

UN 3902:  
Economics of Public Policy Seminar  
Week 2: Methods

Michael Carlos Best

January 27 2026

# The Four Questions of Public Finance

- ▶ Public Finance is the study of the role of the government in the economy. It focuses on four key questions:
  1. When should the government intervene in the economy?
  2. How might the government intervene in the economy?
  3. **What are the effects of government interventions on economic outcomes?**
  4. Why do governments choose to intervene in the way that they do?

# Outline

## The Evaluation Problem and Potential Outcomes

### Instrumental Variables

- IV In The Wild

- IV Pitfalls

  - Weak Instruments

  - Heterogeneity: LATE

### Difference in Differences

- DiD and Endogeneity

- DiD in the Wild

### Regression Discontinuity Designs

- Sharp RDD

- Fuzzy RDD

- RDD in the wild

# The Evaluation Problem

- ▶ Microeconometrics = statistical tools to establish
  - ▶ causality
  - ▶ prediction
- ▶ Answer *economic policy* questions such as:
  - ▶ Do job training programs help participants find jobs and earn higher wages?
  - ▶ How much more do people earn as a result of going to college?
  - ▶ Do higher minimum wages increase unemployment?
  - ▶ Do higher taxes make people work less?
- ▶ Using data, but informed by theory

# The Selection Problem: Why This Isn't Trivial

- ▶ Example 1: Do hospitals make people healthier?
- ▶ “Would you say your health in general is excellent (5), very good (4), good (3), fair (2), poor (1)?”

Group	Sample Size	Mean Health Status	Std. Error
Hospital	7,774	3.21	0.014
No Hospital	90,049	3.93	0.003

Source: National Health Interview Survey (NHIS) 2005, via  
Mostly Harmless Econometrics

- ▶ Difference in means is 0.72 ( $t=58.9$ )
- ▶ So people who go to hospital feel *worse*?

# The Selection Problem: Why This Isn't Trivial

- ▶ Example 2: Do higher tax rates make people work less?

1985 MARGINAL TAX RATE	1985 AGI (\$000) (1)	OBSERVATIONS (2)
22	30.7	800
25	36.1	909
28	42.7	713
33	51.5	771
38	67.5	345
42	94.3	152
45	126.9	45
49	177.7	35
50	479.0	22

Source: Feldstein (1995)

- ▶ So higher taxes make people *richer*?

# The Selection Problem: Why This Isn't Trivial

- Example 3: Do more police officers reduce crime rates?

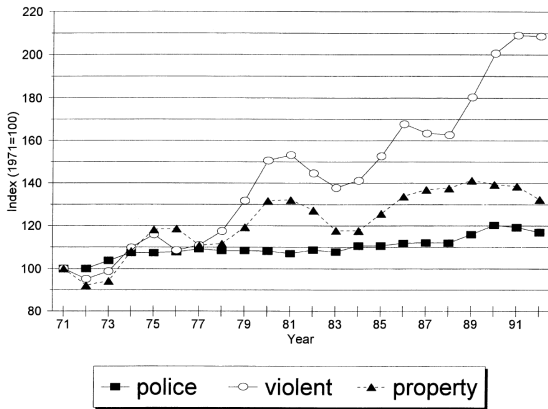


FIGURE 1. TRENDS IN CRIME AND POLICE

Source: Levitt (1997)

- So more police *increase* crime rates?

# Potential Outcomes: A Framework for Thinking About This In

- ▶ Define a binary random variable,  $D_i = \{0, 1\}$ : The **treatment** variable.
- ▶ Outcome of interest  $y_i$  (health status)
- ▶ Does  $D_i$  affect  $y_i$ ?
- ▶ Potential outcome

$$\text{Potential outcome} = \begin{cases} y_{1i} & \text{if } D_i = 1 \\ y_{0i} & \text{if } D_i = 0 \end{cases}$$

- ▶ Hospitals example:
  - ▶  $y_{1i}$  is the health status  $i$  would report if they go to hospital, *irrespective* of whether they actually go
  - ▶  $y_{0i}$  is the health status  $i$  would report if they don't go to hospital, *irrespective* of whether they actually go



# Potential Outcomes: Causal Effects and Observed Outcomes

- ▶ Causal effect of going to hospital on person  $i$  is  $y_{1i} - y_{0i}$
- ▶ Unfortunately, we never observe both  $y_{1i}$  *and*  $y_{0i}$ . Instead we observe

$$\begin{aligned} y_i &= \begin{cases} y_{1i} & \text{if } D_i = 1 \\ y_{0i} & \text{if } D_i = 0 \end{cases} \\ &= y_{0i} + (y_{1i} - y_{0i}) D_i \end{aligned}$$

- ▶ We need to estimate a **counterfactual**: What would  $y_i$  have been if I observed individual  $i$  in the treatment condition she did not in fact experience.
- ▶  $y_{1i} - y_{0i}$  can be very different for different people. But we can learn about averages

## Detour: Expectations

- Recall the **expectation** operator

$$E[y] = \int_{-\infty}^{\infty} Y f_y(Y) dY$$

- and the **conditional expectation** operator

$$\begin{aligned} E[y|x = X] &= \int_{-\infty}^{\infty} Y f_y(Y|x = X) dY \\ &= \int_{-\infty}^{\infty} Y \frac{f_{x,y}(Y, X)}{f_x(X)} dY \end{aligned}$$

# Potential Outcomes: Average Treatment Effects

- ▶ Return to table 1, we observe:
  - ▶  $E[y_i|D_i = 1] = 3.21$
  - ▶  $E[y_i|D_i = 0] = 3.93$
  - ▶  $E[y_i|D_i = 1] - E[y_i|D_i = 0] = -0.72 =$   
 $E[y_{1i}|D_i = 1] - E[y_{0i}|D_i = 1]??$
- ▶ Decompose the observed difference:

$$\underbrace{E[y_i|D_i = 1] - E[y_i|D_i = 0]}_{\text{Observed difference in average health}} = \underbrace{E[y_{1i}|D_i = 1] - E[y_{0i}|D_i = 1]}_{\text{Average treatment effect on the treated}} + \underbrace{E[y_{0i}|D_i = 1] - E[y_{0i}|D_i = 0]}_{\text{Selection bias}}$$

- ▶ We want to estimate  $ATT = E[y_{1i} - y_{0i}|D_i = 1]$  but need to deal with selection bias: Enter microeconometrics

# Randomization and The Selection Problem

- ▶ Randomized Control Trials (RCTs) are becoming increasingly prevalent in applied economics: [[AEA RCT Registry](#)]
- ▶ Randomly assign some units to treatment ( $D_i = 1$ ) and others to “control” ( $D_i = 0$ )
- ▶ Implies that  $D_i$  is *independent* of  $y_{1i}$  and  $y_{0i}$

$$\begin{aligned} E[y_i | D_i = 1] - E[y_i | D_i = 0] &= E[y_{1i} | D_i = 1] - E[y_{0i} | D_i = 0] \\ &= E[y_{1i} | D_i = 1] - E[y_{0i} | D_i = 1] \\ &\quad \text{independence of } D_i, y_{1i} \& y_{0i} \\ &= E[y_{1i} - y_{0i} | D_i = 1] = ATT \end{aligned}$$

the Average Treatment Effect on the Treated (ATT)

- ▶ Moreover, we can simplify further

$$E[y_{1i} - y_{0i} | D_i = 1] = E[y_{1i} - y_{0i}] = ATE$$

the Average Treatment Effect (ATE)

# Outline

The Evaluation Problem and Potential Outcomes

Instrumental Variables

IV In The Wild

IV Pitfalls

Weak Instruments

Heterogeneity: LATE

Difference in Differences

DiD and Endogeneity

DiD in the Wild

Regression Discontinuity Designs

Sharp RDD

Fuzzy RDD

RDD in the wild

# Endogeneity in the potential outcomes framework

- ▶ Return to our potential outcomes framework and generalize it a bit.
- ▶ Assume  $y_{si} = f_i(s)$  where  $y_{si}$  is wages,  $s$  is schooling and  $f_i(\cdot)$  is an individual-specific function that links the two, which we are trying to estimate.
- ▶ A bit more specifically, we would like to estimate  $E[y_{si}|s = X] - E[y_{si}|s = X - 1]$ .
  - ▶ We could think of  $X = 1$  being “go to college” and  $X = 0$  being “do not go to college”
  - ▶ Or,  $s$  could be the number of years of completed schooling, taking on multiple values.

# Endogeneity: Omitted Ability

- ▶ A simple example: Let's assume that

$$f_i(s) = \alpha + \rho s + \eta_i$$

and note that  $\rho$  is the same for everybody (i.e. it's not  $\rho_i$ , we'll relax this later)

- ▶ Let  $\eta_i = \gamma A_i + \nu_i$  where  $A_i$  is (unobserved) ability.
- ▶ Finally, let's assume that  $A_i$  is the only reason  $\eta_i$  and  $s_i$  are correlated:  $E[s_i \nu_i] = 0$
- ▶ That makes the true model

$$y_i = \alpha + \rho s_i + \gamma A_i + \nu_i$$

# Endogeneity: Omitted Ability

- Imagine we just compare people with  $s_i = X$  to people with  $s_i = X - 1$ :

$$\begin{aligned} & \mathbf{E}[y_i | s_i = X] - \mathbf{E}[y_i | s_i = X - 1] \\ &= \mathbf{E}[\alpha + \rho s_i + \gamma A_i + \nu_i | s_i = X] \\ &\quad - \mathbf{E}[\alpha + \rho s_i + \gamma A_i + \nu_i | s_i = X - 1] \\ &= \rho + \gamma \left( \underbrace{\mathbf{E}[A_i | s_i = X] - \mathbf{E}[A_i | s_i = X - 1]}_{\text{probably } \neq 0} \right) \\ &\quad + \underbrace{\mathbf{E}[\nu_i | s_i = X] - \mathbf{E}[\nu_i | s_i = X - 1]}_{=0 \text{ by assumption}} \neq \rho! \end{aligned}$$



# Endogeneity in the potential outcomes framework

- ▶ If we could observe  $A_i$  we'd be fine. We'd regress  $y_i$  on  $s_i$  and  $A_i$  and calculate  $\hat{\rho}$  which has  $\text{plim } \hat{\rho} = \rho$
- ▶ A good instrumental variable(s) allows us to estimate  $\rho$  even without observing  $A_i$
- ▶ A good instrumental variable  $z_i$  requires:
  1. **Relevance:**  $\text{Cov}(s_i, z_i) \neq 0$
  2. **Exclusion Restriction:**  $\text{Cov}(\eta_i, z_i) = 0$ 
    - 2.1 No direct effect:  $z_i$  doesn't belong on the RHS directly
    - 2.2 No indirect effect:  $z_i$  is not correlated with relevant omitted variables

# Endogeneity in the potential outcomes framework

- ▶ With these two assumptions we can write:

$$\begin{aligned}\text{Cov}(y_i, z_i) &= \text{Cov}(\alpha + \rho s_i + \eta_i, z_i) \\ &= \rho \underbrace{\text{Cov}(s_i, z_i)}_{\neq 0 \text{ (relevance)}} + \underbrace{\text{Cov}(\eta_i, z_i)}_{=0 \text{ (exclusion)}}\end{aligned}$$

- ▶ So we have identified  $\rho$ !

$$\rho = \frac{\text{Cov}(y_i, z_i)}{\text{Cov}(s_i, z_i)} = \frac{\text{Cov}(y_i, z_i) / \text{Var}(z_i)}{\text{Cov}(s_i, z_i) / \text{Var}(z_i)}$$

- ▶ The coefficient  $\rho$  is the ratio of the population coefficient in a regression of  $y_i$  on  $z_i$  (we call this the *reduced form*) to the population coefficient in a regression of  $s_i$  on  $z_i$  (we call this the *first stage*)

# Outline

## The Evaluation Problem and Potential Outcomes

## Instrumental Variables

### IV In The Wild

### IV Pitfalls

Weak Instruments

Heterogeneity: LATE

## Difference in Differences

DiD and Endogeneity

DiD in the Wild

## Regression Discontinuity Designs

Sharp RDD

Fuzzy RDD

RDD in the wild

## IV In The Wild

- ▶ Let's go through some examples of prominent papers that have used an Instrumental Variables strategy to identify relationships of economic interest.
- ▶ We'll look at:
  1. Steven D. Levitt - "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime" *American Economic Review* (1997)
  2. Daron Acemoglu, Simon Johnson & James A. Robinson - "The Colonial Origins of Comparative Development: An Empirical Investigation" *American Economic Review* (2001)
  3. Caroline M. Hoxby - "Does Competition among Public Schools Benefit Students and Taxpayers?" *American Economic Review* (2000)

## IV example 1: Do more police reduce crime?

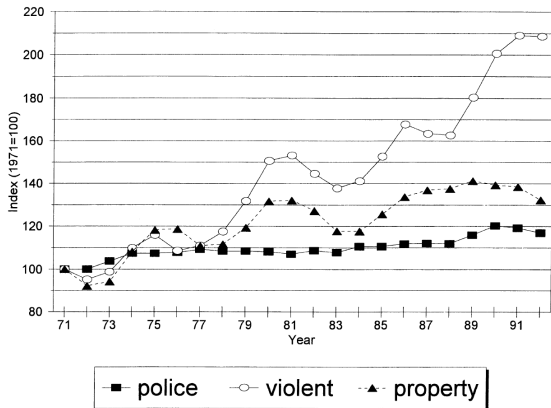


FIGURE 1. TRENDS IN CRIME AND POLICE

Source: Levitt (1997)

- So more police *increase* crime rates?

## IV example 1: Do more police reduce crime?

- ▶ Levitt (1997) studies this.
- ▶ Why might we have an endogeneity problem?
- ▶ Levitt argues he can use the timing of mayoral and gubernatorial elections as an instrument for police hiring.
- ▶ Specifically, his  $z_i$  if a dummy = 1 if a year is an election year in that city.
- ▶ What are the requirements for this to be a valid instrument?

## IV example 1: Do more police reduce crime?

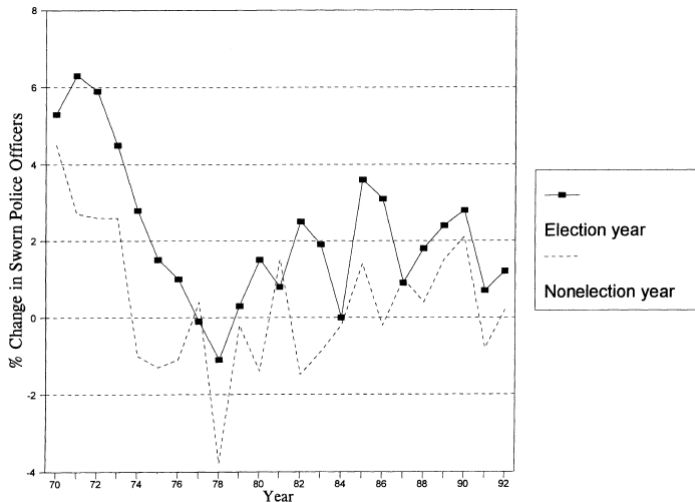


FIGURE 2. YEARLY CHANGES IN SWORN POLICE (ELECTION YEARS VERSUS NONELECTION YEARS)

## IV example 1: Do more police reduce crime?

- ▶ The first stage looks good in a picture. In an equation:

$$\Delta \ln(P_{it}) = \theta_1 M_{it} + \theta_2 G_{it} + \mathbf{X}_{it}\delta + \gamma_t + \lambda_i + \nu_{it}$$

- ▶  $P_{it}$  = # of sworn police officers;  $M_{it}$  is a dummy for a mayoral election;  $G_{it}$  is a dummy for a gubernatorial election;  $\mathbf{X}_{it}$  are covariates.
- ▶ The reduced form is then:

$$\Delta \ln(C_{ijt}) = \Psi_1 M_{it} + \Psi_2 G_{it} + \mathbf{X}_{it}\kappa_j + \gamma_{tj} + \lambda_i + \nu_{ijt}$$

- ▶ where  $j$  indexes different crimes



# IV example 1: Do more police reduce crime?

TABLE 2—THE ELECTION CYCLE AS A PREDICTOR OF CHANGES IN THE POLICE FORCE

Variable	(1) $\Delta \ln$ Sworn officers	(2) $\Delta \ln$ Sworn officers	(3) $\Delta \ln$ Sworn officers	(4) Violent crime	(5) Property crime
Mayoral election year	0.013 (0.004)	0.011 (0.004)	0.012 (0.004)	-0.008 (0.004)	-0.005 (0.003)
Gubernatorial election year	0.025 (0.007)	0.024 (0.007)	0.024 (0.007)	-0.006 (0.006)	-0.009 (0.005)
$\Delta \ln$ Public welfare spending per capita	—	-0.012 (0.009)	-0.013 (0.009)	-0.010 (0.013)	0.008 (0.011)
$\Delta \ln$ Education spending per capita	—	0.088 (0.044)	0.094 (0.045)	0.054 (0.043)	-0.022 (0.035)
$\Delta$ State unemployment rate	—	-0.323 (0.256)	-0.319 (0.258)	-0.286 (0.224)	0.645 (0.182)
$\Delta$ (Percent ages 15–24 in SMSA)	—	2.41 (2.24)	4.69 (2.56)	-4.03 (2.76)	4.24 (2.24)
$\Delta$ (Percent black)	—	0.001 (0.007)	-0.007 (0.010)	-0.018 (0.009)	-0.012 (0.007)
$\Delta$ (Percent female-headed households)	—	-0.003 (0.014)	0.012 (0.019)	-0.002 (0.018)	0.013 (0.014)
Year indicators?	Yes	Yes	Yes	Yes	Yes
City-size indicators?	No	Yes	Yes	Yes	Yes
City-fixed effects?	No	No	Yes	Yes	Yes
P-value: Joint significance of election years?	<0.001	<0.001	<0.01	0.082	0.062
R <sup>2</sup>	0.06	0.09	0.11	0.21	0.37

Notes: Dependent variable in columns (1)–(3) is  $\Delta \ln$  sworn police officers per capita. The dependent variable in columns (4) and (5) is the average  $\Delta \ln$  in the number of crimes per capita over the current and following year. All four violent crime categories are stacked in estimating column (4); all three property crime categories are stacked in estimating column (5). Sample includes 59 large U.S. cities with directly elected mayors, 1970–1992. Number of observations is 1,276 in columns (1)–(3), 4,801 in column (4), and 3,606 in column (5). Columns (4) and (5) allow for heteroskedasticity across crime categories. Year dummies included in all regressions. Three city-size indicators are included in columns (2)–(5); city-fixed effects are included in columns (3)–(5).

## IV example 1: Do more police reduce crime?

- So, Levitt will use  $M_{it}$  and  $G_{it}$  as instruments in the estimating equation

$$\Delta \ln (C_{ijt}) = \beta_{1j} \Delta \ln (P_{ijt}) + \beta_{2j} \Delta \ln (P_{ijt-1}) + \mathbf{X}_{it} \eta_j + \gamma_{tj} + \lambda_i + \varepsilon_{ijt}$$

- Let's look at the results

# IV example 1: Do more police reduce crime?

TABLE 3—ESTIMATES OF THE ELASTICITY OF VIOLENT CRIME RATES WITH RESPECT TO SWORN POLICE OFFICERS

Variable	(1) OLS	(2) OLS	(3) 2SLS	(4) 2SLS	(5) 2SLS	(6) LIML
ln Sworn officers per capita	0.28 (0.05)	-0.27 (0.06)	-1.39 (0.55)	-0.90 (0.40)	-0.65 (0.25)	-1.16 (0.38)
State unemployment rate	-0.65 (0.40)	-0.25 (0.31)	-0.00 (0.36)	-0.19 (0.33)	-0.13 (0.32)	-0.02 (0.33)
ln Public welfare spending per capita	-0.03 (0.02)	-0.03 (0.02)	-0.03 (0.02)	-0.03 (0.02)	-0.02 (0.02)	-0.03 (0.02)
ln Education spending per capita	0.04 (0.07)	0.06 (0.06)	0.02 (0.07)	0.03 (0.07)	0.05 (0.06)	0.03 (0.06)
Percent ages 15–24 in SMSA	1.43 (1.00)	-2.61 (3.71)	-1.47 (4.12)	-2.55 (3.88)	-2.02 (3.76)	-1.50 (3.86)
Percent black	0.010 (0.003)	-0.017 (0.011)	-0.034 (0.015)	-0.025 (0.013)	-0.022 (0.012)	-0.031 (0.013)
Percent female-headed households	0.003 (0.006)	0.007 (0.023)	0.040 (0.030)	0.023 (0.027)	0.018 (0.025)	0.033 (0.027)
Data differenced?	No	Yes	Yes	Yes	Yes	Yes
Instruments:	None	None	Elections	Election * city-size interactions	Election * region interactions	Election * region interactions
P-value of cross-crime restriction on police elasticity	<0.01	<0.01	0.09	0.13	0.33	0.28

# IV example 1: Do more police reduce crime?

TABLE 5—CRIME-SPECIFIC ESTIMATES OF THE EFFECT OF CHANGES IN SWORN OFFICERS

	Murder	Rape	Robbery	Assault	Burglary	Larceny	Motor vehicle theft
OLS (levels)	0.27 (0.06)	-0.07 (0.05)	0.64 (0.05)	0.34 (0.06)	0.08 (0.05)	0.14 (0.05)	0.38 (0.06)
OLS (differences)	-0.60 (0.19)	-0.06 (0.13)	-0.31 (0.10)	0.11 (0.13)	-0.25 (0.08)	-0.10 (0.06)	-0.29 (0.10)
2SLS (elections as instruments)	-3.05 (0.91)	0.67 (1.22)	-1.20 (1.31)	-0.82 (1.20)	-0.58 (1.55)	0.26 (1.66)	-0.61 (1.31)
2SLS (election*city-size interactions as instruments)	-2.09 (0.64)	0.08 (0.84)	-0.38 (0.89)	-0.36 (0.81)	-0.39 (1.06)	0.06 (1.20)	0.14 (0.89)
2SLS (election*region interactions as instruments)	-1.18 (0.39)	-0.11 (0.49)	-0.49 (0.53)	-0.41 (0.50)	-0.11 (0.62)	-0.21 (0.67)	-0.34 (0.53)
LIML (election*region interactions as instruments)	-1.98 (0.59)	-0.27 (0.77)	-0.79 (0.79)	-1.09 (0.73)	-0.05 (0.90)	-0.43 (1.01)	-0.50 (0.80)

## IV example 2: Do better institutions cause growth?

- ▶ One of the biggest questions in economics: Why are some countries rich while others are poor?
- ▶ Can better “institutions” (property rights, less distortionary govt policy) explain these differences?
- ▶ Tons of endogeneity problems!
- ▶ A clever instrument from history:

(potential) settler  
mortality  $\Rightarrow$  settlements

$\Rightarrow$  early institutions  $\Rightarrow$  current institutions

$\Rightarrow$  current performance.

## IV example 2: Do better institutions cause growth?

- ▶ Not even clear the reduced form relationship will be there (in which case very unlikely 2sls will work)

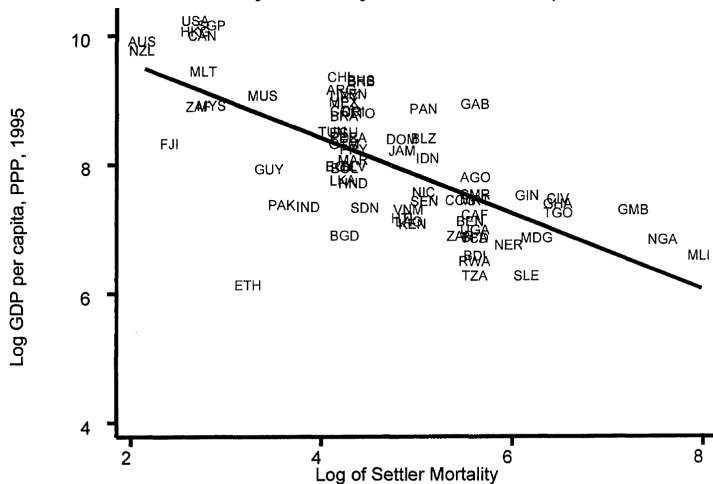


FIGURE 1. REDUCED-FORM RELATIONSHIP BETWEEN INCOME AND SETTLER MORTALITY

## IV example 2: Do better institutions cause growth?

### ► The benchmark: OLS

$$\log(y_i) = \mu + \alpha R_i + \mathbf{X}_i\gamma + \varepsilon_i$$

TABLE 2—OLS REGRESSIONS

	Whole world (1)	Base sample (2)	Whole world (3)	Whole world (4)	Base sample (5)	Base sample (6)	Whole world (7)	Base sample (8)
	Dependent variable is log GDP per capita in 1995						Dependent variable is log output per worker in 1988	
Average protection against expropriation risk, 1985–1995	0.54 (0.04)	0.52 (0.06)	0.47 (0.06)	0.43 (0.05)	0.47 (0.06)	0.41 (0.06)	0.45 (0.04)	0.46 (0.06)
Latitude			0.89 (0.49)	0.37 (0.51)	1.60 (0.70)	0.92 (0.63)		
Asia dummy				−0.62 (0.19)		−0.60 (0.23)		
Africa dummy				−1.00 (0.15)		−0.90 (0.17)		
“Other” continent dummy				−0.25 (0.20)		−0.04 (0.32)		
$R^2$	0.62	0.54	0.63	0.73	0.56	0.69	0.55	0.49
Number of observations	110	64	110	110	64	64	108	61

## IV example 2: Do better institutions cause growth?

► In a picture:

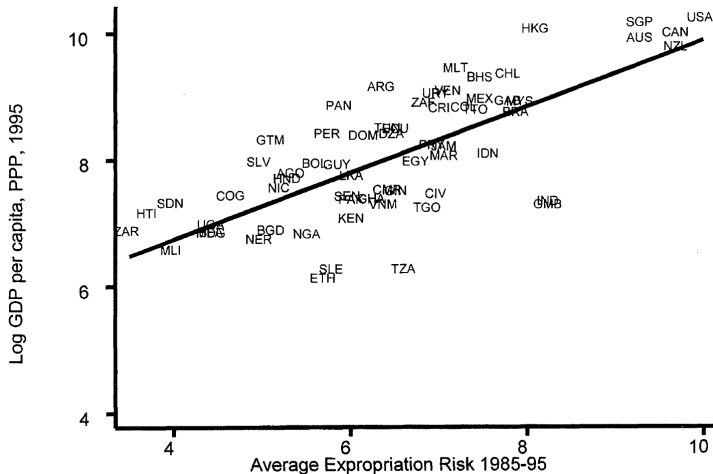


FIGURE 2. OLS RELATIONSHIP BETWEEN EXPROPRIATION RISK AND INCOME



## IV example 2: Do better institutions cause growth?

### ► The first stage:

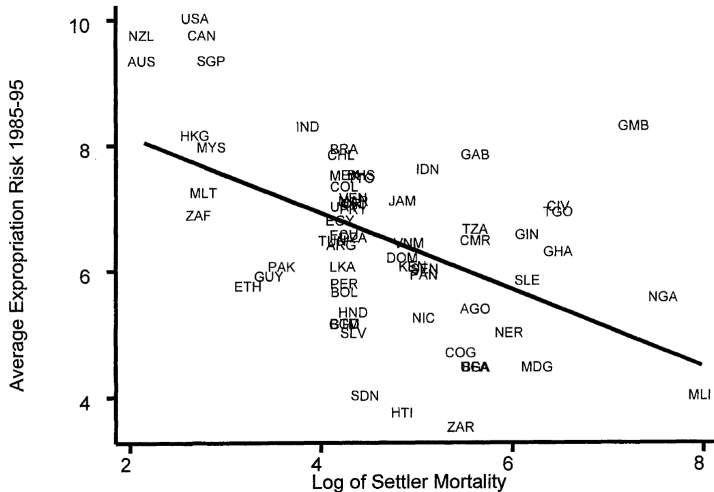


FIGURE 3. FIRST-STAGE RELATIONSHIP BETWEEN SETTLER MORTALITY AND EXPROPRIATION RISK

# IV example 2: Do better institutions cause growth?

TABLE 4—IV REGRESSIONS OF LOG GDP PER CAPITA

	Base sample (1)	Base sample (2)	Base sample without Neo-Europes (3)	Base sample without Neo-Europes (4)	Base sample without Africa (5)	Base sample without Africa (6)	Base sample with continent dummies (7)	Base sample with continent dummies (8)	Base sample, dependent variable is log output per worker (9)
Panel A: Two-Stage Least Squares									
Average protection against expropriation risk 1985–1995	0.94 (0.16)	1.00 (0.22)	1.28 (0.36)	1.21 (0.35)	0.58 (0.10)	0.58 (0.12)	0.98 (0.30)	1.10 (0.46)	0.98 (0.17)
Latitude		−0.65 (1.34)		0.94 (1.46)		0.04 (0.84)		−1.20 (1.8)	
Asia dummy							−0.92 (0.40)	−1.10 (0.52)	
Africa dummy							−0.46 (0.36)	−0.44 (0.42)	
“Other” continent dummy							−0.94 (0.85)	−0.99 (1.0)	
Panel B: First Stage for Average Protection Against Expropriation Risk in 1985–1995									
Log European settler mortality	−0.61 (0.13)	−0.51 (0.14)	−0.39 (0.13)	−0.39 (0.14)	−1.20 (0.22)	−1.10 (0.24)	−0.43 (0.17)	−0.34 (0.18)	−0.63 (0.13)
Latitude		2.00 (1.34)		−0.11 (1.50)		0.99 (1.43)		2.00 (1.40)	
Asia dummy							0.33 (0.49)	0.47 (0.50)	
Africa dummy							−0.27 (0.41)	−0.26 (0.41)	
“Other” continent dummy							1.24 (0.84)	1.1 (0.84)	
R <sup>2</sup>	0.27	0.30	0.13	0.13	0.47	0.47	0.30	0.33	0.28
Panel C: Ordinary Least Squares									
Average protection against expropriation risk 1985–1995	0.52 (0.06)	0.47 (0.06)	0.49 (0.08)	0.47 (0.07)	0.48 (0.07)	0.47 (0.07)	0.42 (0.06)	0.40 (0.06)	0.46 (0.06)
Number of observations	64	64	60	60	37	37	64	64	61

## IV example 3: Does more school choice improve education?

- ▶ Hoxby (2000) studies the effect of school choice on school productivity (both achievement and spending)
- ▶ We want to look at the effect of the number of school districts households can choose from on schools' productivity
  - ▶ Competition among schools improves outcomes
  - ▶ Tiebout sorting of households can reduce achievement
- ▶ Available school choice is endogenous. Why?

# IV example 3: Does more school choice improve education?

- ▶ How do we measure choice?
- ▶ number of districts per student
- ▶ concentration of amount of land taken up by school districts:

$$1 - H_m = 1 - \sum_{k=1}^K \left( \frac{\text{land area}_{km}}{\text{land area}_m} \right)$$

- ▶ concentration of student enrollments:

$$1 - \sum_{k=1}^K (\text{enrollment}_{km} / \text{enrollment}_m)$$

TABLE 1—MEASURES OF TIEBOUT CHOICE

Panel A: Descriptive Statistics			
Measure	Mean	Standard deviation	Standard deviation, controlling for metropolitan-area size <sup>a</sup>
Index of choice among districts, based on enrollment	0.686	0.271	0.250
Index of choice among districts, based on land area	0.761	0.269	0.252
Districts in metropolitan area	21.132	27.611	18.751
Difference in commuting time (minutes) between the district with the third shortest commute and the district with the shortest commute	6.498	8.551	
Index of choice among <i>schools</i> , based on enrollment	0.974	0.069	0.062

## IV example 3: Does more school choice improve education?

- ▶ Hoxby proposes an instrument for the available school choice:
- ▶ Historically, boundaries of school districts were determined by closeness to schools.
- ▶ It's no use being 10 miles from a school if there's a river in between and no bridge.
- ▶ So, she uses the number of rivers in the metropolitan area as an instrument.

TABLE 2—SELECTED COEFFICIENTS FROM THE IMPLIED FIRST-STAGE REGRESSION

	Dependent variable	
	Index of choice among districts (based on enrollment)	Index of choice among schools (based on enrollment)
Number of larger streams in metropolitan area <sup>a</sup>	0.080 (0.040)	-0.040 (0.045)
Number of smaller streams in metropolitan area <sup>a</sup>	0.034 (0.007)	0.004 (0.004)
Population of metropolitan area (thousands)	0.015 (0.013)	0.001 (0.006)
Land area of metropolitan area (thousands of square miles)	0.005 (0.005)	0.003 (0.002)

# IV example 3: Does more school choice improve education?

## ► 2SLS results

TABLE 4—EFFECT OF TIEBOUT CHOICE ON ACHIEVEMENT:  
COEFFICIENT ON INDEX OF CHOICE FOR VARIOUS SPECIFICATIONS

Specification:	Dependent variable:					
	8th-grade reading score	10th-grade math score	12th-grade reading score	ASVAB math knowledge	Highest grade attained	ln(income) at age 32
Base IV specification (see previous table)	3.818 (1.591)	3.061 (1.494)	5.770 (2.208)	2.747 (1.570)	1.381 (0.469)	0.151 (0.072)
Base specification estimated by OLS	−0.236 (0.493)	−0.733 (0.564)	−1.434 (0.650)	2.024 (0.561)	0.323 (0.150)	0.055 (0.029)
Base IV without measures of district heterogeneity	4.649 (1.598)	2.573 (1.478)	6.084 (2.276)	does not apply	does not apply	does not apply
Base IV aggregated to metropolitan-area level	5.137 (3.428)	2.663 (3.419)	7.149 (4.844)	2.860 (4.587)	1.285 (1.229)	0.170 (0.239)
Base IV with choice index based on district land area	4.761 (1.429)	2.875 (1.486)	5.803 (2.179)	2.855 (1.597)	1.516 (0.517)	0.159 (0.073)
Base IV with choice index based on <i>schools'</i> enrollment	61.357 (44.128)	−57.414 (52.959)	−130.577 (95.960)	−18.832 (23.835)	8.031 (12.013)	1.436 (2.341)

## IV example 3: Does more school choice improve education?

TABLE 6—EFFECT OF TIEBOUT CHOICE ON SCHOOL INPUTS AND PRIVATE SCHOOLING:  
COEFFICIENT ON INDEX OF CHOICE FOR VARIOUS SPECIFICATIONS

Specification:	Dependent variable:		
	ln(per-pupil spending)	Student-teacher ratio	Share of students in private school
	−0.076 (0.034)	−2.669 (1.084)	−0.042 (0.018)
Base IV specification (see previous table)	−0.072 (0.022)	0.375 (0.268)	0.006 (0.006)
Base specification estimated by OLS	−0.058 (0.033)	−2.493 (0.994)	−0.067 (0.022)
Base IV without measures of district heterogeneity	−0.064 (0.049)	−2.448 (1.463)	−0.069 (0.031)
Base IV aggregated to metropolitan-area level	−0.101 (0.043)	−2.582 (1.122)	−0.043 (0.020)
Base IV with choice index based on district land area	−0.803 (0.934)	−3.828 (5.372)	−0.180 (0.159)
Base IV with choice index based on <i>schools'</i> enrollment	1.222	0.085	0.144
Test statistic, omnibus overidentification test (distributed $\chi^2_{d.f.=1}$ )			
Test statistic, exogeneity of larger streams variable (distributed $\chi^2_{d.f.=1}$ )	1.021	0.047	0.171

*Notes:* The base specification is shown in the previous table. The notes for that table apply to this table.

*Sources:* Author's calculations based on data from SDDb, CCD, CCDB, GNIS, and USGS maps.

# Outline

## The Evaluation Problem and Potential Outcomes

## Instrumental Variables

IV In The Wild

### IV Pitfalls

Weak Instruments

Heterogeneity: LATE

## Difference in Differences

DiD and Endogeneity

DiD in the Wild

## Regression Discontinuity Designs

Sharp RDD

Fuzzy RDD

RDD in the wild



# Pitfalls

- ▶ IV sounds like a panacea: With a suitable set of instruments, we can wash away endogeneity of our regressors.
  - ▶ That's true, but there are some hitches.
1. **Precision:** Standard errors can get large
  2. **Bias:** If instruments are many and “weak”,  $\hat{\beta}_{2SLS}$  is biased in finite samples, often severely.
  3. **Interpretation:** If effects are different for different people, IV picks up the average effect *for the people for whom the instrument is relevant*

## Weak Instruments

- ▶  $\hat{\beta}_{2SLS}$  is consistent...
- ▶ ...but it is biased in finite samples  $E[\hat{\beta}_{2SLS} - \beta] \neq 0$
- ▶ How does this happen? It comes from the estimation of the first stage
- ▶ In the first stage we regress the endogenous variables on the instruments, with some error.
- ▶ That error comes from the endogenous variables, and is correlated with the errors in the second stage.
- ▶ If the first stage is weak (low F stat, all coefficients nearly 0), then all of the error in the endogenous variables contaminates the fitted values.
- ▶ This error is correlated with the second-stage errors. In fact, it's exactly the correlation that we're trying to get rid of (endogeneity)
- ▶ This correlation causes 2SLS to be biased towards OLS

# Weak Instruments

- ▶ The Limited Information Maximum Likelihood (LIML) estimator is approximately unbiased (`ivregress liml y x...` in stata).
- ▶ Reconsider the QOB instruments:

Table 4.6.2: Alternative IV estimates of the economic returns to schooling

	(1)	(2)	(3)	(4)	(5)	(6)
2SLS	0.105 (0.020)	0.435 (0.450)	0.089 (0.016)	0.076 (0.029)	0.093 (0.009)	0.091 (0.011)
LIML	0.106 (0.020)	0.539 (0.627)	0.093 (0.018)	0.081 (0.041)	0.106 (0.012)	0.110 (0.015)
F-statistic (excluded instruments)	32.27	0.42	4.91	1.61	2.58	1.97
<i>Controls</i>						
Year of birth	✓	✓	✓	✓	✓	✓
State of birth					✓	✓
Age, Age squared		✓		✓		✓
<i>Excluded Instruments</i>						
Quarter of birth	✓	✓				
Quarter of birth*year of birth			✓	✓	✓	✓
Quarter of birth*state of birth					✓	✓
Number of excluded instruments	3	2	30	28	180	178

Notes: The table compares 2SLS and LIML estimates using alternative sets of instruments and controls. The OLS estimate corresponding to the models reported in columns 1-4 is .071; the OLS estimate corresponding to the models reported in columns 5-6 is .067. Data are from the Angrist and Krueger (1991) 1980 Census sample. The sample size is 329,509. Standard errors are reported in parentheses.

## Weak Instruments: Potential Solutions

- ▶ So what should you do in practice? Always be worried about weak instruments! and:
  1. Report your first stage regression. Does it make sense? Are the magnitudes and signs of the coefficients what you expected? If not, the mechanism you're putting forward for the first-stage may not really exist.
  2. Look at the F-stat in the 1st stage. Rule of thumb:  $>10$  good
  3. Use only your best instrument by itself. Remember, adding bad instruments makes F smaller.
  4. Compare your 2SLS estimates to LIML.
  5. Look at your reduced form regressions. The reduced form is proportional to the causal effect you're trying to estimate, so if you can't see the causal relationship in the reduced form, it's probably not there in 2SLS...

# Assuming Constant Effects: Heterogeneity and Non-Linearity

- So far, we have been assuming that the effect we are trying to estimate is constant:

$$y_i = \alpha + \rho s_i + A_i' \gamma + \nu_i$$

- For a binary treatment we assume  $y_{1i} - y_{0i} = \rho$ .  $\rho$  is the same for all individuals: homogeneity
- For multivalued treatments we assume  $y_{si} - y_{s-1i} = \rho$ .
  - $\rho$  is the same for all individuals: homogeneity
  - $\rho$  is the same for all levels of  $s$ : linearity

# Heterogeneity: Internal and External Validity

- ▶ Let's start with heterogeneity (allowing  $\rho$  to be different for different individuals:  $\rho_i$ )
- ▶ Why do we care? It depends what we're after:
- ▶ Our estimates are **internally valid** if we believe the assumptions we need to make for our estimates to be consistent. In this case our estimates uncover the true relationship of interest for the population being studied (at least asymptotically)
- ▶ Our estimates are **externally valid** if we believe that our estimates have strong predictive power in other populations. Do IV estimates using QOB and people born in the 1930s tell us what would happen if we raised the school leaving age in California in 2017?
- ▶ With heterogeneity in  $\rho$  we might worry that we have *internally valid* estimates but that they might lack *external validity*

# Local Average Treatment Effects

- ▶ To think about this issue, let's extend our potential outcomes framework.
- ▶ Take a concrete examples, How does serving in the military affect earnings.

$$y_i = f_i(D_i) + \varepsilon_i$$

- ▶ OLS is biased (why?)
- ▶ Treatment:

$$D_i = \begin{cases} 0 & \text{non-veteran} \\ 1 & \text{veteran} \end{cases}$$

- ▶ Instrument:

$$Z_i = \begin{cases} 0 & \text{eligible for draft during Vietnam (based on SSN)} \\ 1 & \text{ineligible for draft} \end{cases}$$

# Local Average Treatment Effects

- ▶ Before we had two potential outcomes  $y_i(D_i = 0)$  and  $y_i(D_i = 1)$
- ▶ Now generalize this to  $y_i(d, z)$ , the potential outcome when  $D_i = d$  and  $Z_i = z$
- ▶ So causal effect of being a veteran, given draft status is  $y_i(1, Z_i) - y_i(0, Z_i)$
- ▶ Causal effect of draft eligibility, given veteran status is  $y_i(D_i, 1) - y_i(D_i, 0)$



## Local Average Treatment Effects

- ▶ Our instrumental variables strategy relies on the instrument  $Z_i$  setting off a causal chain

$$\underbrace{Z_i \Rightarrow D_i}_{\text{first stage}} \Rightarrow y_i$$

- ▶ So we can think of “potential outcomes” for the treatment  $D_i$  too:

$$\begin{aligned} D_i &= D_{0i} + (D_{1i} - D_{0i}) Z_i \\ &= \pi_0 + \pi_{1i} Z_i + \xi_i \end{aligned}$$

where  $\pi_0 = E[D_{0i}]$ ,  $\pi_{1i} = D_{1i} - D_{0i}$  and  $\xi_i = D_{0i} - E[D_{0i}]$

- ▶  $D_{0i}$ : whether someone with a draft ineligible number would serve in the army
- ▶  $D_{1i}$ : whether someone with a draft eligible number would serve in the army
- ▶ We only observe  $D_i$ , which one we see depends on  $Z_i$

# Local Average Treatment Effects

## Assumption (A.1. Independence)

*The instrument is as good as randomly assigned:*

$$[\{y_i(d, z); \forall d, z\}, D_{1i}, D_{0i}] \perp Z_i$$

- ▶ The instrument is independent of the potential outcomes and potential treatments.
  - ▶ e.g. people with large  $y_i(D_i, 1)$  aren't disproportionately likely to have  $Z_i = 1$ 
    - ▶ SSNs are randomly assigned, so likely to satisfy this.
  - ▶ Implies the reduced form regression of  $y$  on  $Z$  is causal:

$$\begin{aligned} & \mathbb{E}[y_i | Z_i = 1] - \mathbb{E}[y_i | Z_i = 0] \\ &= \mathbb{E}[y_i(D_{1i}, 1) | Z_i = 1] - \mathbb{E}[y_i(D_{0i}, 0) | Z_i] \\ &= \mathbb{E}[y_i(D_{1i}, 1) - y_i(D_{0i}, 0)] \end{aligned}$$

# Local Average Treatment Effects

- ▶ Also implies that the first stage regression of  $D$  on  $Z$  is causal:

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

- ▶ And we'll assume that this first stage regression works:

## Assumption (A.2. First Stage)

$$\mathbf{E}[D_{1i} - D_{0i}] \neq 0$$

# Local Average Treatment Effects

## Assumption (A.2. Exclusion)

*Potential outcomes  $y_i(d, z)$  are functions only of  $d$ :*

$$y_i(d, 1) = y_i(d, 0) \quad \text{for } d = 0, 1$$

- ▶ Of course draft eligibility affects earnings, but *only* through changes in veteran status.
- ▶ In the linear model with constant effects we have been studying, the exclusion and independence assumptions are bundled together in  $E[\varepsilon|\mathbf{Z}] = 0$
- ▶ NB exclusion can fail while independence is satisfied:
  - ▶ e.g. if people know they could be drafted, and stay in school longer to avoid the draft, then  $y_i(d, 1) > y_i(d, 0)$  whether or not they eventually serve.
  - ▶ This is despite SSNs being random.

## Local Average Treatment Effects

- ▶ With the exclusion restriction, we can justify the single index potential outcomes we were using before:

$$y_{1i} \equiv y_i(1, 1) = y_i(1, 0)$$

$$y_{0i} \equiv y_i(0, 1) = y_i(0, 0)$$

so that

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

- ▶ Or, as before, we can write this as

$$y_i = \alpha_0 + \rho_i D_i + \eta_i$$

where  $\alpha_0 = E[y_{0i}]$ ,  $\rho_i = y_{1i} - y_{0i}$  and  $\eta_i = y_{0i} - E[y_{0i}]$

# Local Average Treatment Effects

- ▶ We need one more thing, and then we're ready to go.

## Assumption (A.4. Monotonicity)

*All those affected by the instrument are affected in the same way:*

$$\text{Either } \pi_{1i} \geq 0 \ \forall i \ \text{Or } \pi_{1i} \leq 0 \ \forall i$$

- ▶ So that either  $D_{1i} \geq D_{0i}$  or  $D_{1i} \leq D_{0i}$  for everyone.
- ▶ E.g. nobody who would have joined if they were ineligible stayed out of the army *because* they were draft eligible
- ▶ Combining these 4 things, we can be precise about what the Wald estimator (IV) estimates in a setting with heterogeneity

# Local Average Treatment Effects

## Theorem (LATE Theorem)

*Suppose that*

*(A.1. Independence)  $\{y_i(D_{1i}, 1), y_i(D_{0i}, 0), D_{1i}, D_{0i}\} \perp Z_i$ ;*

*(A.2. First Stage)  $E[D_{1i} - D_{0i}] \neq 0$ ;*

*(A.3. Exclusion)  $y_i(d, 0) = y_i(d, 1) \equiv y_{di}$  for  $d = 0, 1$ ;*

*(A.4. Monotonicity)  $D_{1i} - D_{0i} \geq 0 \forall i$  or vice versa; Then*

$$\begin{aligned}\hat{\beta}^{Wald} &= \frac{E[y_i|Z_i = 1] - E[y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]} = E[y_{1i} - y_{0i} | D_{1i} > D_{0i}] \\ &= E[\rho_i | D_{1i} > D_{0i}]\end{aligned}$$

# Local Average Treatment Effects

- ▶ In light of the LATE theorem, how should we interpret IV estimates?
- ▶ They are the estimates of the effect of the endogenous variable for the subgroup of people whose endogenous variable was affected by the instrument: the **compliant subpopulation**
- ▶ In the veterans case: It is the effect of being a veteran *for those who are veterans because of their draft eligibility* i.e. not for those who would have served anyway
- ▶ So the estimates are *internally valid* for these compliers.
- ▶ What about *external validity*? Well, it depends what you want to use the estimates for. Probably useful for future draft policies, probably less so for effects on volunteers (Iraq vets?)



# Local Average Treatment Effects

Table 4.1.3: Wald estimates of the effects of military service on the earnings of white men born in 1950

Earnings year	Earnings		Veteran Status		Wald Estimate of Veteran Effect
	Mean	Eligibility Effect	Mean	Eligibility Effect	
	(1)	(2)	(3)	(4)	(5)
1981	16,461	-435.8 (210.5)	0.267	0.159 (0.040)	-2,741 (1,324)
1971	3,338	-325.9 (46.6)			-2050 (293)
1969	2,299	-2.0 (34.5)			

Notes: Adapted from Angrist (1990), Tables 2 and 3. Standard errors are shown in parentheses. Earnings data are from Social Security administrative records. Figures are in nominal dollars. Veteran status data are from the Survey of Program Participation. There are about 13,500 individuals in the sample.

# The Compliant Subpopulation

► We can partition people in the data into 3 groups:

1. **Compliers:**  $D_{0i} = 0, D_{1i} = 1$
2. **Always-Takers:**  $D_{0i} = D_{1i} = 1$
3. **Never-Takers:**  $D_{0i} = D_{1i} = 0$

		$D_{0i}$	
		0	1
$D_{1i}$	0	Never-Takers	Defiers (ruled out by monotonicity)
	1	<b>Compliers: LATE</b>	Always-Takers

# The Compliant Subpopulation

- ▶ This lets us think about the effect on the treated  $D_i = 1$  and the untreated  $D_i = 0$
- ▶ If  $D_i = 1$  then by monotonicity it must be the case that *either*
  - ▶  $D_{0i} = 1$  (Always Taker)
  - ▶  $D_{1i} - D_{0i} = 1$  and  $Z_i = 1$  (Complier with  $Z_i = 1$ )
- ▶ With this we can decompose:

$$\underbrace{E[y_{1i} - y_{0i} | D_i = 1]}_{\text{Effect on Treated}} = \underbrace{E[y_{1i} - y_{0i} | D_{0i} = 1]}_{\text{always takers' effect}} \Pr(D_{0i} = 1 | D_i = 1) + \underbrace{E[y_{1i} - y_{0i} | D_{1i} > D_{0i}]}_{\text{compliers' effect}} \Pr(D_{1i} > D_{0i}, Z_i = 1 | D_i = 1)$$

- ▶ So Average Treatment Effect on the Treated (ATT) is a weighted average of LATE and effect on always-takers

# The Compliant Subpopulation

- ▶ What about the effect on the untreated?
- ▶ If  $D_i = 0$  then by monotonicity either
  - ▶  $D_{1i} = 0$  (never taker)
  - ▶  $D_{1i} - D_{0i} = 1$  and  $Z_i = 0$  (Complier with 0)
- ▶ With this we can decompose:

$$\underbrace{E[y_{1i} - y_{0i} | D_i = 0]}_{\text{Effect on Untreated}} = \underbrace{E[y_{1i} - y_{0i} | D_{1i} = 0]}_{\text{never takers' effect}} \Pr(D_{1i} = 0 | D_i = 0) + \underbrace{E[y_{1i} - y_{0i} | D_{1i} > D_{0i}]}_{\text{compliers' effect}} \Pr(D_{1i} > D_{0i}, Z_i = 0 | D_i = 0)$$

- ▶ So Average Treatment Effect on the Untreated (ATU) is a weighted average of LATE and effect on never-takers

# The Compliant Subpopulation

- ▶ These things are important for how we think about external validity
- ▶ An instrument can get us a direct estimate of LATE, but never the always-takers or the never-takers
- ▶ Therefore, we don't usually get the effect on all of the treated or on all of the untreated
- ▶ *Except* in 2 special cases: Instruments that don't allow any always-takers or never-takers
- ▶ E.g.1 twins instrument:
  - ▶ If second birth is twins, then it is impossible to have only 2 children → there are no never-takers.
  - ▶ Since no never-takers, LATE also estimates effect on untreated

# Who are the Compliers?

- ▶ We're worried now that IV estimates of LATE lack external validity.
- ▶ We can't ever be 100% sure that LATE estimates are externally valid, but we can, at least, say something about who the compliers are.
- ▶ In particular, we can say
  1. How many compliers are there? What is  $\Pr(D_{1i} > D_{0i})$ ?
  2. How many of the treated are compliers? What is  $\Pr(D_{1i} > D_{0i} | D_i = 1)$ ?
  3. What are the characteristics of compliers? Are they more likely to be older? Have higher education? etc.

# Who are the Compliers?

► Let's start with how many compliers there are:

1. What is  $\Pr(D_{1i} > D_{0i})$ ?

► By monotonicity  $D_{1i} - D_{0i}$  is either 0 or 1, so

$$\begin{aligned}\Pr(D_{1i} > D_{0i}) &= \mathbb{E}[D_{1i} - D_{0i}] \\ &= \mathbb{E}[D_{1i}] - \mathbb{E}[D_{0i}] \\ &= \mathbb{E}[D_i | Z_i = 1] - \mathbb{E}[D_i | Z_i = 0]\end{aligned}$$

## Who are the Compliers?

2. How many of the treated are compliers? And by implication, how many are always-takers?

- Start by using the definition of conditional probability:

$$\Pr(D_{1i} > D_{0i} | D_i = 1) = \frac{\Pr(D_i = 1 | D_{1i} > D_{0i}) \Pr(D_{1i} > D_{0i})}{\Pr(D_i = 1)}$$

- Amongst the compliers, the only ones who are treated are those who have the instrument switched on.

$$\begin{aligned}\Pr(D_i = 1 | D_{1i} > D_{0i}) &= \Pr(Z_i = 1 | D_{1i} > D_{0i}) \\ &= \Pr(Z_i = 1)\end{aligned}$$

where the second equality follows by independence. So,

$$\Pr(D_{1i} > D_{0i} | D_i = 1) = \frac{\Pr(Z_i = 1) (\mathbb{E}[D_i | Z_i = 1] - \mathbb{E}[D_i | Z_i = 0])}{\Pr(D_i = 1)}$$



# Who are the Compliers?

Table 4.4.2: Probabilities of compliance in instrumental variables studies

Source	Endogenous Variable (D)	Instrument (Z)	Sample	$P[D = 1]$	1st Stage, $P[D_1 > D_0]$	$P[Z = 1]$	$P[D_1 > D_0   D = 1]$	$P[D_1 > D_0   D = 0]$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Angrist (1990)	Veteran Status	Draft eligibility	White men born in 1950	0.267	0.159	0.534	0.318	0.101
			Non-white men born in 1950	0.163	0.060	0.534	0.197	0.033
Angrist and Evans (1998)	More than 2 children	Twins at second birth	Married women aged 21-35 with two or more children in 1980	0.381	0.603	0.008	0.013	0.966
		First two children are of the same sex	Married women aged 21-35 with two or more children in 1980	0.381	0.060	0.506	0.080	0.048
Angrist and Krueger (1991)	High school graduate	Third or fourth quarter birth	Men born between 1930 and 1939	0.770	0.016	0.509	0.011	0.034
Acemoglu and Angrist (2000)	High school graduate	State requires 11 or more years of school attendance	White men aged 40-49	0.617	0.037	0.300	0.018	0.068

Notes: The table shows an analysis of the absolute and relative size of the complier population for a number of instrumental variables. The first-stage, reported in column 6, gives the absolute size of the complier group. Columns 8 and 9 show the size of the complier population relative to the treated and untreated populations.

# Outline

## The Evaluation Problem and Potential Outcomes

## Instrumental Variables

- IV In The Wild

- IV Pitfalls

  - Weak Instruments

  - Heterogeneity: LATE

## Difference in Differences

- DiD and Endogeneity

- DiD in the Wild

## Regression Discontinuity Designs

- Sharp RDD

- Fuzzy RDD

- RDD in the wild

# Outline

## The Evaluation Problem and Potential Outcomes

## Instrumental Variables

- IV In The Wild

- IV Pitfalls

  - Weak Instruments

  - Heterogeneity: LATE

## Difference in Differences

- DiD and Endogeneity

- DiD in the Wild

## Regression Discontinuity Designs

- Sharp RDD

- Fuzzy RDD

- RDD in the wild

## Difference in Differences

- ▶ Let's consider a concrete example of a policy we might evaluate:
- ▶ On April 1 1992 New Jersey raised the state minimum wage from \$4.25 to \$5.05
- ▶ Card & Krueger (1994) collected data on wages and employment from fast food chains in 2/92 and 11/92.
- ▶ They collect the same data for Pennsylvania, a bordering state that did not change the minimum wage.
- ▶ The idea is to compare how wages changed in NJ between Feb and Nov, to how wages changed in PA between Feb and Nov.
- ▶ i.e. calculate the *difference* between the difference between Feb and Nov wages in NJ and the difference between Feb and Nov wages in PA: The *difference in the differences*.

# Difference in Differences

- ▶ Denote employment at restaurant  $i$  in state  $s$  at time  $t$  by  $y_{ist}$
- ▶ In the potential outcomes notation we have been using,  $y_{0ist}$  is the employment at restaurant  $i$  when there is a low minimum wage, and  $y_{1ist}$  is employment when there is a high minimum wage.
- ▶ i.e. in New Jersey we see  $y_{1ist}$  in November, but  $y_{0ist}$  in February.
- ▶ We assume that

$$E[y_{0ist}|s, t] = \gamma_s + \lambda_t$$

where  $s \in \{NJ, PA\}$  and  $t \in \{Feb, Nov\}$

- ▶ And, we assume that

$$E[y_{1ist} - y_{0ist}|s, t] = \delta$$

(which, note, doesn't depend on  $s$  or  $t$ )

# Difference in Differences

- ▶ Then we can write the model as

$$y_{ist} = \gamma_s + \lambda_t + \delta D_{st} + \varepsilon_{ist}$$

where  $D_{st}$  is a dummy variable for high minimum-wage states and time periods (in this example,  $D_{st} = 1$  in NJ in November, and 0 otherwise), and  $E[\varepsilon_{ist}|s, t] = 0$

- ▶ In this case, the Pennsylvania time-difference is

$$\begin{aligned} E[y_{ist}|s = PA, t = Nov] - E[y_{ist}|s = PA, t = Feb] \\ = (\gamma_{PA} + \lambda_{Nov}) - (\gamma_{PA} + \lambda_{Feb}) = \lambda_{Nov} - \lambda_{Feb} \end{aligned}$$

- ▶ And the New Jersey time-difference is

$$\begin{aligned} E[y_{ist}|s = NJ, t = Nov] - E[y_{ist}|s = NJ, t = Feb] \\ = (\gamma_{NJ} + \lambda_{Nov} + \delta) - (\gamma_{NJ} + \lambda_{Feb}) = \lambda_{Nov} - \lambda_{Feb} + \delta \end{aligned}$$

# Difference in Differences

## ► Combining these

$$\begin{aligned} & \{E[y_{ist}|s = NJ, t = Nov] - E[y_{ist}|s = NJ, t = Feb]\} \\ & - \{E[y_{ist}|s = PA, t = Nov] - E[y_{ist}|s = PA, t = Feb]\} \\ & = (\lambda_{Nov} - \lambda_{Feb} + \delta) - (\lambda_{Nov} - \lambda_{Feb}) = \delta \end{aligned}$$

## ► We can easily estimate this by using sample averages::

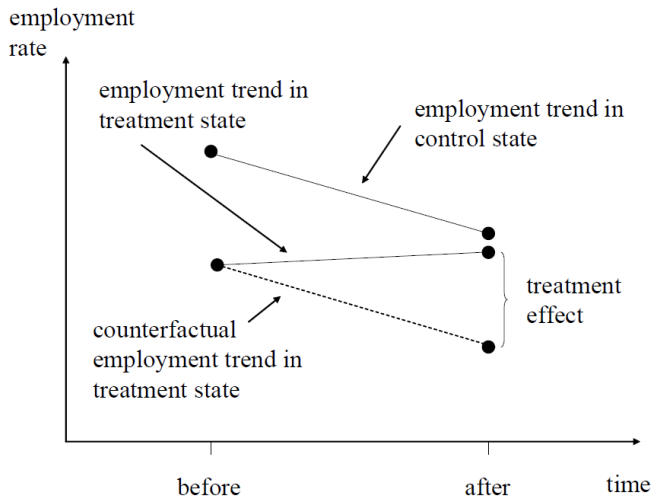
Table 5.2.1: Average employment per store before and after the New Jersey minimum wage increase

Variable	PA (i)	NJ (ii)	Difference, NJ-PA (iii)
1. FTE employment before, all available observations	23.33 (1.35)	20.44 (0.51)	-2.89 (1.44)
2. FTE employment after, all available observations	21.17 (0.94)	21.03 (0.52)	-0.14 (1.07)
3. Change in mean FTE employment	-2.16 (1.25)	0.59 (0.54)	2.76 (1.36)

Notes: Adapted from Card and Krueger (1994), Table 3. The table reports average full-time equivalent (FTE) employment at restaurants in Pennsylvania and New Jersey before and after a minimum wage increase in New Jersey.

# Difference in Differences

- We assume the change in the control state (PA) is what the change in the treatment state (NJ) *would have been* if there had been no treatment. The **Parallel Trends** assumption





# Difference in Differences

- ▶ We can also implement the difference in differences design using a regression:
- ▶ Let  $NJ_s$  be a dummy for restaurants in New Jersey, and  $d_t$  be a dummy for observations from November
- ▶ Then we can write the model as

$$y_{ist} = \alpha + \gamma NJ_s + \lambda d_t + \delta (NJ_s \times d_t) + \varepsilon_{ist}$$

where note that  $NJ_s \times d_t = D_{st}$

## Difference in Differences

- So in terms of the potential outcomes, and the regression coefficients, the 4 groups have

$$E[y_{ist}|s = PA, t = Feb] = \gamma_{PA} + \lambda_{Feb} = \alpha$$

$$E[y_{ist}|s = PA, t = Nov] = \gamma_{PA} + \lambda_{Nov} = \alpha + \lambda$$

$$E[y_{ist}|s = NJ, t = Feb] = \gamma_{NJ} + \lambda_{Feb} = \alpha + \gamma$$

$$E[y_{ist}|s = NJ, t = Nov] = \gamma_{NJ} + \lambda_{Nov} + \delta = \alpha + \gamma + \lambda + \delta$$

- And so we can see that

$$\alpha = E[y_{ist}|s = PA, t = Feb] = \gamma_{PA} + \lambda_{Feb}$$

$$\begin{aligned}\gamma &= E[y_{ist}|s = NJ, t = Feb] - E[y_{ist}|s = PA, t = Feb] \\ &= \gamma_{NJ} - \gamma_{PA}\end{aligned}$$

$$\begin{aligned}\lambda &= E[y_{ist}|s = PA, t = Nov] - E[y_{ist}|s = PA, t = Feb] \\ &= \lambda_{Nov} - \lambda_{Feb}\end{aligned}$$

$$\begin{aligned}\delta &= \{E[y_{ist}|s = NJ, t = Nov] - E[y_{ist}|s = NJ, t = Feb]\} \\ &\quad - \{E[y_{ist}|s = PA, t = Nov] - E[y_{ist}|s = PA, t = Feb]\}\end{aligned}$$

# Difference in Differences

- ▶ We can also extend the model to include individual-level controls and time-varying state level variables:

$$y_{ist} = \gamma_s + \lambda_t + \delta D_{st} + \mathbf{x}'_{it}\beta + \varepsilon_{ist}$$

- ▶ Note that time-varying state level variables could be a source of OVB: If some other characteristics of the state changed at the same time as the policy in the treatment state but not in the control state, then  $\hat{\delta}$  will conflate the effects of this characteristic with the effects of the policy.
- ▶ On the other hand, individual level controls are there only to increase precision.

# Outline

## The Evaluation Problem and Potential Outcomes

## Instrumental Variables

IV In The Wild

IV Pitfalls

Weak Instruments

Heterogeneity: LATE

## Difference in Differences

DiD and Endogeneity

DiD in the Wild

## Regression Discontinuity Designs

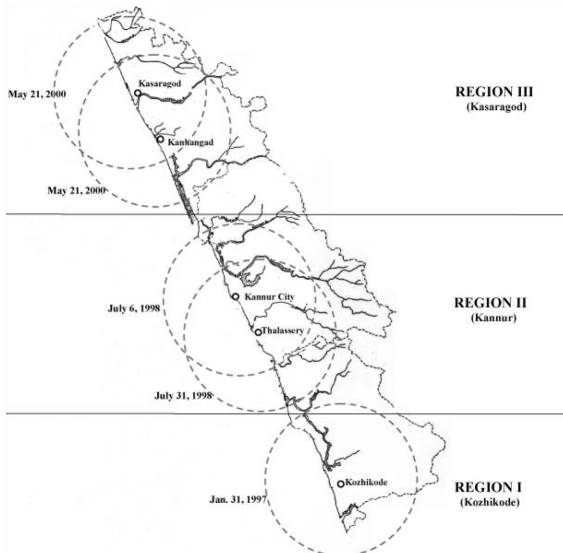
Sharp RDD

Fuzzy RDD

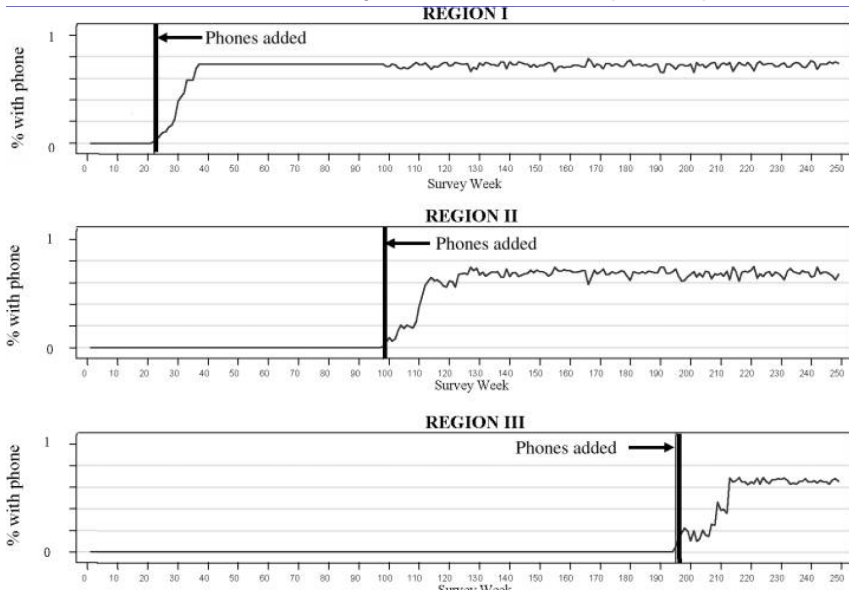
RDD in the wild

# DiD example 1: Mobile Phones in Kerala

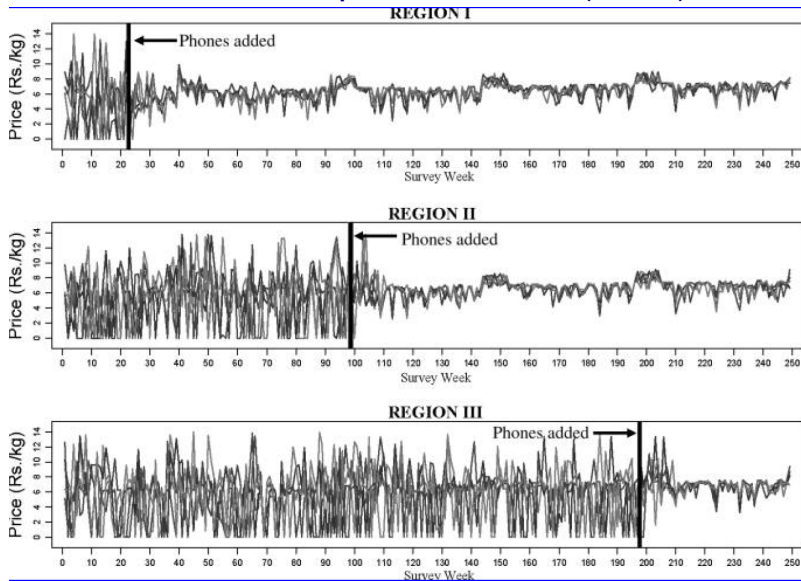
- ▶ Jensen (2007) *natural experiment*: Starting in 1997, mobile phone service was rolled out in the state of Kerala, India.



# DiD example 1: Jensen (2007)



# DiD example 1: Jensen (2007)



# DiD example 1: Jensen (2007)

TABLE IV  
EFFECTS OF MOBILE PHONE SERVICE ON PRICE DISPERSION AND WASTE:  
POOLED TREATMENTS

	(1)	(2)	(3)	(4)	(5)	(6)
	Max-min spread	Coefficient of variation	Percent have waste	Max-min spread	Coefficient of variation	Percent have waste
Phone	-5.0 (0.27)	-.38 (0.03)	-0.048 (0.004)	-5.3 (2.9)	-.41 (0.32)	-0.047 (0.06)
Region I	-0.92 (0.26)	-.06 (0.03)	-0.007 (0.005)	-0.94 (0.26)	-.06 (0.03)	-0.006 (0.005)
Region II	-0.46 (0.21)	-.04 (0.02)	-0.011 (0.004)	-0.46 (0.21)	-.04 (0.02)	-0.011 (0.005)
Period 1	-0.89 (0.29)	-.12 (0.04)	-0.017 (0.008)	-0.84 (0.29)	-.12 (0.03)	-0.016 (0.008)
Period 2	-1.1 (0.32)	-.17 (0.04)	-0.019 (0.008)	-1.0 (0.33)	-.16 (0.04)	-0.018 (0.008)
Period 3	-1.2 (0.40)	-.19 (0.04)	-0.022 (0.009)	-1.2 (0.40)	-.19 (0.04)	-0.021 (0.009)
Fuel cost	0.02 (0.12)	.01 (0.01)	0.001 (0.002)	-0.13 (0.19)	-.02 (0.02)	0.003 (0.005)
Wind/sea index	0.086 (0.051)	.001 (0.004)	-0.002 (0.002)	-0.03 (0.06)	-.01 (0.01)	-0.003 (0.003)
Phone*fuel cost				0.25 (0.14)	.026 (0.014)	-0.003 (0.006)
Phone*wind/sea index				0.19 (0.08)	.021 (0.008)	0.003 (0.005)
Number of observations	747	747	74,700	747	747	74,700



# DiD Example 1: Jensen (2007)

TABLE V  
ESTIMATED EFFECTS OF MOBILE PHONES ON MARKET OUTCOMES:  
SEPARATE TREATMENTS

	Max-min spread	Coefficient of variation	Waste
Estimated effects of adding phones to region I			
(a) Using region II as the control group	-4.8	-.46	-0.064
$(Y_{I,1} - Y_{I,0}) - (Y_{II,1} - Y_{II,0}) = \beta_{RI\_P1}$ $- \beta_{RII\_P1}$	(0.68)	(0.07)	(0.005)
(b) Using region III as the control group	-4.8	-.42	-0.060
$(Y_{I,1} - Y_{I,0}) - (Y_{III,1} - Y_{III,0}) = \beta_{RI\_P1}$ $- \beta_{RII\_P1}$	(0.68)	(0.07)	(0.005)
Estimated effects of adding phones to region II			
(c) Using region I as the control group	-5.8	-.39	-0.039
$(Y_{II,2} - Y_{I,1}) - (Y_{I,2} - Y_{I,1}) = \beta_{RII\_P2}$ $- \beta_{RII\_P1} - \beta_{RI\_P2} + \beta_{RI\_P1}$	(0.43)	(0.05)	(0.003)
(d) Using region III as the control group	-4.9	-.36	-0.038
$(Y_{II,2} - Y_{II,1}) - (Y_{III,2} - Y_{III,1}) = \beta_{RII\_P2}$ $- \beta_{RII\_P1}$	(0.43)	(0.05)	(0.003)
Estimated effects of adding phones to region III			
(e) Using region I as the control group	-4.9	-.38	-0.055
$(Y_{III,3} - Y_{III,2}) - (Y_{I,3} - Y_{I,2}) = \beta_{RI\_P2}$ $- \beta_{RI\_P3}$	(0.48)	(0.05)	(0.004)
(f) Using region II as the control group	-4.7	-.35	-0.054
$(Y_{III,3} - Y_{III,2}) - (Y_{II,3} - Y_{II,2}) = \beta_{RII\_P2}$ $- \beta_{RII\_P3}$	(0.48)	(0.05)	(0.004)

## DiD Example 2: Property Transaction Taxes and Housing Market Activity

- ▶ In September 2008, the UK government cut the property transaction tax “Stamp Duty” from 1% to 0 on houses worth £125,000 – £175,000
- ▶ Surprise-announcement: No anticipation possible
- ▶ Pre-announced end-date: December 31st 2009. Fully anticipated.
- ▶ Does this increase property transactions?
- ▶ How many of these are transactions that would have happened anyway?

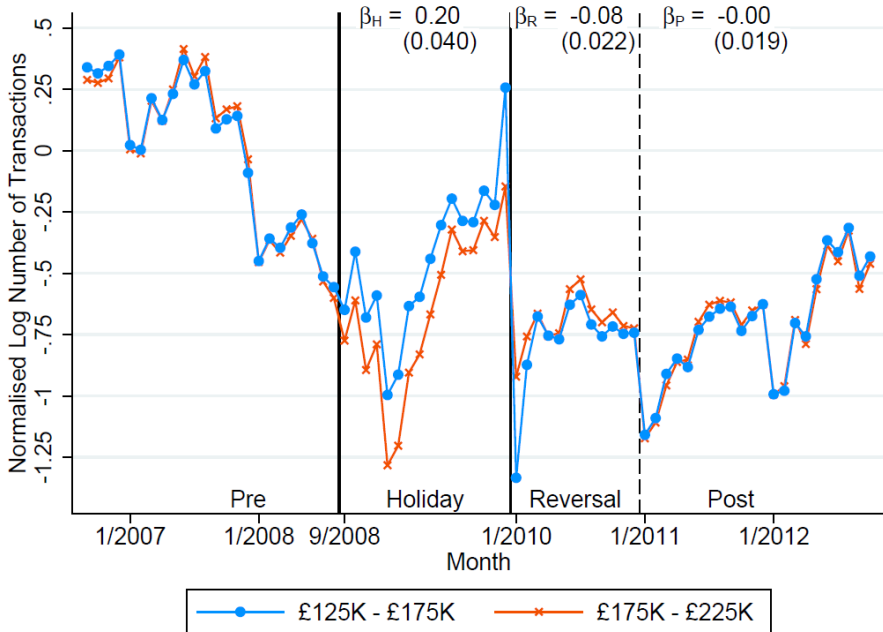
## DiD Example 2: Best & Kleven (2018)

- ▶ Correct for selection: Use bunching estimates to reallocate transactions from treatment to control.
- ▶ Estimate DiD

$$\begin{aligned}n_{it} = & \alpha_0 Pre_t + \alpha_H Hol_t + \alpha_R Rev_t + \alpha_P Post_t + \alpha_T Treated_i \\& + \beta_H Hol_t \times Treated_i + \beta_R Rev_t \times Treated_i \\& + \beta_P Post_t \times Treated_i + \nu_{it}\end{aligned}$$

$i$  is a 5K bin,  $n$  is log # of transactions,  $Pre$  denotes 09/06–08/08,  $Hol$  denotes 09/08–12/09,  $Rev$  denotes 01/10–12/10,  $Post$  denotes 01/11–10/12

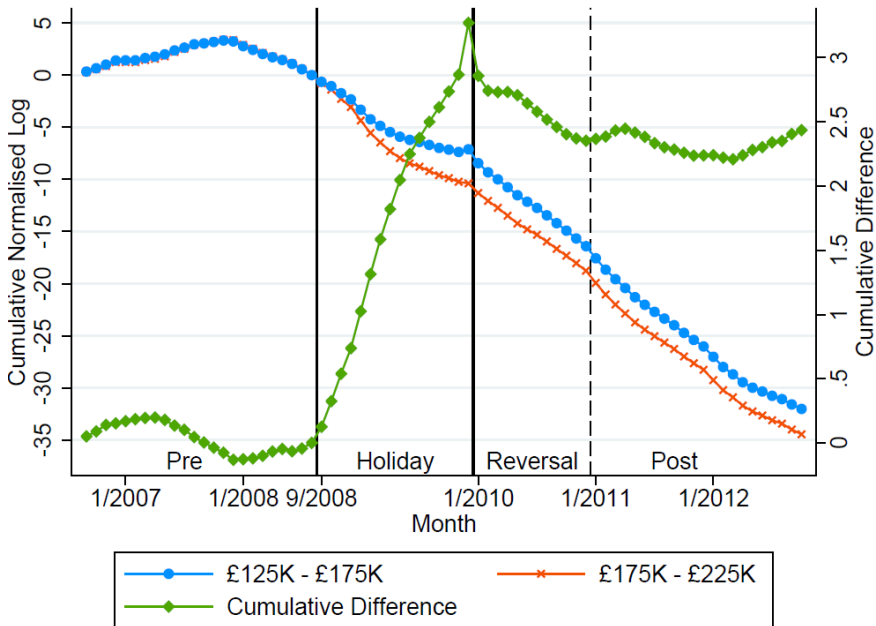
## DiD Example 2: Best & Kleven (2018)



## DiD Example 2: Best & Kleven (2018)

- ▶ Parallel trends
  - ▶ before the tax cut
  - ▶ after 1/2011
- ▶ No anticipation of the cut
- ▶ Anticipation of the end of the stimulus → Spike in transactions

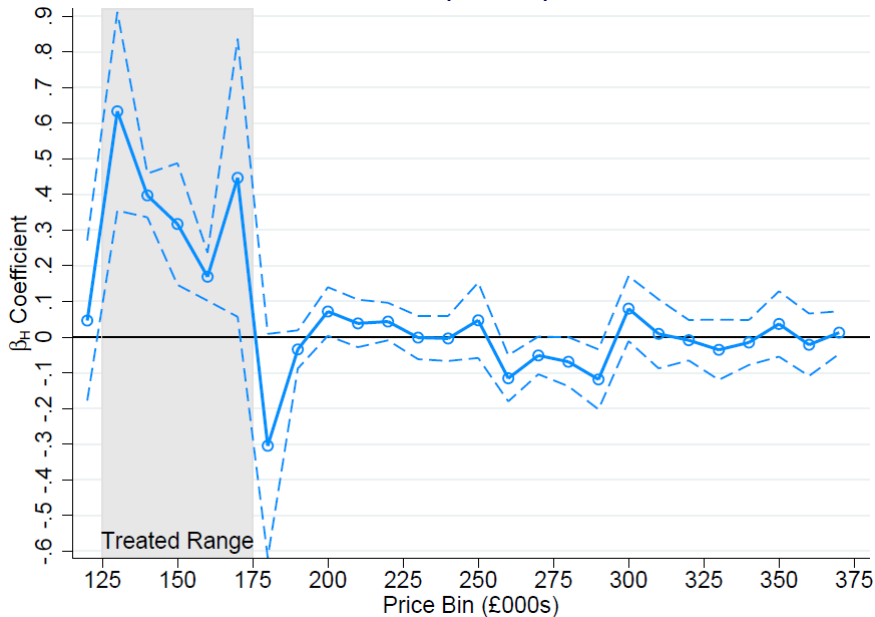
## Best & Kleven (2018): Results



## Best & Kleven (2018): Reversal Date

- ▶ Lull in activity for about a year.
- ▶ Lull doesn't completely offset the gains during the tax holiday.  
Easy to see in cumulative graphs
- ▶ Choice of control group  $£175K - £225K$  is a bit ad-hoc. What about other price ranges
- ▶ Use other price ranges as a **placebo test**.
  1. Pretend that the treatment was in a different price range.
  2. Estimate effect of this pretend treatment → should give zero

## Best & Kleven (2018): Placebo





## DiD Example 3: Synthetic Control

- ▶ What happens when you have many possible control groups?
- ▶ E.g. Abadie, Diamond & Hainmueller (2010) want to study the effect of California's 1988 tobacco control program, Proposition 99.
- ▶ As controls, they want to use other states, but which one(s) should they use?
- ▶ A model: Suppose we observe  $J + 1$  regions, and that only region 1 receives the treatment.
- ▶ We observe each region in time periods  $t = 1, \dots, T$
- ▶ Region 1 receives the treatment from period  $T_0 + 1$  until  $T$

# Synthetic Control

- ▶ Let  $y_{it}^N$  be the potential outcome we *would* observe for region  $i$  at time  $t$  if region  $i$  never receives the treatment
- ▶ Let  $y_{it}^I$  be the potential outcome we *would* observe for region  $i$  at time  $t$  if region  $i$  receives the treatment from periods  $T_0 + 1$  until  $T$
- ▶ Assume the treatment has no impact before period  $T_0 + 1$ :  
 $y_{it}^N = y_{it}^I$  for all  $t = 1, \dots, T_0$  and all  $i = 1, \dots, J + 1$
- ▶ Then  $\alpha_{it} = y_{it}^I - y_{it}^N$  is the effect of the intervention on region  $i$  at time  $t$
- ▶ We only observe

$$y_{it} = y_{it}^N + \alpha_{it}D_{it}$$

where  $D_{it}$  is a dummy for treatment.

# Synthetic Control

- So, for each  $t > T_0$  we'd like to estimate

$$\alpha_{1t} = y_{1t}^I - y_{1t}^N$$

but we only observe  $y_{1t}^I$  so we need to estimate  $y_{1t}^N$ .

- Let's assume the difference in differences model for  $y_{it}^N$

$$y_{it}^N = \delta_t + \boldsymbol{\theta}_t \mathbf{Z}_i + \mu_i + \varepsilon_{it}$$

- Then, we can use any of the other  $J$  regions as a control and construct a series of Diff in Diff estimators. e.g. if we use region  $i = 2$ ,

$$\hat{\alpha}_{1t} = y_{1t} - y_{2t} \quad t = T_0 + 1, \dots, T$$

# Synthetic Control

- ▶ But, we can do better, we can use a weighted average of all the other regions.
- ▶ Let  $\mathbf{W} = (w_2, \dots, w_{J+1})'$  be a  $J \times 1$  vector of weights satisfying  $w_j \leq 0 \ \forall j = 2, \dots, J+1$  and  $\sum_{j=1}^{J+1} w_j = 1$
- ▶ Then any such  $\mathbf{W}$  represents a potential *synthetic control*. For a given  $\mathbf{W}$ , the value of the outcome at time  $t$  is

$$\sum_{j=2}^{J+1} w_j y_{jt} = \delta_t + \boldsymbol{\theta}_t \sum_{j=1}^{J+1} w_j \mathbf{Z}_i + \sum_{j=2}^{J+1} w_j \mu_j + \sum_{j=2}^{J+1} w_j \varepsilon_{jt}$$

- ▶ The optimal weights make the synthetic control look as much as possible like the treatment group in the pre-reform period
- ▶ The authors provide the `synth` package in stata to implement this

# DiD Example 3: Effect of CA Prop 99

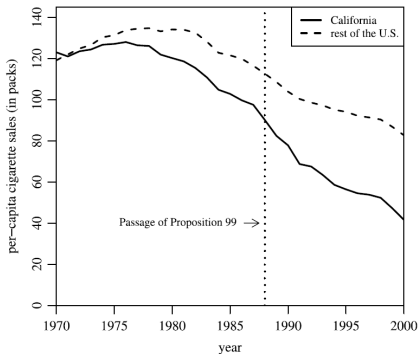


Figure 1. Trends in per-capita cigarette sales: California vs. the rest of the United States.

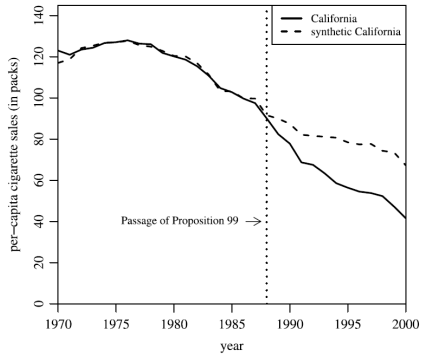


Figure 2. Trends in per-capita cigarette sales: California vs. synthetic California.

# Outline

## The Evaluation Problem and Potential Outcomes

## Instrumental Variables

- IV In The Wild

- IV Pitfalls

  - Weak Instruments

  - Heterogeneity: LATE

## Difference in Differences

- DiD and Endogeneity

- DiD in the Wild

## Regression Discontinuity Designs

- Sharp RDD

- Fuzzy RDD

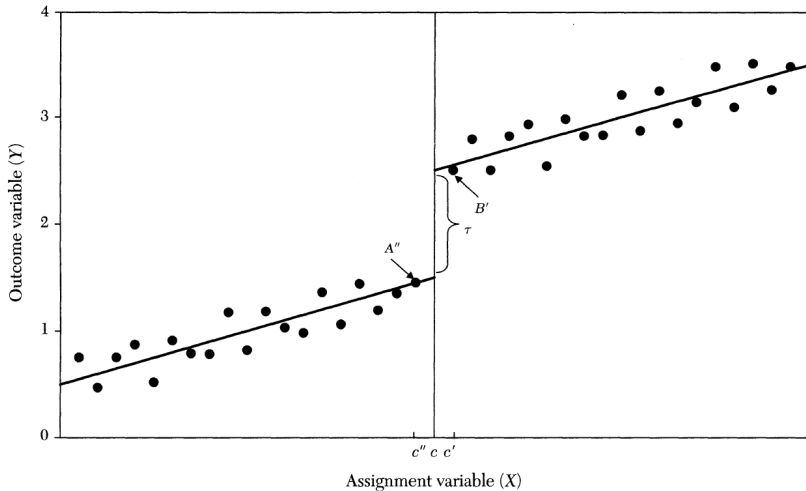
- RDD in the wild

# Regression Discontinuity Designs

- ▶ The world is full of rules containing (arbitrary) thresholds, which often determine whether people receive treatments:
- ▶ Eligibility for Medicare starts at age 65 in USA (Card, Dobkin & Maestas, 2008)  
 $\Rightarrow \text{Insurance} = f(\text{Age})$
- ▶ Classes can't contain more than 40 students in Israel (Angrist & Lavy, 1999)  
 $\Rightarrow \text{Class Size} = f(\text{Cohort Size})$
- ▶ Municipalities with >4K voters had to use electronic voting in Brazil in 1998 (Fujiwara, 2015)  
 $\Rightarrow \text{Electoral Fraud} = f(\text{Municipality population})$
- ▶ People with >36 months of tenure get severance pay in Austria (Card, Chetty & Weber, 2007)  
 $\Rightarrow \text{Cash on Hand during Unemployment} = f(\text{tenure at old job})$
- ▶ Individuals who start work after 1993 pay higher payroll taxes in Greece (Saez, Matsaganis & Tsakloglou, 2012)  
 $\Rightarrow \text{Payroll tax liability} = f(\text{start date})$

# Introduction

$$D_i = \begin{cases} 1 & \text{if } x_i \geq c \\ 0 & \text{if } x_i < c \end{cases}$$





# Introduction

- ▶ The treatment  $D_i$  is a ***discontinuous*** function of the *assignment variable*  $x_i$
- ▶ The outcome we care about  $y_i$  might also depend on  $x_i$
- ▶ But we assume that the relationship between  $y_i$  and  $x_i$  in the absence of the treatment would be smooth (continuous)
- ▶ So we can use a ***regression*** to estimate the relationship between  $y_i$  and  $x_i$ , and if there is a jump at  $x_i = c$ , we attribute that jump to the treatment.

# Sharp vs. Fuzzy RDD

- ▶ RDDs come in 2 basic flavors:

## 1. Sharp

- ▶ The treatment is fully determined by the assignment variable:

$$D_i = \begin{cases} 1 & \text{if } x_i > c \\ 0 & \text{if } x_i \leq c \end{cases}$$

- ▶ *Everyone* with  $x_i > c$  is treated, and *everyone* with  $x_i \leq c$  is untreated

## 2. Fuzzy

- ▶ The *probability* of being treated jumps at  $x_i = c$

$$\Pr(D_i = 1|x_i) = \begin{cases} g_1(x_i) & \text{if } x_i > c \\ g_0(x_i) & \text{if } x_i \leq c \end{cases}$$

where  $g_1(x_i) \neq g_0(x_i)$

- ▶ This ends up meaning that we use the threshold as an *instrument* for the treatment

# Outline

## The Evaluation Problem and Potential Outcomes

## Instrumental Variables

- IV In The Wild

- IV Pitfalls

  - Weak Instruments

  - Heterogeneity: LATE

## Difference in Differences

- DiD and Endogeneity

- DiD in the Wild

## Regression Discontinuity Designs

- Sharp RDD

- Fuzzy RDD

- RDD in the wild

# Sharp RDD

- ▶ When we're using OLS, IV etc. to estimate the effect of  $D_i$  on  $y_i$ , we want to control for observables  $\mathbf{X}$  that we believe are important, mostly because of OVB.
- ▶ We almost always do this *linearly*:  $y_i = \alpha + \rho D_i + \mathbf{x}_i' \beta + \varepsilon_i$ , and this linearity assumption doesn't matter a huge amount
- ▶ By contrast, when we're doing RDD, it matters a *lot* that we get the functional form for the assignment variable right (at least around the threshold).
- ▶ To see why, consider the sharp RDD setting, and assume that the model *is* linear

$$E[y_{0i}|x_i] = \alpha + \beta x_i$$

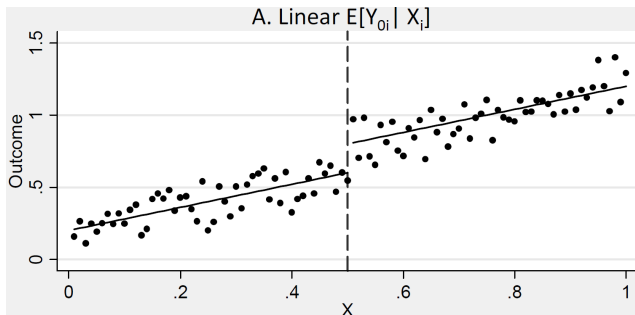
$$y_{1i} = y_{0i} + \rho$$

# Sharp RDD

- So we could run a regression

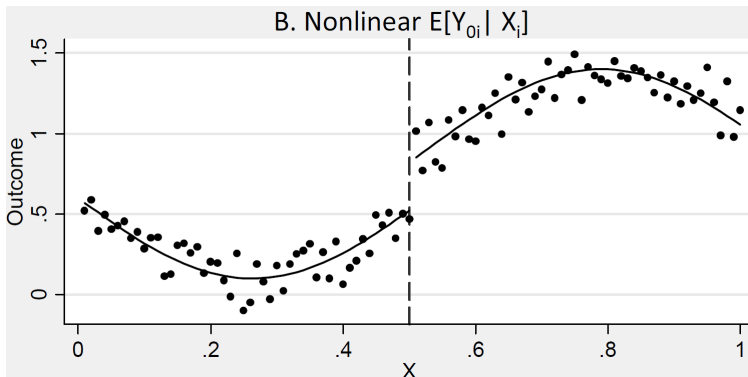
$$y_i = \alpha + \beta x_i + \rho D_i + \eta_i$$

- Difference between RDD and our usual setting:  $D_i$  is not just correlated with  $x_i$  it's a deterministic function of  $x_i$ .
- RDD will capture the causal effects by distinguishing the nonlinear, discontinuous function  $I\{x_i \geq c\}$  from the linear function  $\beta x_i$



# Sharp RDD

- ▶ But, what if the trend relationship  $E[y_{0i}|x_i]$  is nonlinear?
- ▶ Suppose it is some smooth function  $E[y_{0i}|x_i] = f(x_i)$



- ▶ We can construct our RD estimate by estimating

$$y_i = f(x_i) + \rho D_i + \eta_i$$

# Sharp RDD

- So, to be flexible, let's model  $f(x_i)$  with a  $p$ th order polynomial, so that

$$y_i = \alpha + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_p x_i^p + \rho D_i + \eta_i$$

- In fact, we can be even more flexible and allow the polynomial to be different on each side of the threshold

$$E[y_{0i}|x_i] = f_0(x_i) = \alpha + \beta_{01}\tilde{x}_i + \beta_{02}\tilde{x}_i^2 + \dots + \beta_{0p}\tilde{x}_i^p$$

$$E[y_{1i}|x_i] = f_1(x_i) = \alpha + \beta_{11}\tilde{x}_i + \beta_{12}\tilde{x}_i^2 + \dots + \beta_{1p}\tilde{x}_i^p$$

where  $\tilde{x}_i = x_i - c$  so that the coefficient on  $D_i$  in a regression gives you the treatment effect.

# Sharp RDD

- To get a regression for this that we can implement:

$$\mathbb{E}[y_i|x_i] = \mathbb{E}[y_{0i}|x_i] + (\mathbb{E}[y_{1i}|x_i] - \mathbb{E}[y_{0i}|x_i]) D_i$$

- And substituting in  $\mathbb{E}[y_{0i}|x_i] = f_0(x_i)$  and  $\mathbb{E}[y_{1i}|x_i] = f_1(x_i)$ :

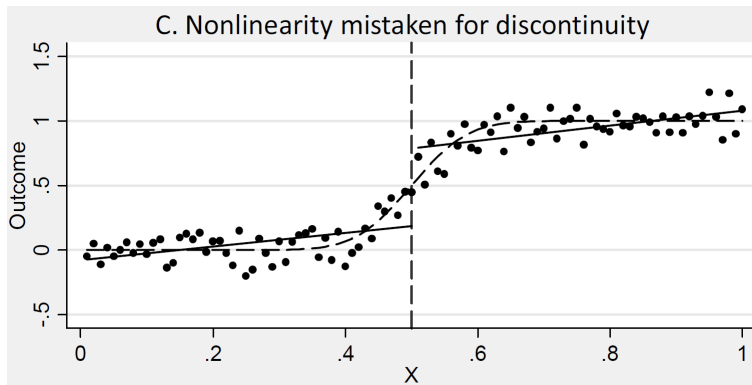
$$\begin{aligned} y_i &= \alpha + \beta_{01}\tilde{x}_i + \beta_{02}\tilde{x}_i^2 + \dots + \beta_{0p}\tilde{x}_i^p \\ &\quad + \rho D_i + (\beta_{11} - \beta_{01}) D_i \tilde{x}_i + (\beta_{12} - \beta_{02}) D_i \tilde{x}_i^2 + \dots \\ &\quad + (\beta_{1p} - \beta_{0p}) D_i \tilde{x}_i^p + \eta_i \\ &= \alpha + \beta_{01}\tilde{x}_i + \beta_{02}\tilde{x}_i^2 + \dots + \beta_{0p}\tilde{x}_i^p \\ &\quad + \rho D_i + \beta_1^* D_i \tilde{x}_i + \beta_2^* D_i \tilde{x}_i^2 + \dots + \beta_p^* D_i \tilde{x}_i^p + \eta_i \end{aligned}$$

where  $\beta_k^* = \beta_{1k} - \beta_{0k}$



# Sharp RDD

- ▶ The validity of the RD estimates will live or die by whether the polynomial “adequately” describes  $E[y_{0i}|x_i]$ : We have to get  $f(x_i)$  approx right, at least around threshold:



# Sharp RDD

- ▶ To make mistakes less likely, we can focus on data only around the threshold, say in an interval  $[c - \Delta, c + \Delta]$
- ▶ In this case

$$\mathbb{E}[y_i | c - \Delta < x_i < c] \simeq \mathbb{E}[y_{0i} | x_i = c]$$

$$\mathbb{E}[y_i | c \leq x_i < c + \Delta] \simeq \mathbb{E}[y_{1i} | x_i = c]$$

- ▶ Which implies that as we zoom in

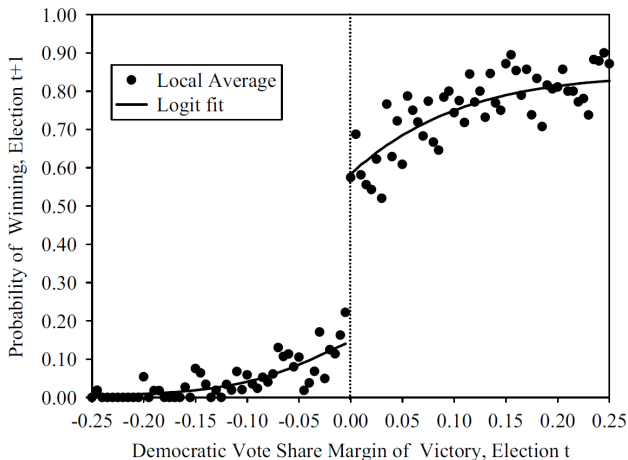
$$\lim_{\Delta \rightarrow 0} \mathbb{E}[y_i | c - \Delta < x_i < c] - \mathbb{E}[y_i | c \leq x_i < c + \Delta] \\ \mathbb{E}[y_{1i} - y_{0i} | x_i = c]$$

# Sharp RDD

- ▶ This is super simple, just compare average outcomes *just below* the threshold to average outcomes *just above* the threshold
- ▶ But:
  - ▶ working in small neighborhoods of the cutoff implies little data...
  - ▶ The sample average is biased near a cutoff (in this case  $c$ ). This can be corrected by using
    - ▶ Local linear regressions (Hahn, Todd & van der Klaauw, 2001)
    - ▶ local polynomial regressions (Porter, 2003)
    - ▶ Essentially, these give more weight to observations that are closer to the cutoff.
- ▶ In practice, people usually rely on polynomials, and then zoom in gradually and hope their estimates don't change.

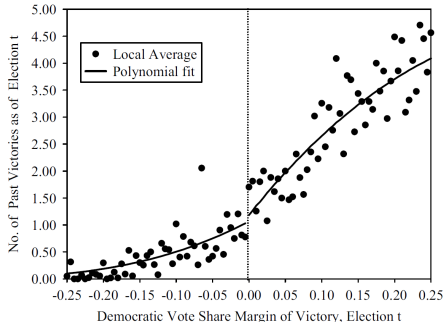
# Sharp RDD

- ▶ An example: Lee (2008) asks: “How big of an advantage to incumbents have in elections?” using a sharp RDD design
- ▶ Incumbency ( $D_i = 1$ ) is determined by the assignment variable vote share difference  $x_i$ :  $D_i = I\{x_i \geq 0\}$



# Sharp RDD

- ▶ Apart from ensuring that the functional form is flexible enough, we also want to make sure there are no omitted variables
- ▶ An omitted variable would be some other characteristic that also jumps discontinuously at  $x_i = c$ .
- ▶ What people usually do is plot RD pictures for other variables that we observe, but that shouldn't be affected by  $D_i$ .
- ▶ In Lee's case, he looks at the number of past election victories: happened before the elections being considered.



# Sharp RDD

- ▶ A second concern is that we need the assignment variable  $x_i$  to be exogenously assigned
- ▶ That is, people can't choose their  $x_i$ . If people choose their  $x_i$  they may sort onto different sides of the threshold in order to get (or avoid) the treatment
- ▶ In the Lee case, we need politicians not to manipulate vote shares in order to win the election. Hopefully, if the vote counts are independently done, this isn't a problem. But Florida 2000...
- ▶ McCrary, 2008 proposes a test for whether or not people are sorting (bunching) around the threshold by looking for discontinuities in the distribution of characteristics that shouldn't be affected by the treatment (gender?, age?, etc.)

# Outline

## The Evaluation Problem and Potential Outcomes

## Instrumental Variables

- IV In The Wild

- IV Pitfalls

  - Weak Instruments

  - Heterogeneity: LATE

## Difference in Differences

- DiD and Endogeneity

- DiD in the Wild

## Regression Discontinuity Designs

- Sharp RDD

- Fuzzy RDD

- RDD in the wild

# Fuzzy RDD

- ▶ In a fuzzy RDD, it's not the treatment itself that jumps at  $x_i = c$ , but the *probability* of receiving the treatment:

$$\Pr(D_i = 1|x_i) = \begin{cases} g_1(x_i) & \text{if } x_i > c \\ g_0(x_i) & \text{if } x_i \leq c \end{cases}$$

where  $g_1(x_i) \neq g_0(x_i)$

- ▶ In principle,  $g_1(x_i)$  and  $g_0(x_i)$  can be anything, as long as they're not the same at  $x_i = c$ . Let's assume that  $g_1(c) > g_0(c)$ .

$$E[D_i|x_i] = \Pr(D_i = 1|x_i) = g_0(x_i) + [g_1(x_i) - g_0(x_i)] T_i$$

where  $T_i = I\{x_i > c\}$



# Fuzzy RDD

- ▶ The natural way to think about this is as an IV, so let's use a 2SLS strategy.
- ▶ Describe  $g_0(x_i)$  and  $g_1(x_i)$  with  $p$ th order polynomials

$$\begin{aligned} E[D_i|x_i] &= \gamma_{00} + \gamma_{01}x_i + \gamma_{02}x_i^2 + \dots + \gamma_{0p}x_i^p \\ &\quad \left[ \pi + (\gamma_{11} - \gamma_{01})x_i + (\gamma_{12} - \gamma_{02})x_i^2 + \dots + (\gamma_{1p} - \gamma_{0p})x_i^p \right] T_i \\ &= \gamma_{00} + \gamma_{01}x_i + \gamma_{02}x_i^2 + \dots + \gamma_{0p}x_i^p \\ &\quad + \pi T_i + \gamma_1^* x_i T_i + \gamma_2^* x_i^2 T_i + \dots + \gamma_p^* x_i^p T_i \end{aligned}$$

- ▶ This means that we can use  $T_i$  and the interaction terms  $x_i T_i$ ,  $x_i^2 T_i$ ,  $\dots$ ,  $x_i^p T_i$  as instruments for  $D_i$

# Fuzzy RDD

- ▶ The simplest Fuzzy RDD uses just  $T_i$  as the instrument.
- ▶ The first stage regression is

$$D_i = \gamma_0 + \gamma_1 x_i + \gamma_2 x_i^2 + \dots + \gamma_p x_i^p + \pi T_i + \xi_{1i}$$

- ▶ Substituting into the relationship between  $y_i$ ,  $x_i$ , and  $D_i$  to get the reduced form relationship

$$\begin{aligned} y_i &= \alpha + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_p x_i^p \\ &\quad + \rho [\gamma_0 + \gamma_1 x_i + \gamma_2 x_i^2 + \dots + \gamma_p x_i^p + \pi T_i + \xi_{1i}] + \eta_i \\ &= \mu + \kappa_1 x_i + \kappa_2 x_i^2 + \dots + \kappa_p x_i^p + \rho \pi T_i + \xi_{2i} \end{aligned}$$

where  $\mu = \alpha + \rho\gamma$ , and the  $\kappa_j = \beta_j + \rho\gamma_j$

# Fuzzy RDD

- ▶ Similarly to the sharp RDD, we can zoom in around the threshold and get non-parametric estimates.
- ▶ The reduced form is

$$E[y_i | c \leq x_i < c + \Delta] - E[y_i | c - \Delta \leq x_i < c] \simeq \rho\pi$$

- ▶ While the first stage is

$$E[D_i | c \leq x_i < c + \Delta] - E[D_i | c - \Delta \leq x_i < c] \simeq \pi$$

- ▶ And so,

$$\lim_{\Delta \rightarrow 0} \frac{E[y_i | c \leq x_i < c + \Delta] - E[y_i | c - \Delta \leq x_i < c]}{E[D_i | c \leq x_i < c + \Delta] - E[D_i | c - \Delta \leq x_i < c]} = \rho$$

- ▶ The sample analog of this is a Wald estimator using  $T_i$  as an instrument for  $D_i$

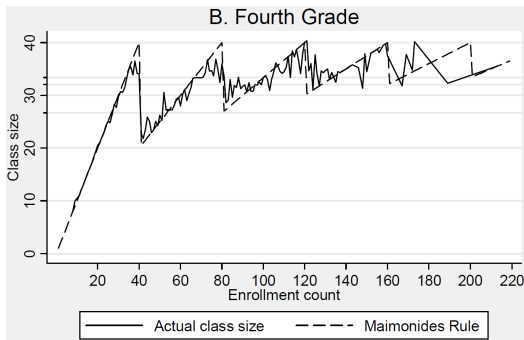
# Fuzzy RDD

- ▶ RDDs, fuzzy or sharp, are very powerful ways to get credible estimates of causal effects  $y_{1i} - y_{0i}$
- ▶ Only assumption is the smoothness of the counterfactual  $E[y_0|x_i]$
- ▶ Of course, they are *Local* average treatment effects (LATEs): The compliers are people around the threshold.
- ▶ If we think effects are heterogeneous, then we need to be careful if we want to extrapolate from RDD estimates.

# Fuzzy RDD

- ▶ Angrist & Lavy (1999) study the impact of class size on educational attainment
- ▶ “Maimonides’ rule” in Israel: class size  $< 40$  students.
- ▶ This rule implies that class  $c$ ’s size at school  $s$ ,  $m_{sc}$  as a function of enrollment  $e_s$  is

$$m_{sc} = \frac{e_s}{\text{int} \left[ \frac{e_s - 1}{40} \right] + 1}$$



# Fuzzy RDD

Table 6.2.1: OLS and fuzzy RD estimates of the effects of class size on fifth grade math scores

	OLS			2SLS				
				Full sample		Discontinuity samples		
						+/- 5		+/- 3
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean score		67.3			67.3		67.0	67.0
(s.d.)		(9.6)			(9.6)		(10.2)	(10.6)
<i>Regressors</i>								
Class size	.322 (.039)	.076 (.036)	.019 (.044)	-.230 (.092)	-.261 (.113)	-.185 (.151)	-.443 (.236)	-.270 (.281)
Percent disadvantaged		-.340 (.018)	-.332 (.018)	-.350 (.019)	-.350 (.019)	-.459 (.049)	-.435 (.049)	
Enrollment			.017 (.009)	.041 (.012)	.062 (.037)		.079 (.036)	
Enrollment squared/100					-.010 (.016)			
Segment 1 (enrollment 36-45)								-12.6 (3.80)
Segment 2 (enrollment 76-85)								-2.89 (2.41)
Root MSE	9.36	8.32	8.30	8.40	8.42	8.79	9.10	10.2
R-squared	.048	.249	.252					
N		2,018			2,018		471	302

Notes: Adapted from Angrist and Lavy (1999). The table reports estimates of equation

(6.2.6) in the text using class averages. Standard errors, reported in parentheses, are corrected for within-school correlation.

# Outline

## The Evaluation Problem and Potential Outcomes

## Instrumental Variables

- IV In The Wild

- IV Pitfalls

  - Weak Instruments

  - Heterogeneity: LATE

## Difference in Differences

- DiD and Endogeneity

- DiD in the Wild

## Regression Discontinuity Designs

- Sharp RDD

- Fuzzy RDD

- RDD in the wild

# RDD Example 1: Card, Dobkin & Maestas (2008)

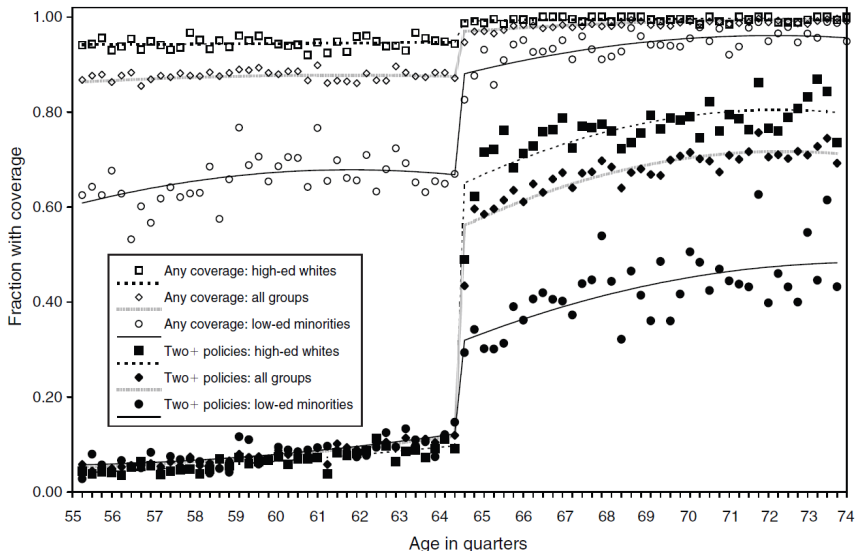


FIGURE 1. COVERAGE BY ANY INSURANCE AND BY TWO OR MORE POLICIES, BY AGE AND DEMOGRAPHIC GROUP



# RDD Example 1: Card, Dobkin & Maestas (2008)

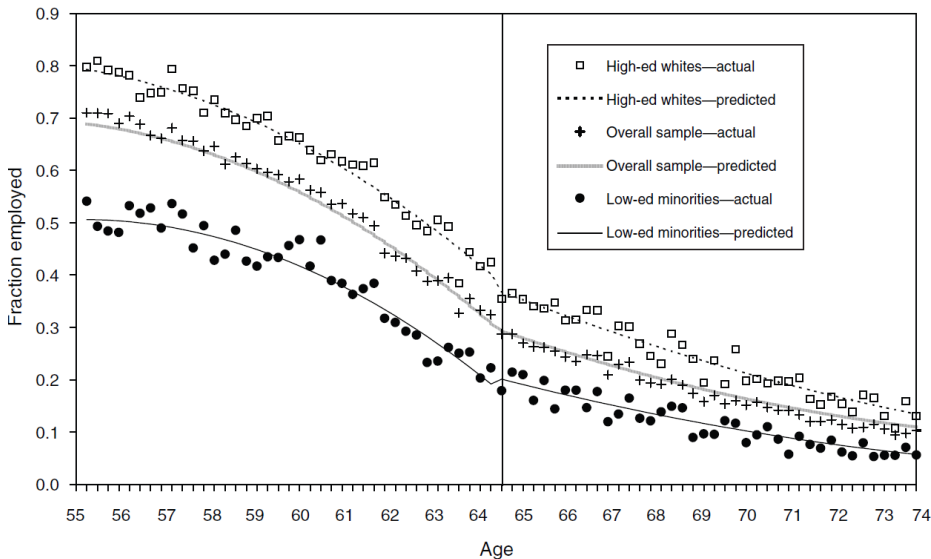


FIGURE 2. EMPLOYMENT RATES BY AGE AND DEMOGRAPHIC GROUP (1992–2003 NHIS)

# RDD Example 1: Card, Dobkin & Maestas (2008)

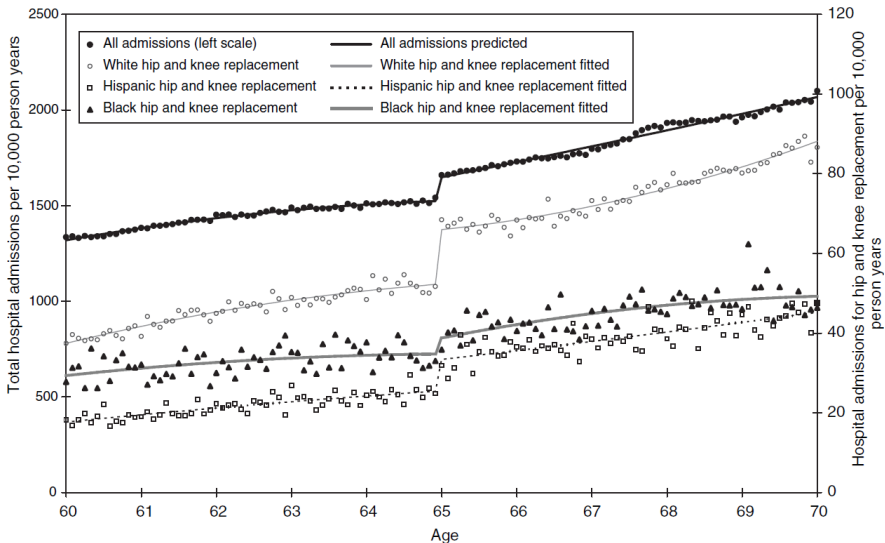
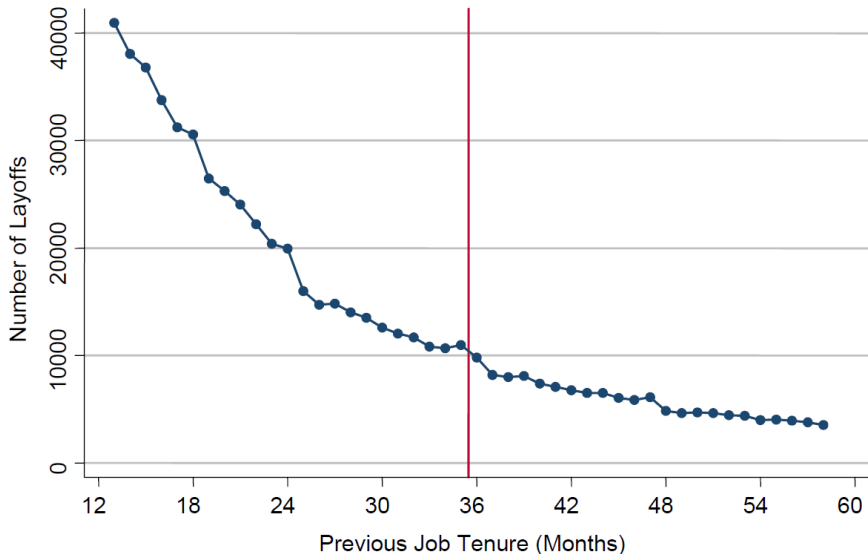


FIGURE 3. HOSPITAL ADMISSION RATES BY RACE/ETHNICITY

# RDD Example 2: Card, Chetty & Weber (2007)

Frequency of Layoffs by Job Tenure



# RDD Example 2: Card, Chetty & Weber (2007)

Effect of Severance Pay on Nonemployment Durations



# RDD Example 2: Card, Chetty & Weber (2007)

Effect of Severance Pay on Subsequent Wages

