An Introduction to Difference-in-Differences

Rajeev Dehejia

Natural/ ‘quasi’ experiments

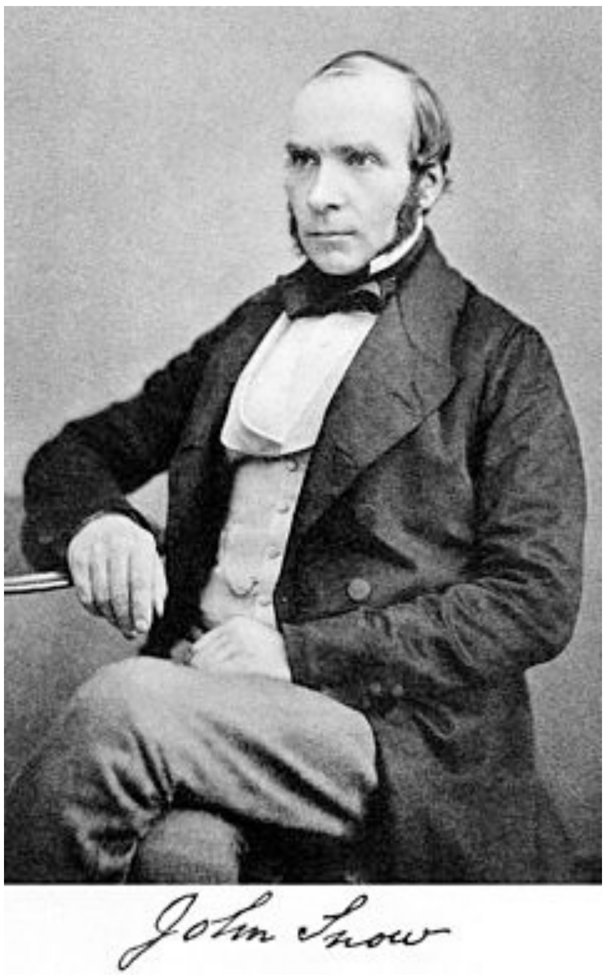
- Used to refer to situation that is not experimental but is ‘as if’ it was.
- Not a precise definition – saying your data is a ‘natural experiment’ makes it sound better.
- Refers to case where variation in T is ‘good variation’ (directly or indirectly via instrument).
- A Famous Example: London, 1854.

The case of the Broad Street pump

- Regular cholera epidemics in 19th century London.
- Widely believed to be caused by *miasma* or ‘bad air’.
- John Snow thought ‘bad water’ was the cause.
- Experimental design would be to randomly give some people good water and some bad water.
  - Ethical Problems with this…

Jo(h)n Snow

John Snow…not…



Jon Snow

Soho outbreak
August/September 1854

- People closest to Broad Street Pump most likely to die.
- But breathe same air so does not resolve air vs. water hypothesis.
- Nearby workhouse had own well and few deaths.
- Nearby brewery had own well and no deaths (workers all drank beer).

John Snow’s Dot Map: Clusters of Cholera

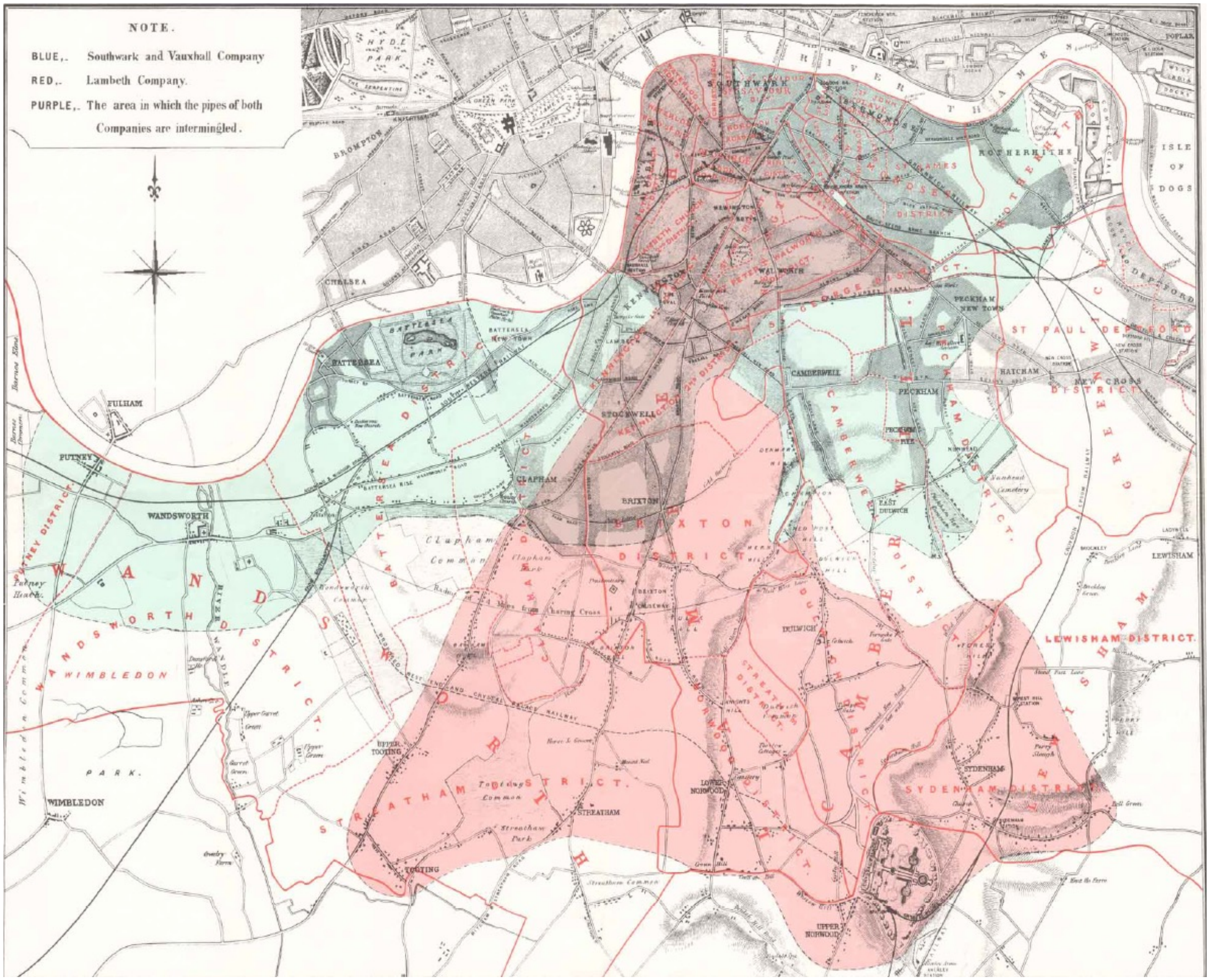
Why is this a natural experiment?

- Variation in water supply ‘as if’ it had been randomly assigned – other factors (‘air’) held constant.
- Can then estimate treatment effect using difference in means.
- Or run regression of death on water source distance to pump, other factors.
- Strongly suggests water the cause.
- Woman died in Hampstead, niece in Islington.

What’s that got to do with it?

- Aunt liked taste of water from Broad Street pump.
- Had it delivered every day.
- Niece had visited her.
- Investigation of well found contamination by sewer!
- This is non-experimental data but analyzed in a way that makes a very powerful case – no theory either.

John Snow again…



The grand experiment

- Water supplied to households by competing private companies.
- **Sometimes different companies supplied households in same street.**

Why is this important? What makes these features ideal for a natural experiment?

- In south London two main companies:
  - Lambeth Company (water supply from Thames Ditton, 22 miles upstream).
  - Southwark and Vauxhall Company (water supply from Thames).

In 1853/54 cholera outbreak

- Death Rates per 10000 people by water company:
  - Lambeth 10
  - Southwark and Vauxhall 150
- Might be water but perhaps other factors.
- Snow compared death rates in 1849 epidemic:
  - Lambeth 150
  - Southwark and Vauxhall 125
- In 1852 Lambeth Company had changed supply from Hungerford Bridge.

What would be good estimate of effect of clean water?

	1849	1853/54	Difference
Lambeth	150	10	-140
Vauxhall and Southwark	125	150	25
Difference	-25	140	-165

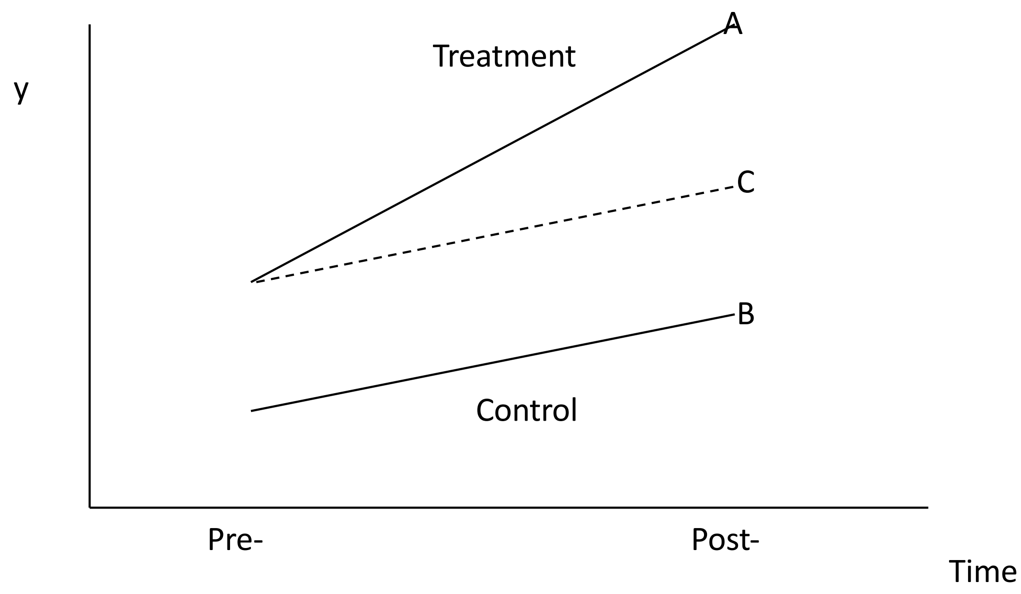
This is basic idea of differences-in-differences

- Have already seen idea of using differences to estimate causal effects.
  - Treatment/control groups in experimental data
- Often would like to find ‘treatment’ and ‘control’ group that can be assumed to be similar in every way except receipt of treatment.
- This may be very difficult to do.

A weaker assumption is…

- Assume that, in absence of treatment, difference between 'treatment' and 'comparison' group is constant over time.
- With this assumption can use observations on treatment and control group pre- and post-treatment to estimate causal effect.
- Idea:
  - Difference pre-treatment is 'normal' difference
  - Difference post-treatment is 'normal' difference + causal effect
  - Difference-in-difference is causal effect

A graphical representation



What is a d-in-d estimate?

- Standard differences estimator is AB.
- But ‘normal’ difference estimated as CB.
- Hence D-in-D estimate is AC.
- Note: assumes trends in outcome variables the same for treatment and comparison groups.
- This is not (directly) testable
- With two periods can get no idea of plausibility but can with more periods.

Difference-in-differences setup

Definition

Causal effect for unit i **at** time t is

- $Y_{1i(t)} - Y_{0i(t)}$

Observed outcomes $Y_{i(t)}$ are realized as

- $Y_{i(t)} = Y_{0i(t)}(1 - T_{i(t)}) + Y_{1i(t)}\ T_{i(t)}$
- Fundamental problem of causal inference!
- If T occurs only after $t = 0$ ($T_i = T_{i\,(1)}$ and $Y_{i(0)} = Y_{0i(0)}$) we have $Y_{i(1)} = Y_{0i(1)} \bullet (1 - T_i) + Y_{1i(1)} \bullet T_i$

Note: "t" is an observation at a unique point in time. Me today is not the same as me tomorrow, so any measurement of an outcome for the same unit at different dates/times would need to be marked as unique observations

Here, t=0 can roughly translate to "baseline observation period"

Estimand (ATET)

Focus on estimating the average effect of the treatment on the treated: $\alpha_{ATET} = E[Y_{1(1)} - Y_{0(1)} \,|\, T = 1]$

Difference-in-differences setup

Estimand (ATET)

$\alpha_{ATET} = E[Y_1(1) - Y_0(1) \mid T = 1]$

	Post-Period (t=1)	**Pre-Period (t=0)**
Treated T=1	$E[Y_1(1) \mid T = 1]$	$E[Y_0(0) \mid T = 1]$
Comparison T=0	$E[Y_0(1) \mid T = 0]$	$E[Y_0(0) \mid T = 0]$

Problem

Missing potential outcome: $E[Y_0(1) \mid T = 1]$, *i.e., what is the average post-period outcome for the treated in the absence of the treatment?*

Difference-in-differences setup

Estimand (ATET)

$\alpha_{ATET} = E[Y_1(1) - Y_0(1) \mid T = 1]$

	Post-Period (t=1)	**Pre-Period (t=0)**
Treated T=1	$E[Y_1(1) \mid T = 1]$	$E[Y_0(0) \mid T = 1]$
Control T=0	$E[Y_0(1) \mid T = 0]$	$E[Y_0(0) \mid T = 0]$

Strategy: Before vs. After

- Use: $E[Y(1) \mid T = 1] - E[Y(0) \mid T = 1]$
- Assumes: $\color{red}{E[Y_0(1) \mid T = 1]} = E[Y_0(0) \mid T = 1]$

Difference-in-differences setup

Estimand (ATET)

$\alpha_{ATET} = E[Y_1(1) - Y_0(1) \,|\, T = 1]$

	Post-Period (t=1)	**Pre-Period (t=0)**
Treated T=1	$E[Y_1(1) \,|\, T = 1]$	$E[Y_0(0) \,|\, T = 1]$
Control T=0	$E[Y_0(1) \,|\, T = 0]$	$E[Y_0(0) \,|\, T = 0]$

Strategy: Treated vs. Control in Post-Period

- Use: $E[Y(1) \,|\, T = 1] - E[Y(1) \,|\, T = 0]$
- Assumes: $E[Y_0(1) \,|\, T = 1] = E[Y_0(1) \,|\, T = 0]$

Difference-in-differences setup

Estimand (ATET)

$\alpha_{ATET} = E[Y_1(1) - Y_0(1) \,|\, T = 1]$

	Post-Period (t=1)	**Pre-Period (t=0)**
Treated T=1	$E[Y_1(1) \,|\, T = 1]$	$E[Y_0(0) \,|\, T = 1]$
Control T=0	$E[Y_0(1) \,|\, T = 0]$	$E[Y_0(0) \,|\, T = 0]$

Strategy: Difference-in-Differences

- Use:

  $\{E[Y(1) \,|\, T = 1] - E[Y(1) \,|\, T = 0]\} -$

  $\{E[Y(0) \,|\, T = 1] - E[Y(0) \,|\, T = 0]\}$
- Assumes: $E[Y_0(1) - Y_0(0) \,|\, T = 1] = E[Y_0(1) - Y_0(0) \,|\, T = 0]$

Causal model

- Y(i,j), where i=1/0 treated/control and j=1/0 post/pre.
  - $Y(1,0) = \alpha$
  - $Y(1,1) = \alpha + \tau + \delta$
  - $Y(0,0) = \beta$
  - $Y(0,1) = \beta + \tau$
- Assumes τ same across treatment and control.

Statistical model

- Y(i,j), where i=1/0 treated/control and j=1/0 post/pre.
  - $Y = a + b\text{Treated} + c\text{Post} + d\ \text{Treated x Post} + e$
  - $Y_{(0,0)} = \beta = a \rightarrow \boldsymbol{a=\beta}$
  - $Y_{(1,0)} = \alpha = a+b \rightarrow a+b=\alpha \rightarrow \boldsymbol{b= \alpha - a = \alpha - \beta}$
  - $Y_{(0,1)} = \beta + \tau = a + c \rightarrow c = \beta + \tau - a = \beta + \tau - \boldsymbol{\beta} \boldsymbol{=} \tau$
  - $Y_{(1,1)} = \alpha + \tau + \delta = a+b+c+d$

    $\rightarrow d = \alpha + \tau + \delta - a - b - c = \alpha + \tau + \delta - \boldsymbol{\beta} \boldsymbol{-} \boldsymbol{\alpha} \boldsymbol{+} \boldsymbol{\beta} \boldsymbol{-} \tau$

    $\rightarrow d = \alpha + \tau + \delta - \boldsymbol{\beta} \boldsymbol{-} \boldsymbol{\alpha} \boldsymbol{+} \boldsymbol{\beta} \boldsymbol{-} \tau = \delta$
- Treatment effect: $Y_{(1,1)} - Y_{(0,1)} - (Y_{(0,1)} - Y_{(0,0)}) = \delta$.
- Assumes τ same across treatment and control.

Extension 1: Panel Data

- Suppose you have a unit-level panel, i.e., observe not only treatment and comparison groups in pre and post period, but same units before and afterwards.
- Then you can estimate

$$Y = a + b\text{Treated}_{t=2} + c\text{Post} + d\ \text{Treated x Post} + e \quad \text{if } t=2$$
$$\rightarrow Y = (a+c) + (b+d)Treated + e$$

$$Y = a + b\text{Treated}_{t=2} + c\text{Post} + d\ \text{Treated x Post} + e \quad \text{if } t=1$$
$$\rightarrow Y = a + bTreated + e$$

$$\Delta Y = c + d\ \text{Treated} + (e_{t=2} - e_{t=1})$$

Extension 1: Panel Data

- This is simply 'differences' estimator applied to the difference.
- To implement this need to have repeat observations on the same individuals.
- May not have this – individuals observed pre- and post-treatment may be different.

A comparison of the two methods

- Where have repeated observations could use both methods.
- Will give same parameter estimates.
- But will give different standard errors.
- ‘Levels’ version will assume residuals are independent – unlikely to be a good assumption.
- Can deal with this by:
  - Clustering
  - Or estimating ‘differences’ version

Extension 2: other regressors

- Can put in other regressors as before.
- Perhaps should think about way in which they enter the estimating equation.
- E.g. if level of W affects level of y then should include ΔW in differences version.
- Another issue may be if your treatment is *selected* by participants then only the worst off individuals elect the treatment—not comparable to general effect of policy.
  - Covariates can help with this.

Extension 3: testing differential trends in treatment and control groups

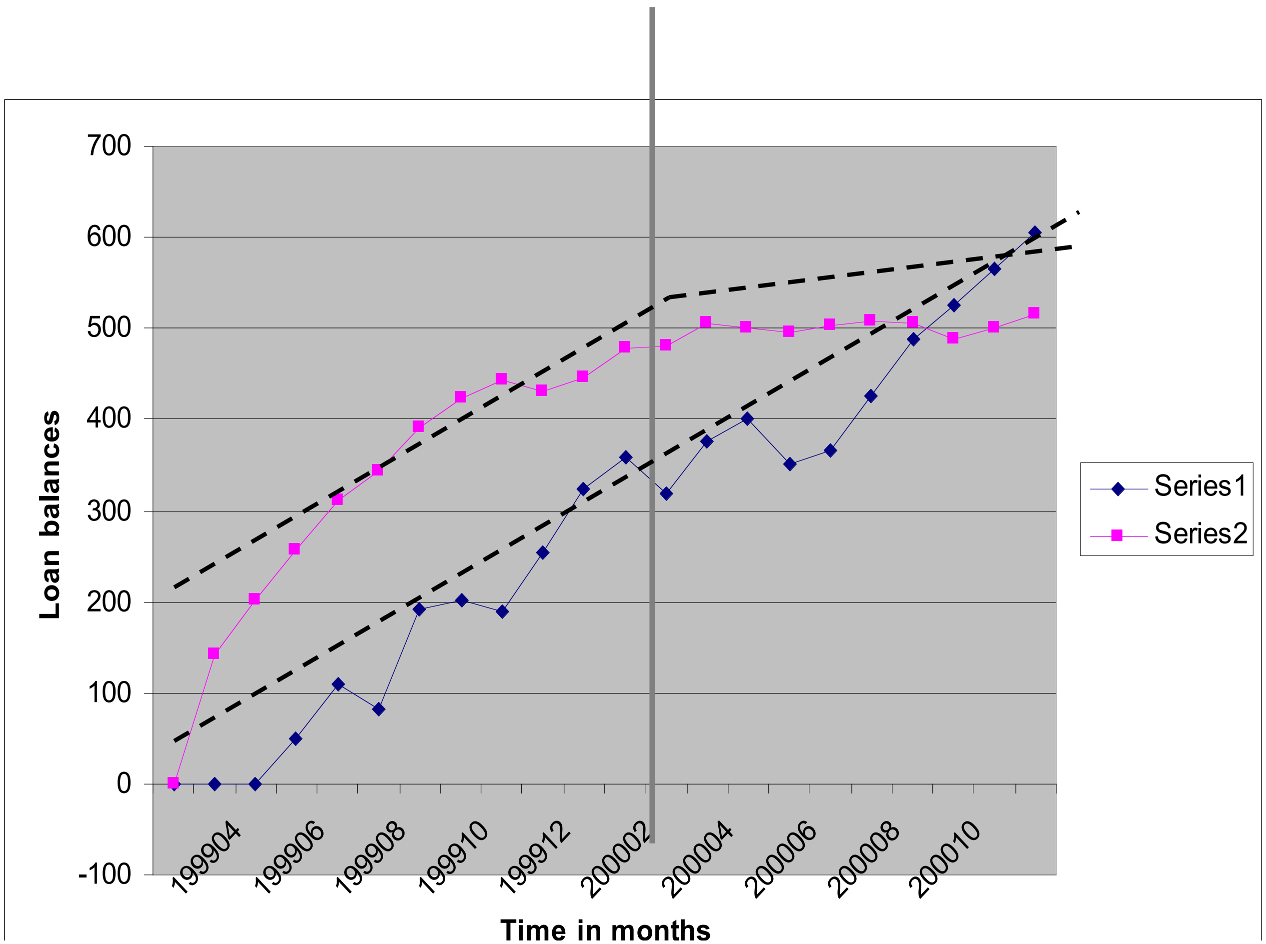
- Key assumption underlying validity of D-in-D estimate is that differences between treatment and control group would have remained constant in absence of treatment.
- Can never test this.
- With only two periods can get no idea of plausibility.
- But can with more than two periods.

Example 1: Do interest rates matter?

Dehejia, Morduch, Montgomery

- Looking at the effect of interest rates on loan demand and banking activity.
- It has been argued that poor borrowers will not sensitive, because they are so credit constrained.
- We investigated empirically.

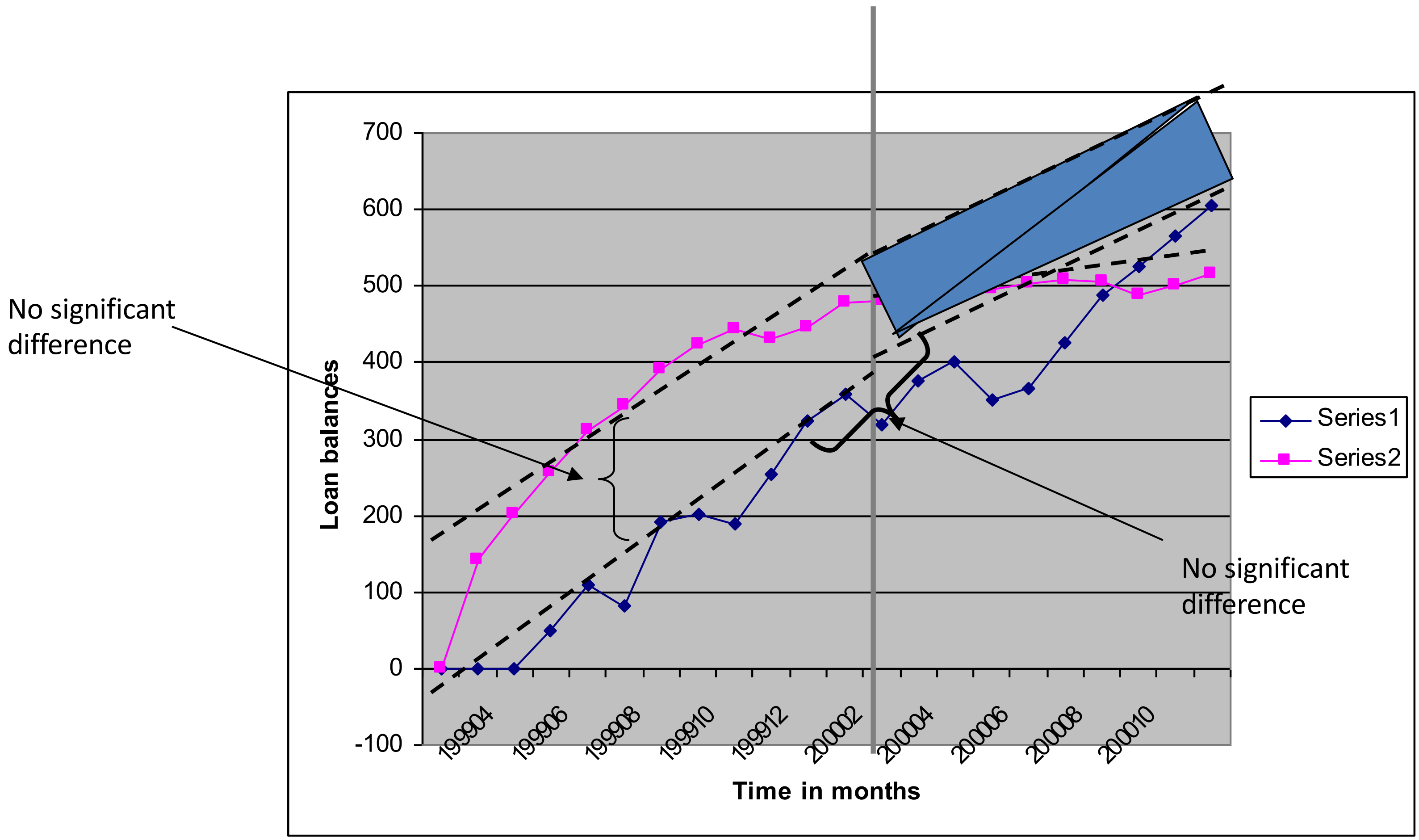
Trends



700
600
500
400
300
200
100
0
-100
Loan balances
199904
199906
199908
199910
199912
200002
200004
200006
200008
200010
Time in months
Series1
Series2

Trend

Common linear trend in pre-period (regression test)



Average loan balances, *Safe*Save

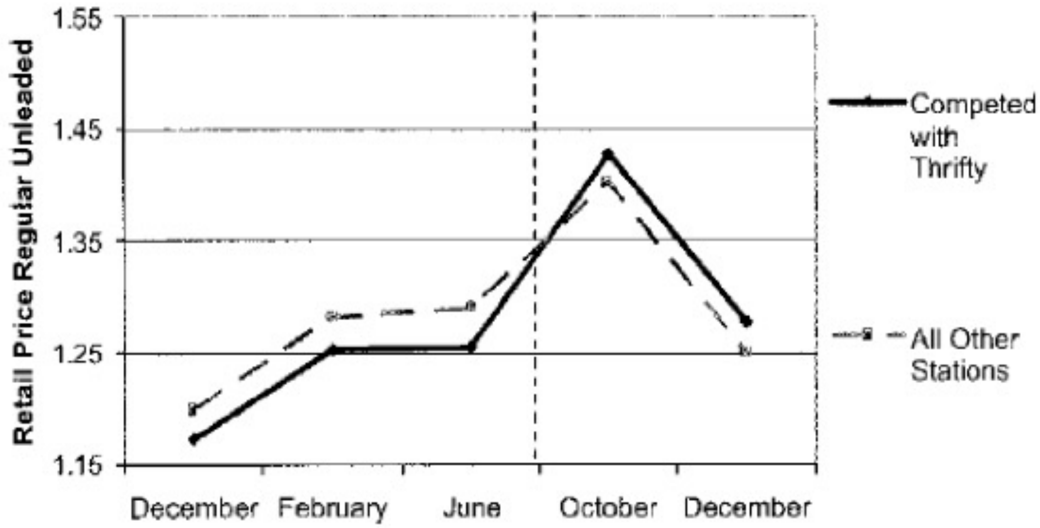
Example 2

- "Vertical Relationships and Competition in Retail Gasoline Markets", by Justine Hastings, *American Economic Review*, 2004
- Interested in effect of vertical integration on retail petrol prices

- Investigates take-over in CA of independent 'Thrifty' chain of petrol stations by ARCO (more integrated)
  - Treatment Group: petrol stations < 1mi from 'Thrifty'
  - Control group: petrol stations > 1mi from 'Thrifty'

- Lots of reasons why these groups might be different so D-in-D approach seems a good idea

This picture contains relevant information…



(a) Los Angeles

- Can see D-in-D estimate of +5c per gallon.

- Also can see trends before and after change very similar – D-in-D assumption valid.

Extension 4: Ashenfelter’s dip

- `Pre-program ***dip***', for participants
  - Related to the idea of *mean reversion*: individuals experience some idiosyncratic shock.
  - May enter program when things are especially bad.
  - Would have improved anyway (reversion to the mean).
  - Might be able to deal with this with longer earnings history and <u>not</u> using one-period-pre-treatment earnings in comparison.

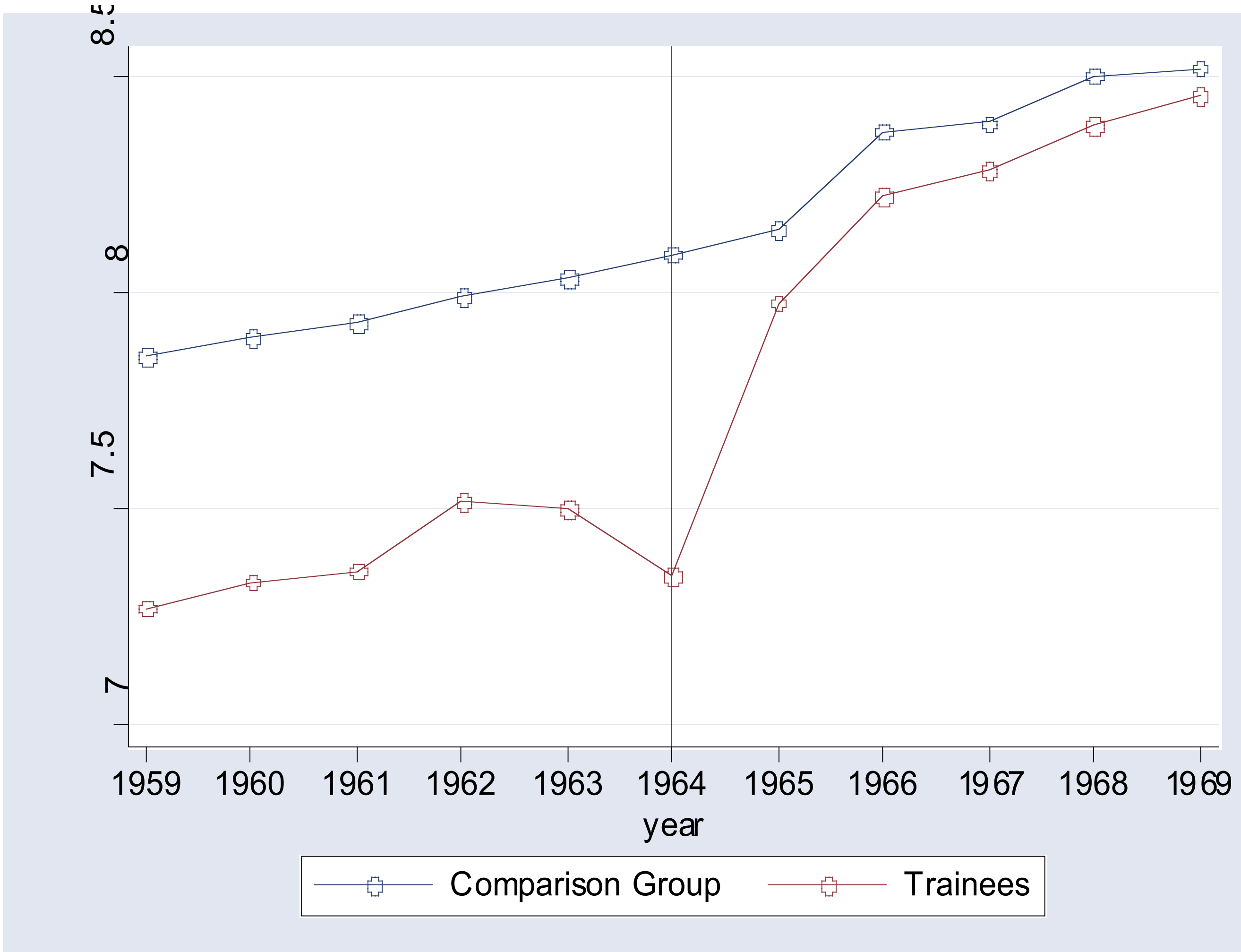
Example 3: the dip

- Interested in effect of government-sponsored training (MDTA) on earnings

- Treatment group are those who received training in 1964

- Control group are random sample of population as a whole

Earnings for period 1959-69



Things to note..

- Earnings for trainees very low in 1964 as training not working in that year – should ignore this year.

- Simple D-in-D approach would compare earnings in 1965 with 1963.

- But earnings of trainees in 1963 seem to show a 'dip' – so D-in-D assumption probably not valid.

- Probably because those who enter training are those who had a bad shock (e.g. job loss).

Extension 5

- *Compositional differences*: In repeated cross-sections we do not want that the composition of the sample changes between periods.
  - Test: Distribution of (T, X) should be similar for the pre-treatment and post-treatment periods.
  - You can control in the regression.

Extension 6

- Clustered standard errors

```
reg yvar xvar j xvar*j,
cl(groupvar)
```

*In **stata,** these are the "bear bones" of a diff-in-diff regression with levels and clustered standard errors.*

- *yvar = outcome variable*
- *xvar = treatment group variable*
- *j = time period variable (i.e. before (0), after (1))*
- *xvar*j = "levels" interaction*
- *groupvar = variable 'above' the unit of observation that you expect influences outcomes for units therein; common example: if observations are recorded at the student level throughout a large school, you'd want to cluster standard errors at the classroom level to adjust for likely correlation between classroom membership and residuals*

Extension 7

- DDD

Example 4:
Mandated Maternity Benefits (Gruber, 1994)

TABLE 3—DDD ESTIMATES OF THE IMPACT OF STATE MANDATES ON HOURLY WAGES

Location/year	Before law change	After law change	Time difference for location
A. *Treatment Individuals: Married Women, 20–40 Years Old:*			
Experimental states	1.547 (0.012) [1,400]	1.513 (0.012) [1,496]	−0.034 (0.017)
Nonexperimental states	1.369 (0.010) [1,480]	1.397 (0.010) [1,640]	0.028 (0.014)
Location difference at a point in time:	0.178 (0.016)	0.116 (0.015)	
Difference-in-difference:	−0.062 (0.022)		
B. *Control Group: Over 40 and Single Males 20–40:*			
Experimental states	1.759 (0.007) [5,624]	1.748 (0.007) [5,407]	−0.011 (0.010)
Nonexperimental states	1.630 (0.007) [4,959]	1.627 (0.007) [4,928]	−0.003 (0.010)
Location difference at a point in time:	0.129 (0.010)	0.121 (0.010)	
Difference-in-difference:	−0.008: (0.014)		
DDD:	**−0.054** **(0.026)**		

Extension 8

- In some cases you may have not only individual unit data (so a true panel) but your data may span many time periods, with the treatment not necessarily starting at the same point in time for different units.

Example 5
Dehejia and Lleras-Muney

- Look at the effect of bank branching laws and deposit insurance by states on a range of outcomes at the state level (agricultural, manufacturing output).
- Two challenges:
  - States adopt banking laws at different points in time.
  - Even then worry that *timing* of adoption of bank laws is not random (i.e., differential trends).

Table 1

Presence of State Branching (Selected Years) and Deposition Insurance Regulations

State	Branching 1900	1909	1919	1929	1939	Deposit Insurance
Alabama					x	
Arizona	x	x	x	x	x	
Arkansas	x				x	
California	x	x	x	x	x	
Colorado						
Connecticut					x	
Delaware	x	x	x	x	x	
Florida	x	x	x			
Georgia	x	x	x	x	x	
Idaho					x	
Illinois						
Indiana					x	
Iowa					x	
Kansas						1909–29
Kentucky		x	x	x	x	
Louisiana	x	x	x	x	x	
Maine	x	x	x	x	x	
Maryland			x	x	x	
Massachusetts		x	x	x	x	
Michigan	x			x	x	
Minnesota						
Mississippi				x	x	1914–30
Missouri						
Montana					x	
Nebraska						1911–30
Nevada	x				x	
New Hampshire						
New Jersey				x	x	
New Mexico					x	
New York	x	x	x	x	x	
North Carolina	x			x	x	
North Dakota					x	1917–29
Ohio				x	x	
Oklahoma						1908–23
Oregon	x	x	x		x	
Pennsylvania			x	x	x	
Rhode Island	x	x	x	x	x	
South Carolina				x	x	
South Dakota					x	1916–27
Tennessee	x	x	x	x	x	
Texas						1910–27
Utah					x	
Vermont				x	x	
Virginia				x	x	
Washington	x	x	x		x	1917–21
West Virginia						
Wisconsin	x				x	
Wyoming						

Solution to different timing of *T*

- So *T* is time varying, i.e., T_{it}.
- Use state and year **fixed effects**.
  - State FE means we are looking at effect of *T* within states over time.
  - Year FE means we are taking out common time trends across treatment and control.
  - So really diff-in-diffs.

Basic results

Table 5

Effect of Banking Laws on Economic Outcomes

	Deposit Insurance: Before 1920	Deposit Insurance: After 1920	Branching	N	Within-State R^2
Agricultural outcomes:					
Log number of farms	−.046	.044	−.054		
	(.061)	(.051)	(.040)	336	.21
Log acres agricultural land	−.14**	.082**	−.057*		
	(.041)	(.034)	(.029)	336	.09
Value of machines per acre	1,354**	−642**	462		
	(413)	(271)	(552)	288	.56
Value of crops per farm	−2.57e−04**	−1.52e−04	6.13e−05		
	(1.05e−04)	(2.91e−04)	(1.01e−04)	237	.10
Cash receipts per farm			.49**		
			(.12)	96	.60
Manufacturing outcomes:					
Employment per establishment	8.35**	−4.71*	5.44**		
	(2.67)	(2.08)	(2.30)	672	.48
Log of real annual wage earnings per worker	.051	−.056^{+}	.030*		
	(.033)	(.030)	(.015)	672	.99
Value added per capita	.025**	−.049**	.027**		
	(.007)	(.010)	(.009)	672	.50
Human capital outcomes:[a]					
Males aged 10–15 working and not in school (%)	.013**	−.002*	−.013**		
	(.004)	(8.28e−04)	(.005)	240	.61
Females aged 10–15 working and not in school (%)	.005	−.001^{+}	−.005		
	(.003)	(6.70e−04)	(.004)	240	.37

Note. Results are for fixed effects regressions that are weighted using state population, include state and year fixed effects, do not include controls, and are clustered at the state level. Nominal values are deflated using the wholesale price index, base 1947–49. Robust standard errors are in parentheses.

Solutions to second challenge: predicting entry into laws

Table 3

Predicting Passage of Branching and Insurance Laws, Linear Probability Models

Dependent Variable	Branching (1)	Insurance (2)	Branching (3)	Insurance (4)	Branching (5)	Insurance (6)	Branching (7)	Insurance (8)	Branching (9)	Insurance (10)	Branching (11)	Insurance (12)
Percent urban	.46	−.31*	−.99*	.064	.050	−.63	−.59	−.88^{+}	−.43	−.72^{+}	−.18	−.65
	(.31)	(.14)	(.46)	(.21)	(.98)	(.43)	(.85)	(.47)	(.86)	(.40)	(.97)	(.44)
Population (millions)	.0256	.00438	−.0183	.0114	−.00405	.00534	−.00541	.00805	−.0275	−.00046	.0153	.00449
	(.0188)	(.00737)	(.0160)	(.00832)	(.0767)	(.0123)	(.0597)	(.0117)	(.0672)	(.00958)	(.0734)	(.0130)
Log lagged deposits per state bank			.35**	−.072^{+}	−.001	.16**					−.054	.15**
			(.059)	(.044)	(.16)	(.064)					(.15)	(.063)
Lagged banks per square mile			−4.45	−.88	2.96	−9.08					5.65	−7.80
			(6.22)	(2.18)	(12.7)	(6.92)					(13.0)	(6.47)
Indicator for credit contractions			.022	.009	−.039	.037**					−.045^{+}	.036**
			(.025)	(.009)	(.026)	(.015)					(.026)	(.015)
Proportion of state banks					1.28^{+}	1.29*					1.04^{+}	1.25*
					(.66)	(.60)					(.61)	(.59)
Ratio of state to national bank capital/asset ratios					.007	.020					.021	.025
					(.13)	(.048)					(.11)	(.049)
Lagged growth of value added per firm							−.086**	−.013			−.082*	−.003
							(.034)	(.019)			(.037)	(.020)
Lagged employment per firm							.008*	.002**			.007*	.001
							(.003)	(9.50e-04)			(.003)	(7.13e−04)
Lagged farm size							.28	−.084			.34	−.037
							(.39)	(.26)			(.38)	(.20)
Indicator of Democratic control									.018	.050	.044	.050
									(.064)	(.031)	(.057)	(.030)
Governor and legislature controlled									−.012	−.001	−.016	−.005
									(.042)	(.013)	(.038)	(.012)
N	1,968	1,968	1,920	1,920	1,920	1,920	1,920	1,920	1,968	1,968	1,920	1,920
R^2	.12	.06	.35	.11	.67	.53	.68	.48	.66	.46	.69	.53

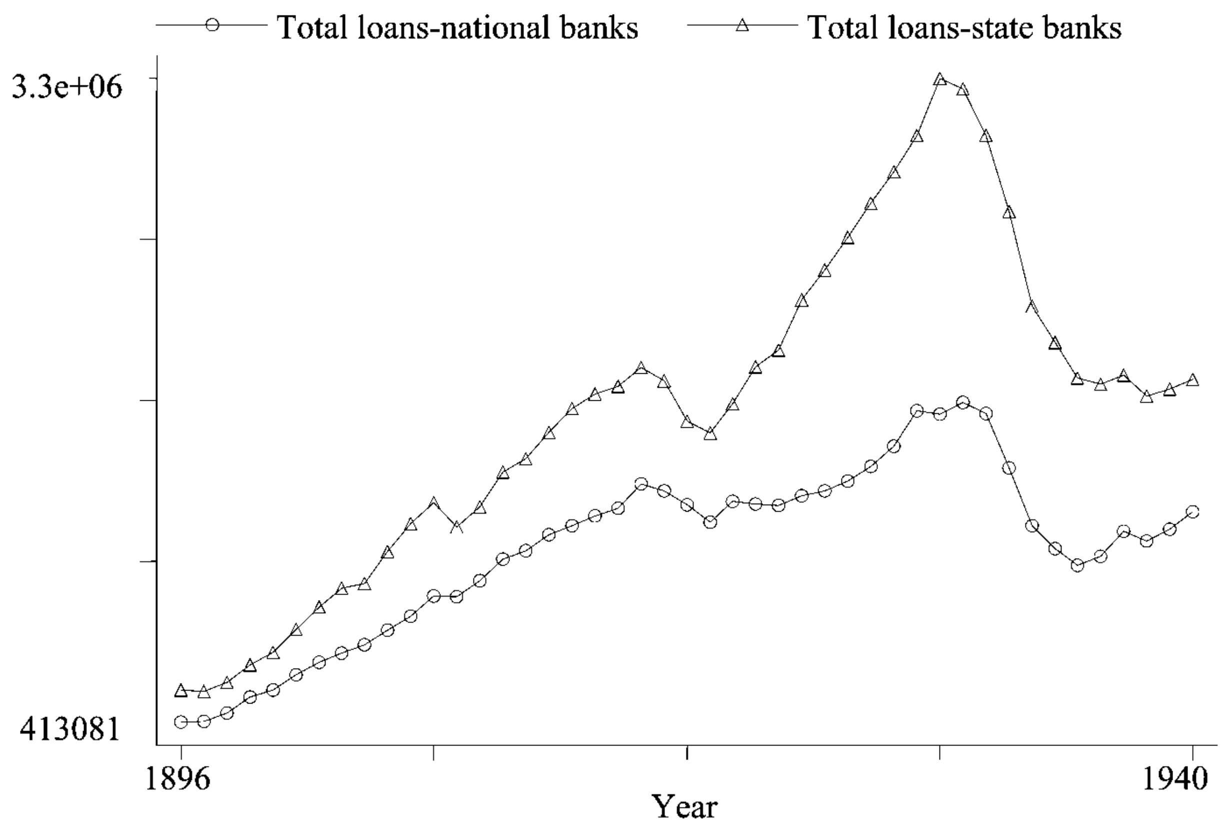


Figure 2. Evolution of loans, by national and state banks. All monetary values are deflated using the wholesale price index. The base period is 1947–49.

Controlling for national banks

Table 6

Fixed Effects Results with National Bank Controls, 1930 and Earlier

	Deposit Insurance				
	Before 1920	After 1920	Branching	N	Within-State R^2
Bank outcomes:					
Growth of national bank loans	−.003	−.020	.007	1,488	.59
	(.008)	(.016)	(.010)		
Growth of state bank loans	.029*	−.14**	.031**	1,488	.62
	(.016)	(.026)	(.010)		
Agricultural outcomes:					
Log number of farms	−.006	.051	−.12**	240	.34
	(.050)	(.053)	(.024)		
Log acres agricultural land	−.051$^+$	.085*	−.12**	240	.18
	(.030)	(.036)	(.024)		
Value of machines per acre	748**	−708**	562	240	.63
	(287)	(275)	(567)		
Value of crops per farm	−2.27e-04	−1.51e−04	1.26e−04	189	.15
	(1.55e−04)	(3.12e-04)	(1.68e−04)		
Cash receipts per farm			.37**	96	.28
			(.15)		
Manufacturing outcomes:					
Employment per establishment	2.38	−4.90*	6.26**	432	.53
	(2.28)	(2.23)	(1.73)		
Log of real annual wage earnings per worker	−.004	−.060$^+$	.044**	432	.99
	(.016)	(.034)	(.018)		
Value added per capita	.019$^+$	−.050**	.026*	432	.56
	(.011)	(.011)	(.013)		
Human capital outcomes:[a]					
Males aged 10–15 working and not in school (%)	−.014	.014$^+$	−.002*	192	.62
	(.011)	(.008)	(.001)		
Females aged 10–15 working and not in school (%)	−.001	.002	−.001	192	.34
	(.008)	(.005)	(8.57e−04)		

Note. Regressions are weighted using state population, include state and year fixed effects, are clustered at the state level, and control for the growth of national bank assets. Nominal values are deflated using the wholesale price index, base 1947–49. Robust standard errors are in parentheses.

By the way, what's the "first stage"

- Not IV so no formal first stage.
- But there is an implicit first stage. If we are to believe that banking laws affect farms and manufacturing firms, we must believe that there is a direct effect of banking laws on bank behavior.

The “first stage”

Table 4

Effect of Banking Laws on Banking Outcomes

Dependent Variable	States That Never Repealed: Branches (per Million Residents) (1)	States That Never Repealed: Banks That Branch (per Million Residents) (2)	States That Repealed: Branches (per Million Residents) (3)	States That Repealed: Banks That Branch (per Million Residents) (4)	Growth Rate of Loans (5)	Growth Rate of Loans, Excluding Outliers (6)	Growth Rate of Loans (7)
Branching passed	10.1**	4.27**					
	(3.22)	(1.81)					
Branching allowed	14.6**	6.28*					
	(4.40)	(2.74)					
Branching repealed			−19.5*	−4.88**			
			(9.85)	(.77)			
Branching not allowed			−1.8	−2.88			
			(8.31)	(2.37)			
Deposit insurance before 1920					.037**	.033**	
					(.015)	(.010)	
Deposit insurance after 1920					−.14**	−.12**	
					(.028)	(.020)	
Branching					.021*	.022**	
					(.011)	(.009)	
Years with deposit insurance							−.019*
							(.009)
(Years with deposit insurance)^2							.002
							(.001)
Years with branching							.011**
							(.005)
(Years with branching)^2							−.0013**
							(5.65e−04)
N	140	140	56	56	1,968	1,932	1,968
R^2	.70	.49	.67	.54	.62	.63	.62

Difference-in-differences: summary

- A very useful and widespread approach
- Validity does depend on assumption that trends would have been the same in absence of treatment
- Can use other periods to see if this assumption is plausible or not
- Uses 2 observations on same unit (not even individual) – most rudimentary form of panel data

Additional resources

Goings on in #EconTwitter…

- Tulane econ prof. Patrick Button made a (technical!) but handy flow chart explaining clustering and SE issues for his PhD students
- Seattle U prof. Nick C. Huntington-Klein makes cool animations of graphs, like this one showing the impact of adding different fixed effects to a model
- David McKenzie at the World Bank recently wrote a (very accessible) blog about how different papers demonstrate the parallel trends assumption

Current research and discussion

- Andrew Goodman-Bacon (Vanderbilt) and Jan Marcus (Universität Hamburg) advocate for the role of **diff-in-diff designs in evaluating COVID policies**. This is a good read for anyone interested in understanding the threats to validity and opportunities in COVID policy evaluations. "Using Difference-in-Differences to Identify Causal Effects of COVID-19 Policies" (May 2020)

- Not so recent (2001) but great example of d-in-d: Esther Duflo's "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment"

- On policies that seek to reduce discrimination by preventing employers from conducting background checks: "Does 'Ban the Box' Help or Hurt Low-Skilled Workers? Statistical Discrimination and Employment Outcomes When Criminal Histories are Hidden" Jennifer Doleac, Benjami Hansen (2016)

What we'll cover in recitation

- ✓ Stata code and interpretation of output for Diff-in-Diff
- ✓ Questions and essential concepts for problem set and replication (please prepare questions)
- ✓ Overview of readings/ discussion of methods and key take-aways (come prepared with questions or discussion points!)
- ✓ More discussion on Diff-in-Diff in the field/employers that use this evaluation experience
- ✓ Not covered in recitation, but if you are in the quantitative capstone, it's likely that you'll use this design. Please be in touch if you'd like to chat

Email Michelle at dimarm01@nyu.edu with questions beforehand if desired

Extra topics – won't be discussed in class
Extension 9

- When you have multiple pre- and post-periods you can estimate what is known as an event study model.
- You can do this for a pre-post and D-i-D
- Pre-post intuition: estimate the time fixed effect period by period.
- D-i-D intuition: estimate time fixed effects and time fixed effects x treatment indicator

Extension 9

Formally

$$Y = a + b\text{Treated} + c\text{Post} + d\ \text{Treated x Post} + e$$

becomes…

$$Y = a + c_1 T_1 + \ldots + c_R T_R + c_{R+1} T_{R+1} + \ldots + c_F T_F +$$
$$d_1 T_1 \text{xTr} + \ldots + d_R T_R \text{xTr} + d_{R+1} T_{R+1} \text{xTr} + \ldots + d_F T_F \text{xTr} + e$$

where T_x are time dummies and time runs from 1 to F, and treatment kicks in at R+1

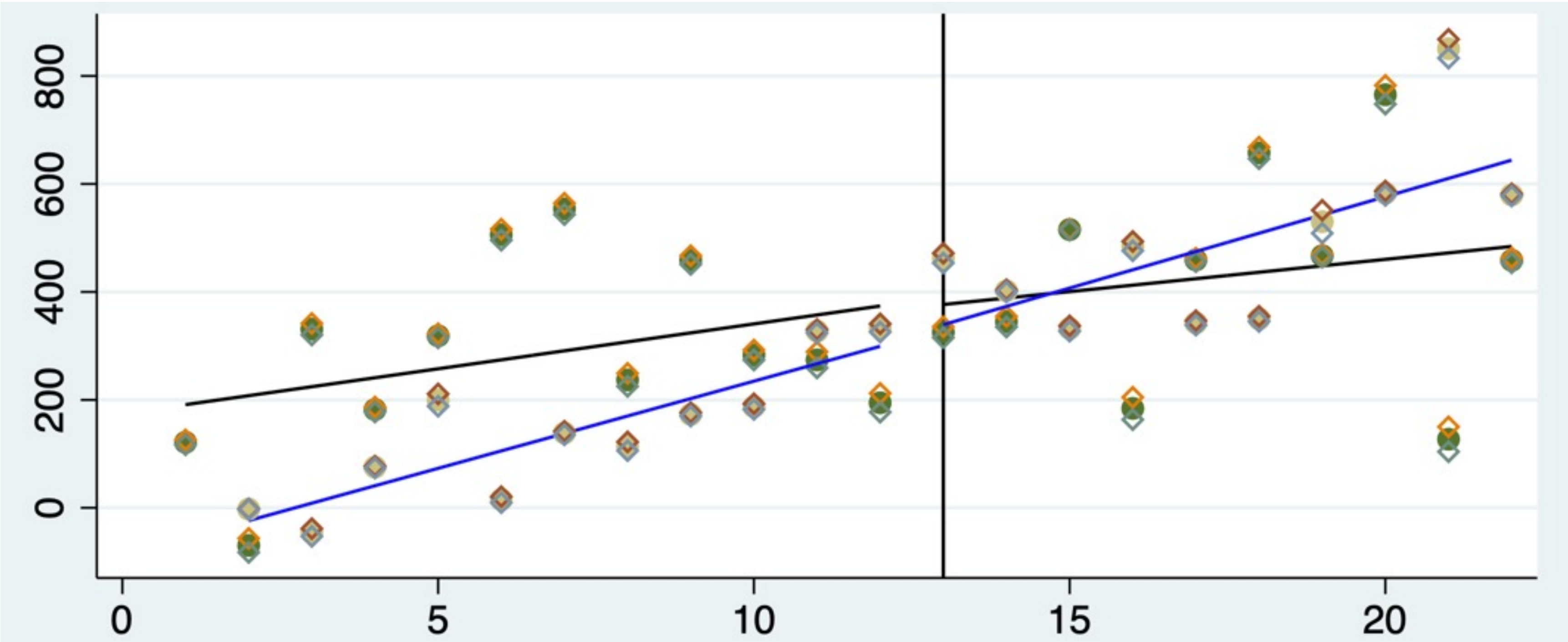
Extension 9

$$Y = a + c_1 T_1 + \ldots + c_R T_R + c_{R+1} T_{R+1} + \ldots + c_F T_F + d_1 T_1 \mathrm{xTr} + \ldots + d_R T_R \mathrm{xTr} + d_{R+1} T_{R+1} \mathrm{xTr} + \ldots + d_F T_F \mathrm{xTr} + e$$

- Comparing $c_1 \ldots c_R$ with $d_1 \ldots d_R$ compares pre-trends
- Comparing $c_{R+1} \ldots c_F$ with $d_{R+1} \ldots d_f$ gives the treatment effect

Extension 9

- SafeSave example looks a bit like this



Extension 10

- It turns out that the two-way fixed effects approach creates its own challenges.
- In particular, one has to extend the assumptions for this to work.
- Since these assumptions often don't hold, there is a new set of methods for DiD.

The problem, in short

Augment the causal model with covariates

- Y(i,j), where i=1/0 treated/control and j=1/0 post/pre.
  - $Y(1,0) = \alpha + \theta_0 X$
  - $Y(1,1) = \alpha + \tau + \delta + \theta_1 X$
  - $Y(0,0) = \beta + \theta_0 X$
  - $Y(0,1) = \beta + \tau + \theta_1 X$
- Assumes τ same across treatment and control.
- Even then our usual DiD depends on $\theta_1 X - \theta_0 X$

The problem, in short

Augment the causal model with covariates

- Gets even worse when we consider heterogenous treatment effects by group

Many solutions have been proposed

- One simple one is to estimate the year-group FE in a first step, and then estimate the model
  - Intuition: the fact that the FE are estimated on the whole sample means that there isn't a clean pre-post-treatment world – some of the treatment can be baked into the FE

First stage

- Estimation: First stage

$$Y_{git} = \lambda_g + \gamma_t + \varepsilon_{gi}t$$

using only $T_g = 0$ (non-treatment group), retaining the fixed effects. Collect the λ_g and γ_t .

Second stage

- Estimation: Second stage
- $\hat{y}_{git} = y_{git} - \hat{\lambda}_g - \hat{\gamma}_t$
  $\hat{y}_{git}$ =a +b $D_{gt} + e_{git}$
- Idea: the first stage cleanses out the FE using only the non-treatment group.
- Then after cleansing that out, you estimate the model.

An alternative: DRDiD

- An alternative is the doubly-robust DiD
- Idea: DiD with two-way FE yields a valid ATE if there is no heterogeneity in treatment effect across groups.
- Alternatively what if we used propensity score to adjust for covariate differences, and then simply take mean differences?
- Each if valid *if* its assumptions are correct.
- But if we combine both, we can say the combination if valid if *either* assumption is true – doubly robust.

The propensity score DiD

- Hopefully looks familiar-ish!

$$\widehat{\delta}^{ipw} = \frac{1}{E_N[D]} E\left[\frac{D - \widehat{p}(X)}{1 - \widehat{p}(X)}(Y_1 - Y_0)\right]$$

- Then combine this with a regression to control for covariate differences – can’t be written easily as a regression.

Intuition

- Take the pre-post difference for treated, and the adjust to take out the time trend

$$\widehat{\delta}^{OR} = \overline{Y}_{1,1} - \left[\overline{Y}_{1,0} + \frac{1}{n^T} \sum_{i|D_i=1} \left(\widehat{\mu}_{0,1}(X_i) - \widehat{\mu}_{0,0}(X_i) \right) \right]$$

- But rather than just T for control group pre vs post, you adjust for covariate differences.
- Now imagine doing this and IPW at the same time.
- Fortunately, exists as Stata package

Extension 11

- Fuzzy DiD
- Similar idea to fuzzy RD
- Rather than sharp treated-untreated groups, the treatment is higher in the treatment group compared to the comparison group
- Solution?

$$Wald_{DiD} = \frac{\left(E[Y_k|Post] - E[Y_k|Pre]\right) - \left(E[Y_U|Post] - E[Y_U|Pre]\right)}{\left(E[D_k|Post] - E[D_k|Pre]\right) - \left(E[D_U|Post] - E[D_U|Pre]\right)}$$

Extension 11

- This turns out to work with stronger assumptions.
- de Chaisemartin and d'Haultfœuille proposed an estimator that works with weaker assumptions
- If you're interested there is a `fuzzydid` Stata program