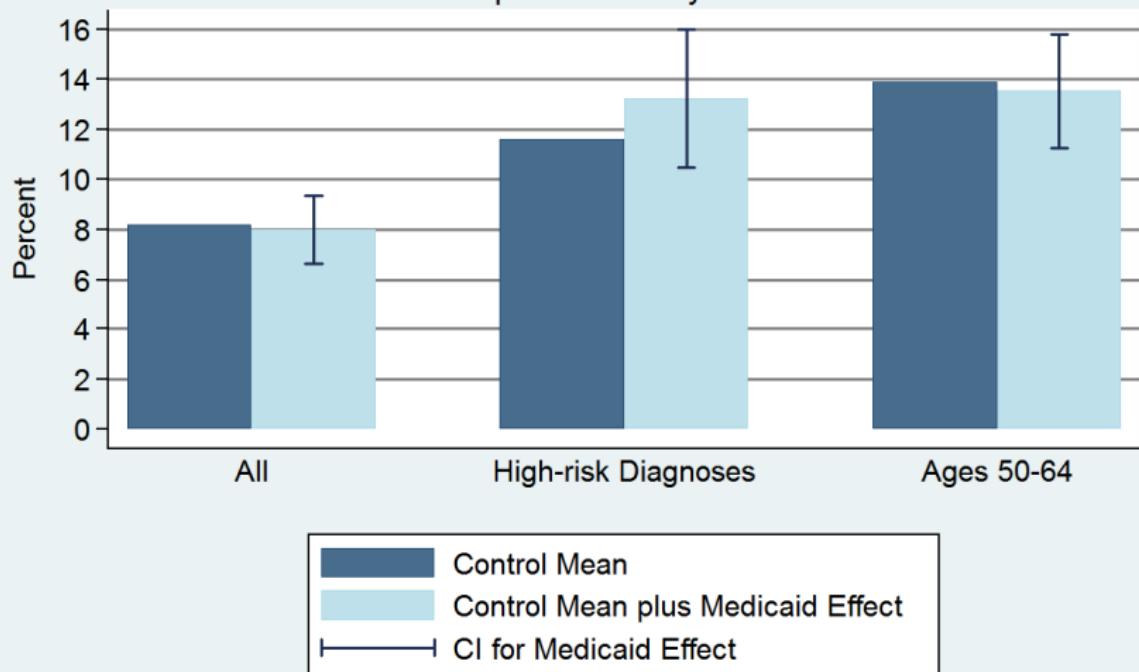


Blood Pressure Inperson Survey Data



Framingham Risk Scores

Inperson Survey Data



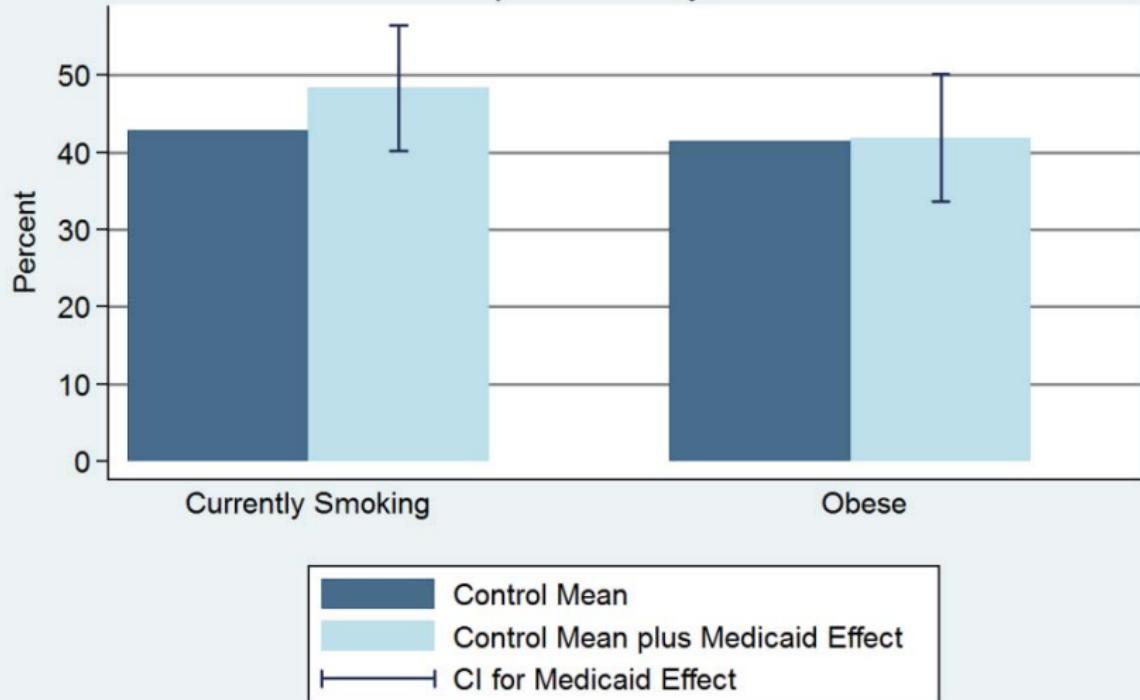
Framingham Risk Score gives the 10 year predicted risk of cardiovascular disease.

Results on specific conditions

- Large reductions in depression
 - Increases in diagnoses and medication
 - In-person estimate of –9 percentage points in being depressed
- Glycated hemoglobin
 - Increases in diagnosis and medication
 - No significant effect on HbA1c; wide confidence intervals
- Blood pressure and cholesterol
 - No significant effects on diagnosis or medication
 - No significant effects on outcomes
- Framingham risk score
 - No significant effect (in general or sub-populations)

Smoking and Obesity

Inperson Survey Data



Summary

- One to two years after expanded access to Medicaid:
 - Increases in health care use and associated costs
 - Increases in compliance with recommended preventive care
 - Improvements in quality and access
 - Reductions in financial strain
 - Improvements in self-reported health
 - Improvements in depression
 - No significant change in specific physical measures
- Sense of the relative magnitude of the effects
 - Use and access, financial benefits, general health, depression
 - Physical measures of specific chronic conditions

Extrapolation to Obamacare (ACA) Expansion

- Context quite relevant for health care reform:
 - States can choose to cover a similar population in planned 2014 Medicaid expansions (up to 138% of federal poverty line)
- But important caveats to bear in mind
 - Oregon and Portland vs. US generally
 - Voluntary enrollment vs. mandate
 - Partial vs. general equilibrium effects
 - Short-run (1-2 years) vs. medium or long run
- We will revisit this again later in the difference-in-differences section when discussing Miller, et al. (2019)

Updating Priors based on Study's Findings

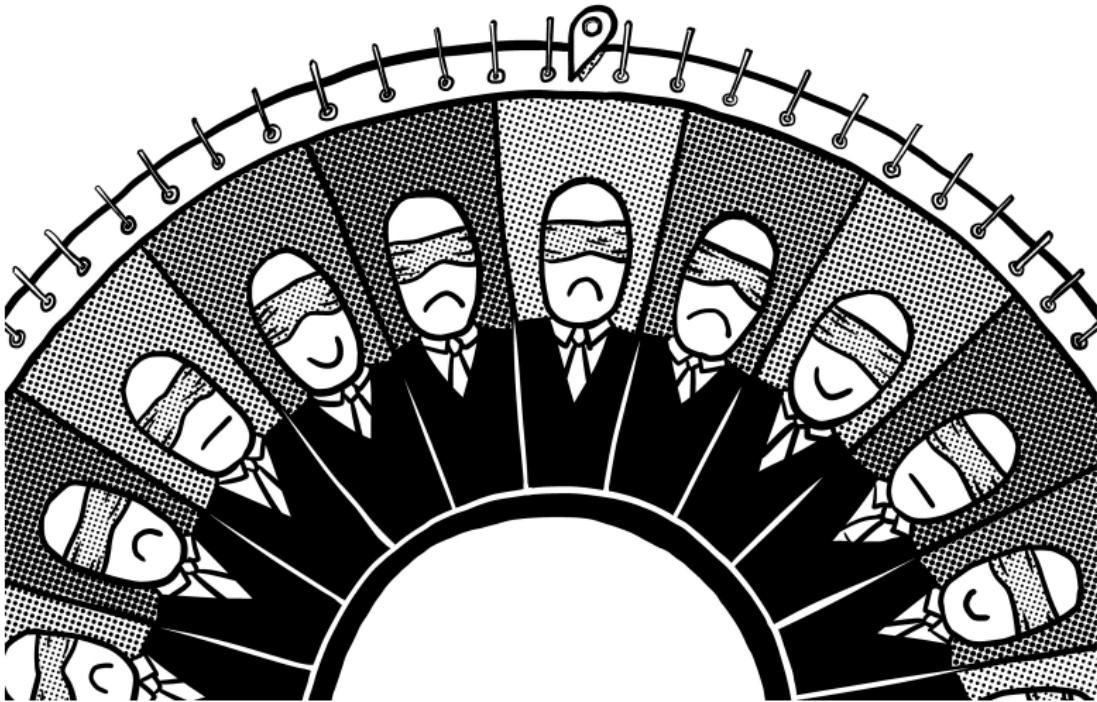
- "Medicaid is worthless or worse than no insurance'"
 - Studies found increases in utilization and perceived access and quality
 - Reductions in financial strain, improvement in self-reported health
 - Improvement in depression
 - Can reject large declines in several physical measures
- "Health insurance expansion saves money"
 - In short run, studies showed increases in utilization and cost and no change in ED use
 - Increases in preventive care, improvements in self-reported health, improvements in depression

Conclusion

- Effects of expanding Medicaid likely to be manifold
 - Hard to establish with observational data and often misleading
- Expanding Medicaid generates both costs and benefits
 - Increased spending
 - Measurably improves *some* aspects of health but not others
 - Important caveats about generalizability
 - Weighing them depends on policy priorities
- Further research on alternative policies needed
 - Many steps in pathway between insurance and outcome
 - Role for innovation in insurance coverage
 - Complements to health care (e.g., social determinants)

Judge fixed effects designs

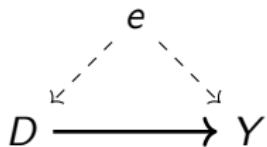
- Imagine the following:
 - ① A person moves through a pipeline and hits a critical point where treatment occurs as a result of some decision-maker
 - ② There are many different decision-makers and you're assigned randomly to one of them
 - ③ Each decision-maker differs in terms of their *leniency* in assigning the treatment
- Very popular in criminal justice bc of how often judges are randomly assigned to defendants (Kling 2006; Mueller-Smith 2015; Dobbie, et al. 2018) or even children to foster care case workers (Doyle 2007; Doyle 2008)



Juvenile incarceration

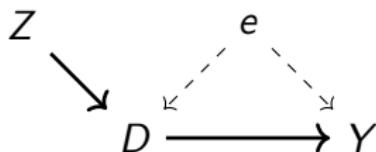
- Aizer and Doyle (2015) were interested in the causal effect of juvenile imprisonment on future crime and human capital accumulation
- Extremely important policy question given the US has the world's highest incarceration rate and prison population of any country in the world by a significant margin (500 prisoners per 100,000, over 2 million adults imprisoned, 4.8 million under supervision)
- High rates of incarceration extend to juveniles: in 2010, the stock of juvenile detainees stood at 70,792, a rate of 2.3 per 1,000 aged 10-19.
- Including supervision, US has a juvenile corrections rate 5x higher than the next highest country, South Africa

Confounding



- We are interested in the causal effect of juvenile incarceration (D) on life outcomes, like adult crime and high school completion
- But youth *choose* to commit crimes, and that choice may be due to unobserved criminogenic factors like poverty or underlying criminal propensities which are themselves causing those future outcomes

Leniency as an instrument



- Aizer and Doyle (2015) propose an instrument - the propensity to convict by the judge the youth is randomly assigned
- If judge assignment is random, and the various assumptions hold, then the IV strategy identifies the local average treatment effect of juvenile incarceration on life outcomes

The Main Idea

- “Plausibly exogenous” variation in juvenile detention stemming from the random assignment of cases to judges who vary in their sentencing
- Consider two juveniles randomly assigned to two different judges with different incarceration tendencies (Scott and Bob)
- Random assignment ensures that differences in incarceration between Scott and Bob are due to the judge, not themselves, because remember, they’re identical

Data

- 35,000 juveniles administrative records over 10 years who came before a juvenile court in Chicago (Juvenile Court of Cook County Delinquency Database)
- Data were linked to public school data for Chicago (Chicago Public Schools) and adult incarceration data for Illinois (Illinois Dept. of Corrections Adult Admissions and Exits)
- They wanted to know the effect of juvenile incarceration on high school completion (2nd data needed) and adult crime (3rd data needed) using randomized judge assignment (1st data needed)
- They need personal identifying information in each data set to make this link (i.e., name, DOB, address)

Preview of findings

- Juvenile incarceration decreased high school graduation by 13 percentage points (vs. 39pp in OLS)
- Increased adult incarceration by 23 percentage points (vs. 41pp in OLS)
- Marginal cases are high risk of adult incarceration and low risk of high school completion as a result of juvenile custody
- Unlikely to ever return to school after incarcerated, but when they do return, they are more likely to be classified as special ed students, and more likely to be classified for special ed services due to behavioral/emotional disorders (as opposed to cognitive disability)

“Plausibly” exogenous

- Very common in these studies for the assignment to some decision-maker to be *arbitrary* but not clearly random (i.e., not random no. generator)
- In this case, juveniles charged with a crime are assigned to a calendar corresponding to their neighborhood and calendars have 1-2 judges who preside over them
- 1/5 of hearings are presided over by judges who cover the calendar when the main judge can't, known as swing judges
- Judge assignment is a function of the sequence with which cases happen to enter into the system and judge availability that is set in advance
- No scope for which judge you see first; conversations with court administrators confirm its random

Structural equation

$$Y_i = \beta_0 + \beta_1 JI_i + \beta_2 X_i + \varepsilon_i$$

where X_i is controls and ε_i is an error term. In this, juvenile incarceration is likely correlated with the error term.

This is the “long” causal model. But note, from the prior DAG, we cannot control for e because it is unobserved. But it is confounding the estimation of juvenile incarceration’s effect on outcomes.

Incarceration Propensity as an Instrument

- The instrument is based on the randomized judge equalling the propensity to incarcerate from the randomly assigned judge
- “Leave-one-out mean”

$$Z_{j(i)} = \left(\frac{1}{n_{j(i)} - 1} \right) \left(\sum_{k \neq i}^{n_{j(i)} - 1} \tilde{J}_{l_k} \right)$$

- The $n_{j(i)}$ terms is the total number of cases seen by judge k , and \tilde{J}_{l_k} is equal to 1 if the juvenile was incarcerated during their first case
- Thus the instrument is the judge's incarceration among first cases based on all their other cases
- It's basically a judge fixed effect given the likelihood two judges have precisely the same propensity is small

Information about the instrument

- There are 62 judges in the data, and the average number of initial cases per judge is 607
- Substantial variation in the data - raw measure ranges from 4% to 21%
- Residualized measure based on controls still has substantial variation from 6% to 18%
- Variation comes from two sources: variation among the regular (nonswing) judges (80% of cases) and variation from the swing judges (20% of cases)

Distribution of IV

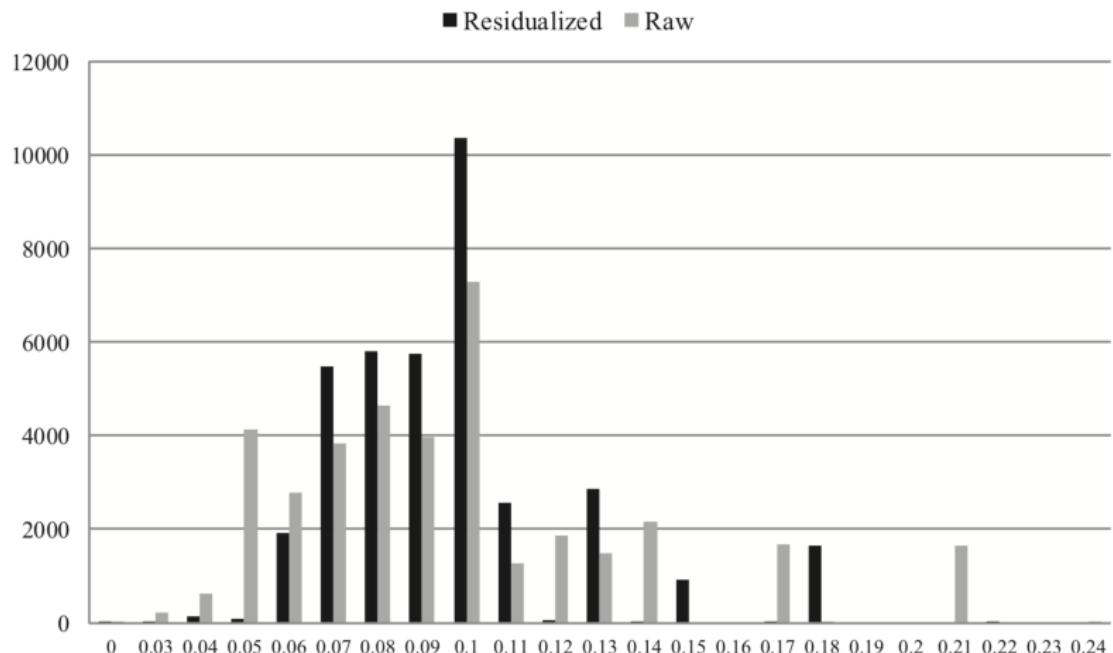


FIGURE I
Distribution of Z: Judge Incarceration Rate

Balance test

TABLE II
INSTRUMENT VERSUS JUVENILE CHARACTERISTICS

	Z distribution			Middle vs.	Top vs.
	Bottom	Middle	Top	bottom	bottom
	tercile	tercile	tercile	p-value	p-value
Z: first judge's leave-out mean incarceration rate in first cases	0.062	0.094	0.147	(.000)	(.000)
Juvenile characteristics					
Male	0.827	0.830	0.833	(.561)	(.311)
African American	0.724	0.737	0.742	(.096)	(.249)
Hispanic	0.189	0.176	0.172	(.061)	(.272)
White	0.078	0.079	0.078	(.833)	(.957)
Other race/ethnicity	0.009	0.008	0.007	(.352)	(.345)
Special education	0.241	0.237	0.252	(.549)	(.130)
U.S. census tract poverty rate	0.264	0.265	0.265	(.572)	(.696)
Age at offense	14.8	14.8	14.8	(.437)	(.434)
P(Juvenile incarceration X)	0.219	0.221	0.220	(.251)	(.516)
Observations	37,692				

First stage

TABLE III
FIRST STAGE

	(1)	(2)	(3)
Dependent variable: juvenile incarcerations		OLS	
First judge's leave-out mean incarceration rate among first cases	1.103 (0.102)	1.082 (0.095)	1.060 (0.097)
Demographic controls	No	Yes	Yes
Court controls	No	No	Yes
Observations	37,692		
Mean of dependent variable	0.227		

High school completion

TABLE IV
JUVENILE INCARCERATION AND HIGH SCHOOL GRADUATION

	Dependent variable: graduated high school						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Full CPS sample				Juvenile court sample		
	OLS	OLS	Inverse propensity score weighting	OLS	OLS	2SLS	2SLS
Juvenile incarceration	-0.389 (0.0066)	-0.292 (0.0065)	-0.391 (0.0055)	-0.088 (0.0043)	-0.073 (0.0041)	-0.108 (0.044)	-0.125 (0.043)
Demographic controls	No	Yes	Yes	No	Yes	No	Yes
Court controls	N/A	N/A	N/A	No	Yes	No	Yes
Observations	440,797	440,797	420,033	37,692			
Mean of dependent variable	0.428	0.428	0.433	0.099			

Adult crime

TABLE V
JUVENILE INCARCERATION AND ADULT CRIME

	Dependent variable: entered adult prison by age 25						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Full CPS sample				Juvenile court sample		
	OLS	OLS	Inverse propensity score weighting	OLS	OLS	2SLS	2SLS
Juvenile incarceration	0.407 (0.0082)	0.350 (0.0064)	0.219 (0.013)	0.200 (0.0072)	0.155 (0.0073)	0.260 (0.073)	0.234 (0.076)
Demographic controls	No	Yes	Yes	No	Yes	No	Yes
Court controls	N/A	N/A	N/A	No	Yes	No	Yes
Observations	440797	440797	420033	37692			
Mean of dependent variable	0.067	0.067	0.057	0.327			

Crime type

TABLE VI
JUVENILE INCARCERATION AND ADULT CRIME TYPE

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: entered adult prison by age 25 for crime type					
	Homicide			Violent		
	OLS	OLS	2SLS	OLS	OLS	2SLS
Juvenile incarceration	0.051 (0.0031)	0.021 (0.0030)	0.035 (0.030)	0.138 (0.0046)	0.061 (0.0050)	0.149 (0.041)
Sample	Full CPS	Juvenile court	Juvenile court	Full CPS	Juvenile court	Juvenile court
Mean of dep. var.: JI = 0	0.008	0.043	0.043	0.024	0.121	0.121
Observations	440,797	37,692	37,692	440,797	37,692	37,692
	Property			Drug		
Juvenile incarceration	0.079 (0.0040)	0.047 (0.0038)	0.142 (0.044)	0.183 (0.011)	0.078 (0.0068)	0.097 (0.052)
Sample	Full CPS	Juvenile Court	Juvenile Court	Full CPS	Juvenile Court	Juvenile Court
Mean of dep. var.	0.013	0.060	0.060	0.034	0.176	0.176
Observations	440,797	37,692	37,692	440,797	37,692	37,692

High school transfers

TABLE VIII
INTERMEDIATE SCHOOLING OUTCOMES: HIGH SCHOOL TRANSFERS

Dependent variable:	(1)	(2)	(3)	(4)	(5)	(6)
	Ever present in CPS school at least 1 year after Initial hearing		Transferred to another CPS high school in years after hearing		Ultimate transfer: adult correctional facility	
	OLS	2SLS	OLS	2SLS	OLS	2SLS
Juvenile incarceration	-0.025 (0.0063)	-0.215 (0.069)	0.055 (0.010)	-0.115 (0.243)	0.127 (0.006)	0.243 (0.060)
Mean of dependent variable	0.666		0.242		0.175	
Observations	37,692		18,195		37,692	

Developing emotional problems

TABLE IX
INTERMEDIATE SCHOOLING OUTCOMES: SPECIAL EDUCATION STATUS

Dependent variable:	Special education type observed in years after initial hearing					
	Any Special Education		Emotional/ behavioral disorder		Learning disability	
	OLS	2SLS	OLS	2SLS	OLS	2SLS
Juvenile incarceration	-0.024 (0.004)	-0.003 (0.037)	0.027 (0.003)	0.133 (0.043)	-0.040 (0.004)	-0.097 (0.039)
Mean of dependent variable	0.193		0.082		0.085	
Observations	29,794					

Concluding remarks

- Sad, but important, paper - the marginal kid shouldn't have been incarcerated
- More generally, leniency designs are very powerful and very common if you know how to look for them
- Bottleneck, influential decision-makers, discretion - these are the three elements of the design

Comments on judge fixed effects

- Leave-one-out average propensity of the decision-maker, or some residualized instrument, is very common
- More often you'll see jackknife IV (JIVE) which drops observations while running regressions to improve finite sample bias
- The biggest threats aren't exclusion probably (though sometimes), but monotonicity
- Might judges be harsh in some situations (violent crimes) but lenient in others (female defendants, first time offenders)

Tests for violations

- New paper by Frandsen, Lefgren and Leslie (2019) proposes a test
- They show that the identifying assumptions imply a conditional expectation of the outcome of interest given the judge assignment is a continuous function of the judge propensity
- They propose a two-part test that generalizes the Sargan-Hansen over identification test and assesses whether treatment effects across judge propensities are possible
- Software available on Emily Leslie's website

Multi-dimensional instrument

- Peter Hull in a cautionary note notes that while combining judge fixed effects into a single propensity is numerically equivalent, it's still a series of dummies
- Therefore it's very important to keep in mind the lessons we learned from weak instruments – the more weak instruments you have when a parameter is overidentified, the larger the bias
- It's ongoing at the moment to think about ways to improve instrument selection, but not settled
- I encourage you to read Peter's note on his website and begin thinking about this yourself

Discussion questions

- When working on a judge fixed effects project, write down an IV DAG
- Whereas monotonicity cannot be visualized to my knowledge on a DAG, exclusion can – so what does an exclusion violation mean in this context?
- Use logic and conversations with those administering the program to answer the following – what does monotonicity mean in this context and how might it be violated?

Empirical exercise

- Let's estimate the effect of cash bail on defendant outcomes using 2SLS and JIVE
- Excellent paper by Megan Stevenson
- -bail.do- and -bail.r- in dropbox and github

Hidden curriculum
Foundational causality stuff
Regression discontinuity designs
Instrumental variables
Two way fixed effects estimator
Differences-in-differences
Comparative case studies
Matching and weighting
Concluding remarks

Intuition
Two stage least squares
Weak instruments
Practical IV Tips
Heterogeneity and the LATE
Lottery designs
Judge fixed effects
Bartik

Bartik instruments

- Bartik instruments are very common in trade, immigration, and labor
- They may be a bit challenging to defend
- But new work by Goldsmith-Pinkham, et al. (2020) and Borusyak, et al. (2019) provide insight into how these work

Bartik instruments

- Perloff (1957) showed industry shares could be used to predict income levels
- Freeman (1980) also used the change in industry composition as an instrument for labor demand
- Timothy Bartik's (1991) study of regional labor markets in which he used the instrument
- Due to Bartik (1991) detailed exposition of the logic of how the national growth shares created variation in labor market demand gets it named after him
- Also called shift-share instruments

Shift-share

"Obvious candidates for instruments are variables shifting MSA labor demand. In this book, only one type of demand shifter is used to form instrumental variables: the share effect from a shift-share analysis of each metropolitan area and year-to-year employment change. A shift-share analysis decomposes MSA growth into three components: a national growth component, which calculates what growth would have occurred if all industries in the MSA had grown at the all-industry national average; a share component, which calculates what extra growth would have occurred if each industry in the MSA had grown at that industry's national average; and a shift component, which calculates the extra growth that occurs because industries grow at different rates locally than they do nationally."

So what is it?

The idea behind a Bartik instrument is to measure the change in a region's labor demand due to changes in the national demand for different industries' products

$$Y_{I,t} = \alpha + \delta I_{I,t} + \rho X_{I,t} + \varepsilon_{I,t}$$

where $Y_{I,t}$ is log wages in location I (e.g., Detroit) in time period t (e.g., 2000) among native workers, $I_{I,t}$ are immigration flows in region I at time period t and $X_{I,t}$ are controls that include region and time fixed effects among other things.

Endogeneity

$$Y_{I,t} = \alpha + \delta I_{I,t} + \rho X_{I,t} + \varepsilon_{I,t}$$

- The parameter δ as elsewhere is some average treatment effect of the immigration flows' effect on native wages.
- The problem is that it is almost certainly the case that immigration flows are highly correlated with the disturbance term such as the time varying characteristics of location I (e.g., changing amenities)

Shares

- Bartik instruments are created by interacting initial “shares” of geographic regions, prior to the immigration flow you’re studying, with national growth rates
- The deviations of a region’s growth from the US national are explained by deviations from national averages
- These deviations are due to shares because the national growth is the same otherwise

Shift-Share Instrument

$$B_{I,t} = \sum_{k=1}^K z_{I,k,t^0} m_{k,t}$$

- The first term is the share variable and the second term is the shift variable.
 - **Share:** z_{I,k,t^0} are the “initial” t^0 share of immigrants from source country k (e.g., Mexico) in location I (e.g., Detroit)
 - **Shift:** $m_{k,t}$ is the change in immigration from country k (e.g., Mexico) into the US as a whole.
- B is the predicted flow of immigrants into destination I (e.g., Detroit) and is a weighted average of the national inflow rates from each country
- Weights depend on the initial distribution of immigrants.

Estimator

- 2SLS: Regress endogenous $I_{l,t}$ (immigration) onto the controls and our Bartik instrument (B)
- Using the fitted values from that regression, \hat{B} , regress $Y_{l,t}$ onto $\hat{I}_{l,t}$
- Under what assumptions does this recover the impact of immigration flows onto log wages?

Two views

- Goldsmith-Pinkham, et al. (2020): the share view
- Borusyak, et al. (2019): the shift view

Exogenous share view

"The Bartik instrument is 'equivalent' to using local industry shares as instruments, and so the exogeneity condition should be interpreted in terms of the shares."

- GMM results showing that the shifts affect the strength of the first stage, but the shares provide the exogenous variation
- If you are exploiting differential exposure to a common shock, then probably the variation comes from shares not shifts
- Strict exogeneity conditional on observables, which means you have to argue why initial shares exogenous to error

Many shifting values

- Bartik is practically challenging to interpret because of the large number of shifting values; US has over 400 industries for instance
- Multiply these over many time periods and the exclusion restriction becomes hard to defend
- Goldsmith-Pinkham, et al. (2020) provide several suggestions for evaluating the central identifying assumption: over-identification tests for instance (but this fails with heterogenous treatment effects)

Rotemberg weights

- Decomposition of the Bartik estimator into a weighted combination of estimates where each share is an instrument
- Weights are called “Rotemberg weights” which sum to one
- Higher valued weights indicate that those instruments are responsible for more of the identifying variation in the design itself
- Now you can focus on the top weights rather than, for instance, 400 shares
- If high weighted areas pass some basic specific tests, then confidence may be more defensible
- <https://github.com/paulgp/shift-share>

Exogenous shift view

"Ultimately, the plausibility of the exogenous shocks framework, as with the alternative framework of Goldsmith-Pinkham, et al. (2020) based on exogenous shares, depends on the shift-share IV application. We encourage practitioners to use shift-share instruments based on an a priori argument supporting the plausibility of either one of these approaches; various diagnostics and tests of the framework that is most suitable for the setting may then be applied. While Borusyak, et al. (2019) develops such procedures for the "shocks" view, Goldsmith-Pinkham, et al. (2020) provide different tools for the "shares" view."

Exogenous shift view

- Exogenous shares are sufficient, but not necessary, for identifying causal effects from Bartik designs
- Temporal shocks may provide exogenous sources of variation
- Borusyak, et al. (2019) show that exogenous independent shocks to many industries allow Bartik designs to identify causal effects regardless of whether the shrars are exogenous
- Shocks must be uncorrelated with the bias of the shares