Lecture 4:
Difference in Difference, 2 of 2

September 27, 2023

Course Administration

1. Graded summaries through late last week
2. Lab after class this week
3. PS 2 due next week
4. Any problem set 2 issues?
5. If you haven't identified a replication paper, I'm nervous
6. Any other issues?

Today

Relaxing diff-in-diff: event study

1. Simplest possible event study
2. Diff-in-diff event study
3. Estimating trends
4. Testing for trends
5. Important things we don't cover

Today

Relaxing diff-in-diff: event study

1. Simplest possible event study
2. Diff-in-diff event study
3. Estimating trends
4. Testing for trends
5. Important things we don't cover

Janssen and Zhang

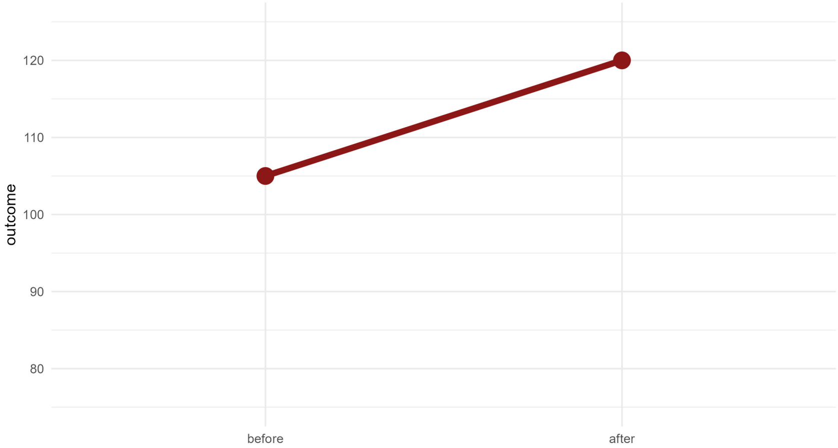
1. Diff-in-diff specification
2. Event study specification

1. Simplest Event Study

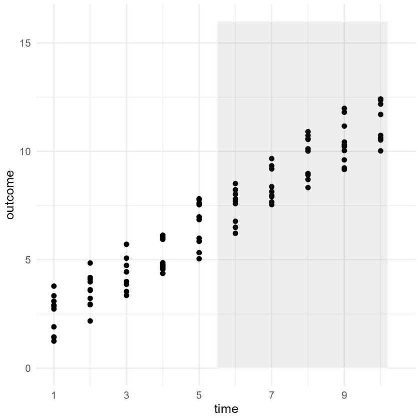
Basic Set-Up

- We want to know the impact of X on Y
- Over time, the treatment X changes – increases, decreases, appears, disappears
- Compare outcomes Y before and after change in X
- Examples, please!

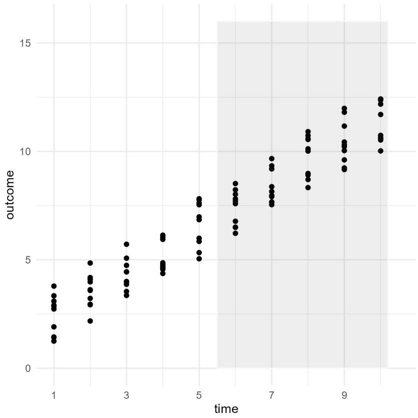
Last Week: Only Before and After



More Dots: Observe Each i in each time t



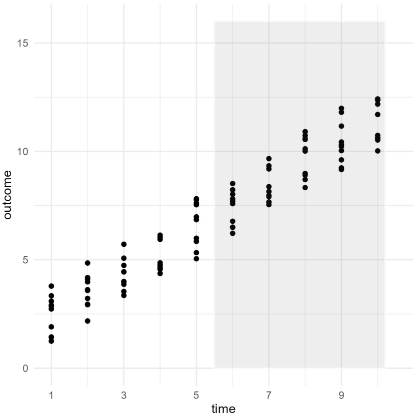
More Dots: Observe Each i in each time t



- All i are treated
- At all times $t > T_0$

Equation to estimate average Y after?

More Dots: Observe Each i in each time t



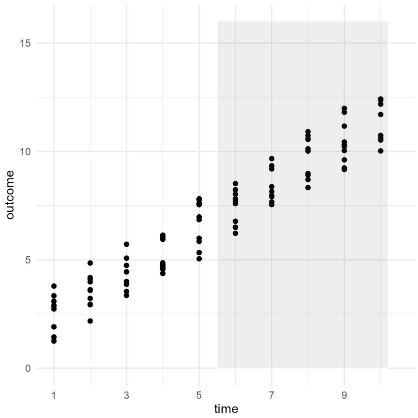
- All i are treated
- At all times $t > T_0$

Equation to estimate average Y after?

$$Y_{i,t} = \beta_0 + \beta_1 after_t + \epsilon_{i,t}$$

where after$_t$ is 1 for years $t > T_0$.

More Dots: Observe Each i in each time t



- All i are treated
- At all times $t > T_0$

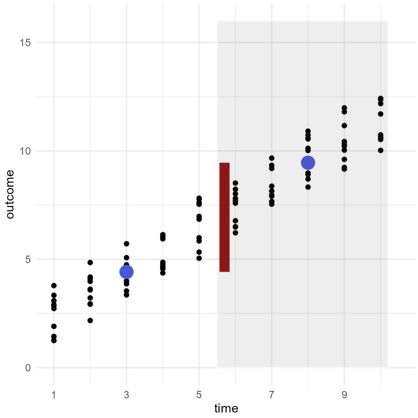
Equation to estimate average Y after?

$$Y_{i,t} = \beta_0 + \beta_1 after_t + \epsilon_{i,t}$$

where $after_t$ is 1 for years $t > T_0$.

What does β_1 report?

What β_1 Reports

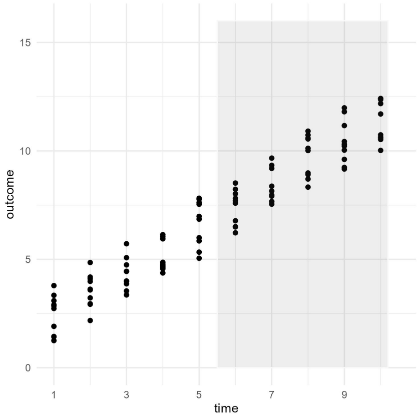


- All i are treated
- At all times $t > T_0$

$$Y_{i,t} = \beta_0 + \beta_1 after_t + \epsilon_{i,t}$$

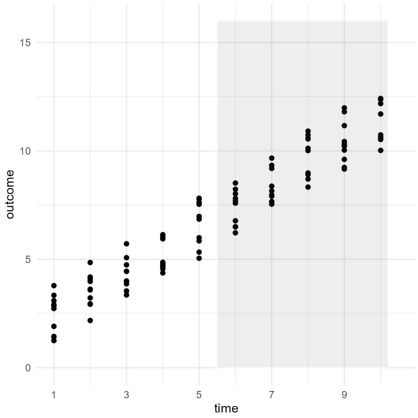
where after$_t$ is 1 for years $t > T_0$

Estimating the Impact of Time Granularly



How do we estimate the impact of treatment in each period individually?

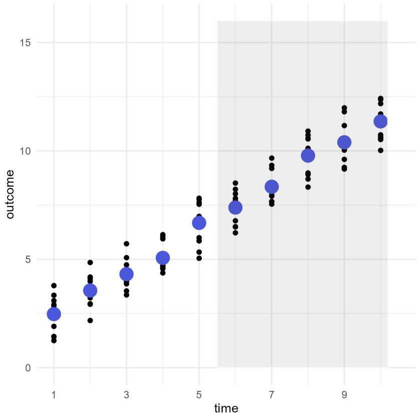
Estimating the Impact of Time Granularly



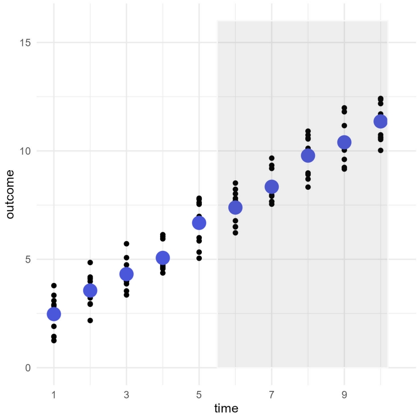
How do we estimate the impact of treatment in each period individually?

$$Y_{i,t} = \beta_0 + \beta_{1,t} 1\{\text{time} = t\}_t + \epsilon_{i,t}$$

Raw Data: Event Study Diagram



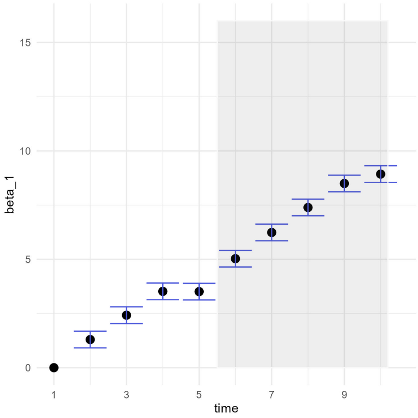
Raw Data: Event Study Diagram



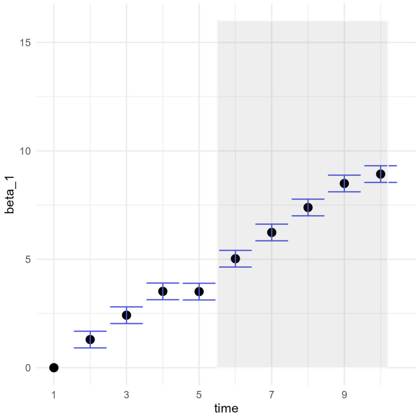
$$Y_{i,t} = \beta_0 + \beta_{1,t} 1\{\text{time} = t\}_t + \epsilon_{i,t}$$

- Regression coefficients should measure these means in the raw data
- What do you think a plot of $\beta_{1,t}$ should look like?

Regression Coefficients: Event Study Diagram

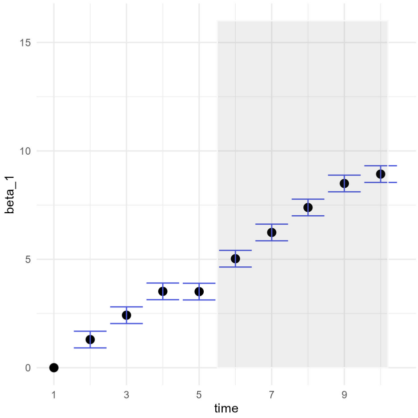


Regression Coefficients: Event Study Diagram



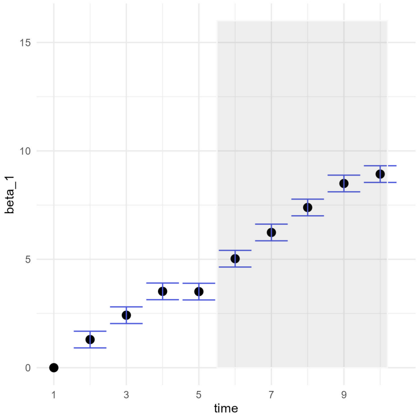
- Everything is relative to mean in year 1

Regression Coefficients: Event Study Diagram



- Everything is relative to mean in year 1
- Why might comparing pre- and post blue dots not give the causal impact of X on Y?

Regression Coefficients: Event Study Diagram



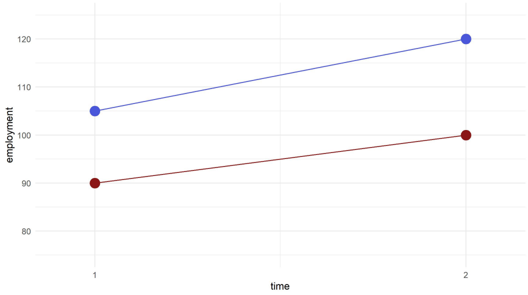
- Everything is relative to mean in year 1
- Why might comparing pre- and post blue dots not give the causal impact of X on Y?
- Given what we learned last class, how can we fix?

2. Diff-in-diff Event Study

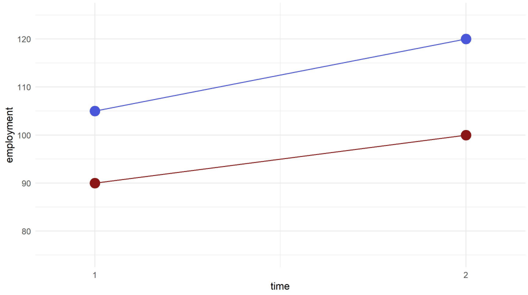
Basic Set-Up

- We want to know the impact of X on Y
- Over time, the treatment X changes – increases, decreases, appears, disappears
- **Some units experience a change in X – are treated – and others are not**
- Compare outcomes Y before and after change in X
- Examples, please!

Review: How We Do This with Just Before and After



Review: How We Do This with Just Before and After

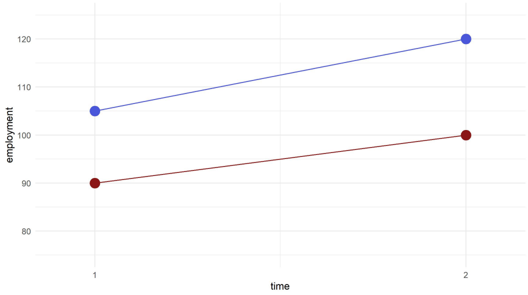


Equation to estimate impact of treatment?

- For treated i assign $\text{treated}_i = 1$
- Treatment at all times $t > T_0$

Equation to estimate diff-in-diff?

Review: How We Do This with Just Before and After



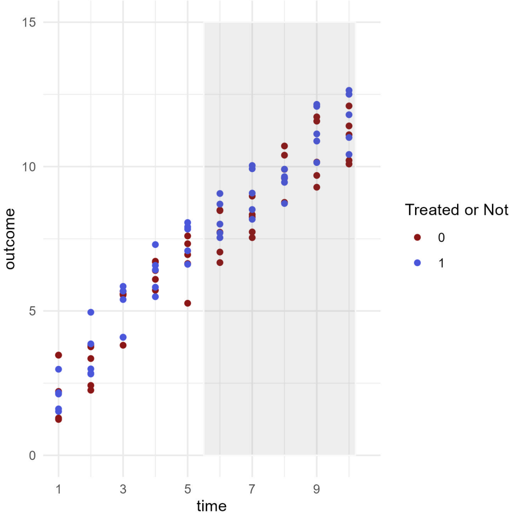
Equation to estimate impact of treatment?

- For treated i assign $\text{treated}_i = 1$
- Treatment at all times $t > T_0$

Equation to estimate diff-in-diff?

$$\begin{aligned} Y_{i,t} &= \beta_0 + \beta_1 \text{treated}_i * \text{after}_t \\ &+ \beta_2 \text{treated}_i + \beta_3 \text{after}_t + \epsilon_{i,t} \end{aligned}$$

Treated and Untreated in an Event Study Framework



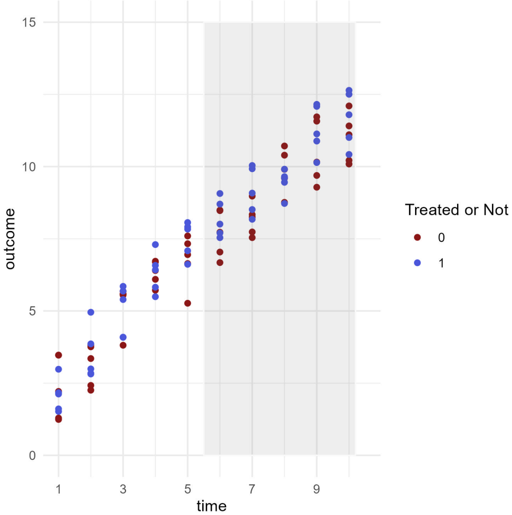
- For treated i assign $\text{treated}_i = 1$
- Treatment at all times $t > T_0$

If we estimate treatment impact via diff-in-diff equation

$$
\begin{aligned}
Y_{i,t} &= \beta_0 + \beta_1 \text{treated}_i * \text{after}_t \\
&+ \beta_2 \text{treated}_i + \beta_3 \text{after}_t + \epsilon_{i,t}
\end{aligned}
$$

what does it compare?

Treated and Untreated in an Event Study Framework



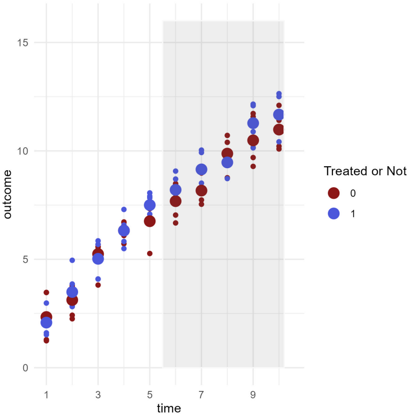
- For treated i assign $\text{treated}_i = 1$
- Treatment at all times $t > T_0$

If we estimate treatment impact via diff-in-diff equation

$$\begin{aligned} Y_{i,t} &= \beta_0 + \beta_1 \text{treated}_i * \text{after}_t \\ &+ \beta_2 \text{treated}_i + \beta_3 \text{after}_t + \epsilon_{i,t} \end{aligned}$$

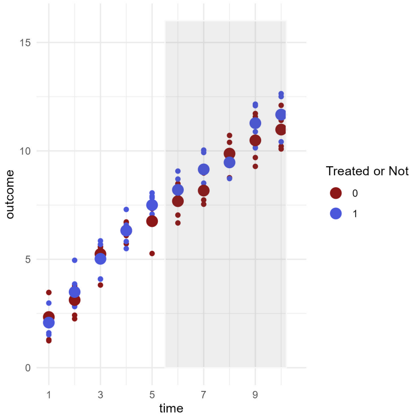
what does it compare? Comparison is **still** all before vs all after, but relative to untreated

Estimating the Impact of Time Granularly: Event Study



Can we estimate the impact of each period individually?

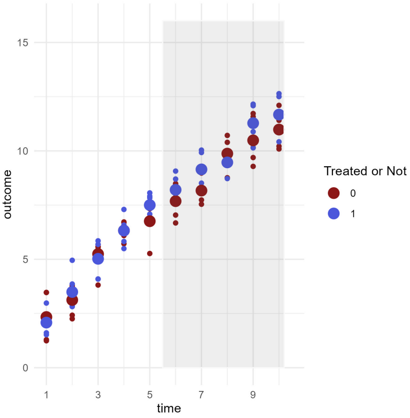
Estimating the Impact of Time Granularly: Event Study



Can we estimate the impact of each period individually?

$$\begin{aligned} Y_{i,t} &= \beta_0 + \beta_{1,t}\text{treated}_i * 1\{\text{time} = t\}_t \\ &+ \beta_2\text{treated}_i + \beta_{3,t}1\{\text{time} = t\}_t + \epsilon_{i,t} \end{aligned}$$

Estimating the Impact of Time Granularly: Event Study

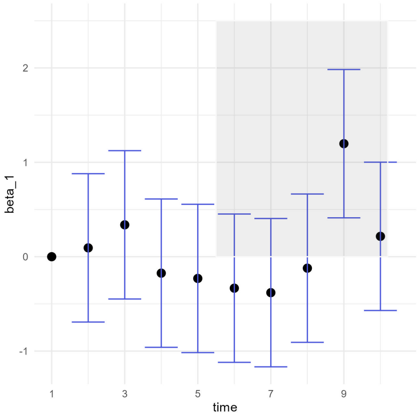


Can we estimate the impact of each period individually?

$$
\begin{aligned}
Y_{i,t} &= \beta_0 + \beta_{1,t}\text{treated}_i * 1\{\text{time} = t\}_t \\
&+ \beta_2\text{treated}_i + \beta_{3,t}1\{\text{time} = t\}_t + \epsilon_{i,t}
\end{aligned}
$$

What do you expect $\beta_{1,t}$ to be given this figure?

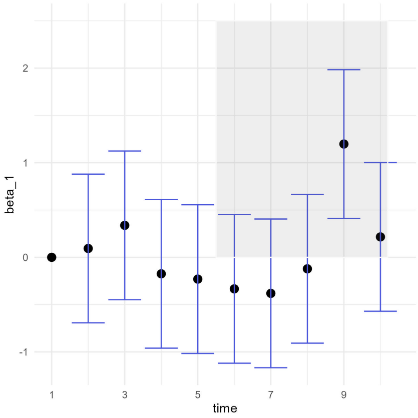
Estimating the Impact of Time Granularly: Regression Coefficients



Plot $\beta_{1,t}$:

$$Y_{i,t} = \beta_0 + \beta_{1,t}\text{treated}_i * 1\{\text{time} = t\}_t + \beta_2\text{treated}_i + \beta_{3,t}1\{\text{time} = t\}_t + \epsilon_{i,t}$$

Estimating the Impact of Time Granularly: Regression Coefficients



Plot $\beta_{1,t}$:

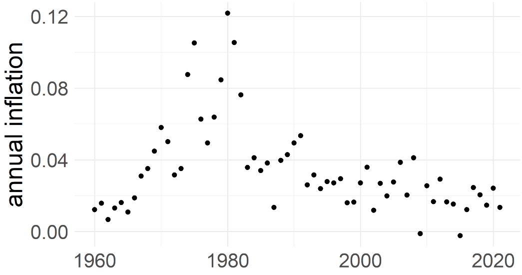
$$Y_{i,t} = \beta_0 + \beta_{1,t}\text{treated}_i * 1\{\text{time} = t\}_t + \beta_2\text{treated}_i + \beta_{3,t}1\{\text{time} = t\}_t + \epsilon_{i,t}$$

But ...

- you may care about the change in trends
- you may want to estimate the effect net of trends

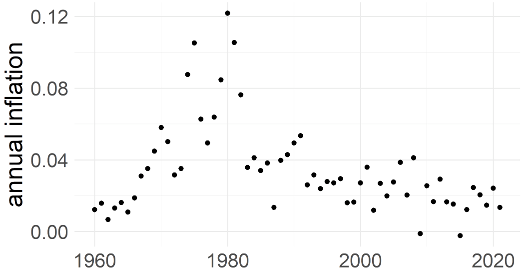
3. Estimating Trends

On Trends



How do we calculate a linear trend for these data?

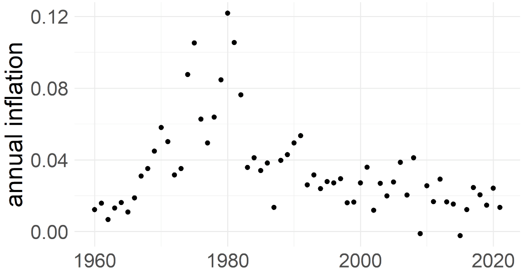
On Trends



How do we calculate a linear trend for these data?

$$\text{inflation}_t = \alpha_0 + \alpha_1 \text{year}_t + \epsilon_t$$

On Trends



How do we calculate a linear trend for these data?

$$\text{inflation}_t = \alpha_0 + \alpha_1 \text{year}_t + \epsilon_t$$

Graph $\alpha_0 + \alpha_1 * \text{year}_t$ where year_t is $\{1, 2, 3, ...\}$

Just To Be Clear on Data

year	inflation	year2
1980	0.12	1
1981	0.10	2
1982	0.07	3
1983	0.03	4

$$\text{inflation}_t = \alpha_0 + \alpha_1 \text{year}_t + \epsilon_t$$

and

$$\text{inflation}_t = \gamma_0 + \gamma_1 \text{year2}_t + \epsilon_t$$

yield $\alpha_1 = \gamma_1$

Just To Be Clear on Data

year	inflation	year2
1980	0.12	1
1981	0.10	2
1982	0.07	3
1983	0.03	4

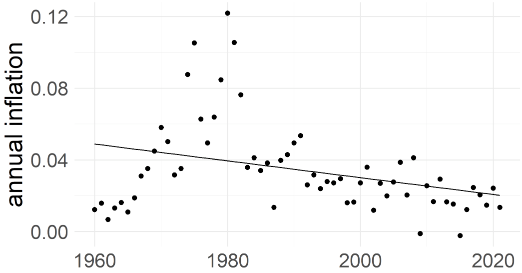
$$\text{inflation}_t = \alpha_0 + \alpha_1 \text{year}_t + \epsilon_t$$

and

$$\text{inflation}_t = \gamma_0 + \gamma_1 \text{year2}_t + \epsilon_t$$

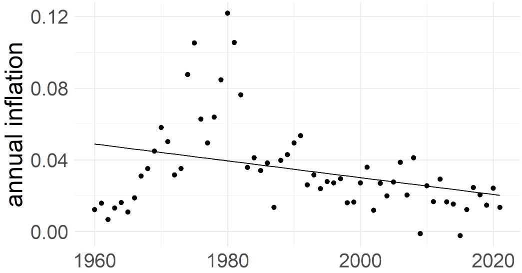
yield $\alpha_1 = \gamma_1$, but not $\alpha_0 = \gamma_0$

Adding Trends



What's odd about this line?

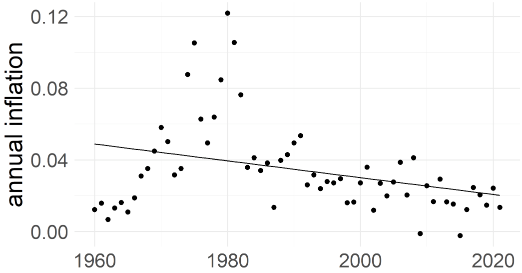
Adding Trends



What's odd about this line?

Make two lines

Adding Trends



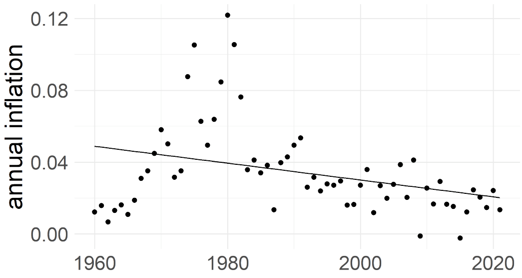
What's odd about this line?

Make two lines

$$\text{inflation}_t = \delta_0 + \delta_1 A_t + \delta_2 \text{year}_t + \delta_3 A_t * \text{year}_t + \epsilon_t$$

where A_t is 1 if $\text{year}_t > T_0$ and 0 otherwise.

Adding Trends



What's odd about this line?

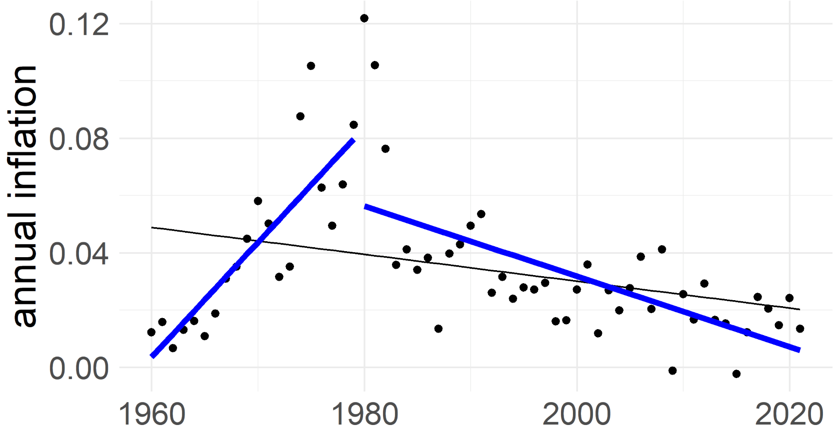
Make two lines

$$\text{inflation}_t = \delta_0 + \delta_1 A_t + \delta_2 \text{year}_t + \delta_3 A_t * \text{year}_t + \epsilon_t$$

where A_t is 1 if $\text{year}_t > T_0$ and 0 otherwise.

What might you want T_0 to be?

Separate Trends



Adding Linear Trends

What is a linear trend?

- a variable that increases linearly for each unit of time – here a year
- the calendar year is a trend variable
- this is different than a fixed effect

Adding Linear Trends

What is a linear trend?

- a variable that increases linearly for each unit of time – here a year
- the calendar year is a trend variable
- this is different than a fixed effect

ID	year	t1	t2
A	1990	1	5
A	1991	2	10
A	1992	3	15
B	2000	11	55
B	2001	12	60
B	2002	13	65

4. Validity Tests

Validity Tests

- Parallel trends in the absence of treatment is unobservable
- But you can assess parallel trends pre-treatment
- This is precisely estimable

Adding a Pre-Treatment Trend

Suppose you start with

$$Y_{i,t} = \beta_0 + \beta_1 \text{treated}_i * \text{after}_t + \beta_2 \text{treated}_i + \beta_3 \text{after}_t + \epsilon_{i,t}$$

and you want to test for pre-treatment trends. What do you do?

Adding a Pre-Treatment Trend

Suppose you start with

$$Y_{i,t} = \beta_0 + \beta_1 \text{treated}_i * \text{after}_t + \beta_2 \text{treated}_i + \beta_3 \text{after}_t + \epsilon_{i,t}$$

and you want to test for pre-treatment trends. What do you do?

- Use only data from before treatment
- Estimate

$$Y_{i,t} = \alpha_0 + \alpha_1 \text{year}_t + \alpha_2 \text{treated}_i + \alpha_3 \text{treated}_i * \text{year}_t + \epsilon_{i,t}$$

Adding a Pre-Treatment Trend

Suppose you start with

$$Y_{i,t} = \beta_0 + \beta_1 \text{treated}_i * \text{after}_t + \beta_2 \text{treated}_i + \beta_3 \text{after}_t + \epsilon_{i,t}$$

and you want to test for pre-treatment trends. What do you do?

- Use only data from before treatment
- Estimate

$$Y_{i,t} = \alpha_0 + \alpha_1 \text{year}_t + \alpha_2 \text{treated}_i + \alpha_3 \text{treated}_i * \text{year}_t + \epsilon_{i,t}$$

- What do we expect if there is no pre-treatment trend?

Adding a Pre-Treatment Trend

Suppose you start with

$$Y_{i,t} = \beta_0 + \beta_1 \text{treated}_i * \text{after}_t + \beta_2 \text{treated}_i + \beta_3 \text{after}_t + \epsilon_{i,t}$$

and you want to test for pre-treatment trends. What do you do?

- Use only data from before treatment
- Estimate

$$Y_{i,t} = \alpha_0 + \alpha_1 \text{year}_t + \alpha_2 \text{treated}_i + \alpha_3 \text{treated}_i * \text{year}_t + \epsilon_{i,t}$$

- What do we expect if there is no pre-treatment trend? $\alpha_3 = 0$

Additional Validity Tests

- Add unit-specific time trends. If these kill the effect, what does this tell us?
  - for example, you have state by year data
  - looking for the impact of a policy that hits some states and not others

Additional Validity Tests

- Add unit-specific time trends. If these kill the effect, what does this tell us?
    - for example, you have state by year data
    - looking for the impact of a policy that hits some states and not others
- Triple difference – not always possible

5. Important Things We Don't Cover

Time Is Limited, So We Skip Important Things

A non-exhaustive list includes

1. How serial correlation can inflate estimates. See Bertrand, et al., 2004
2. Heterogeneous treatment effects + differential treatment timing can bias estimates – large current literature

Opioids and Event Studies

Order of Events

1. Paper background
2. Diff-in-diff strategies
   2.1 independent vs chain, geographic fixed effects
   2.2 exploit independents that change to chain
   2.3 independent vs chain, before and after reformulation

Paper Basics

- What are the two key pharmacy types?

Paper Basics

- What are the two key pharmacy types?
- What is the causal research question?

Paper Basics

- What are the two key pharmacy types?
- What is the causal research question?
- What are the potential challenges to identification? or, why don't we just compare outcomes at independents and chains?

Paper Basics

- What are the two key pharmacy types?
- What is the causal research question?
- What are the potential challenges to identification? or, why don't we just compare outcomes at independents and chains?

Data

Paper Basics

- What are the two key pharmacy types?
- What is the causal research question?
- What are the potential challenges to identification? or, why don't we just compare outcomes at independents and chains?

Data

- morphine equivalent doses (MEDs)
- by pharmacy
- by month

Paper Basics

- What are the two key pharmacy types?
- What is the causal research question?
- What are the potential challenges to identification? or, why don't we just compare outcomes at independents and chains?

Data

- morphine equivalent doses (MEDs)
- by pharmacy
- by month
- What is the unit of observation?

E1: Independents vs Everyone Else

$$Y_{it} = \beta \text{Indep}_i + \mu_t + \gamma_{FE} + \epsilon_{it}$$

- $Y_{i,t}$ MED at pharmacy i at time t
- Indep_i: 1 if independent
- μ_t: year-month FE
- γ_{FE}: place FE

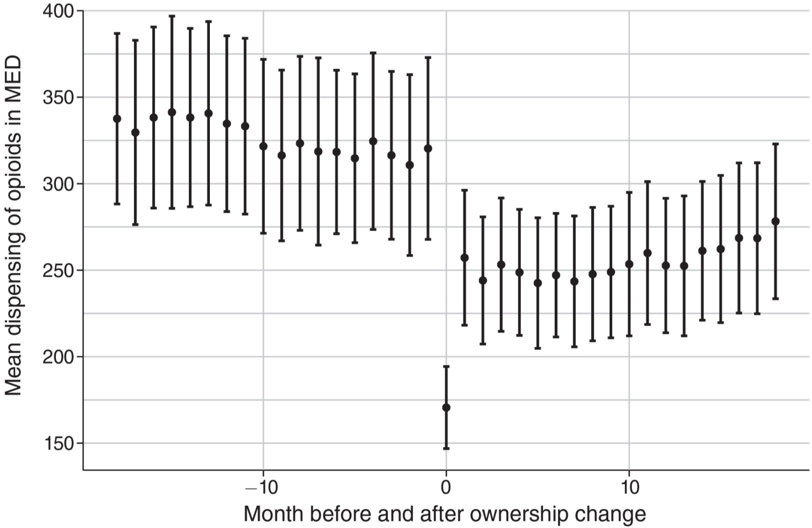
E1: Independents vs Everyone Else

$$Y_{it} = \beta \text{Indep}_i + \mu_t + \gamma_{FE} + \epsilon_{it}$$

- $Y_{i,t}$ MED at pharmacy i at time t
- Indep_i: 1 if independent
- μ_t: year-month FE
- γ_{FE}: place FE

	(1)	(2)	(3)	(4)
Independent	50.131	51.362	107.826	128.016
	(4.908)	(4.912)	(5.551)	(5.875)
Constant	306.488			
	(2.109)			
Year-month fixed effects	No	Yes	Yes	Yes
County fixed effects	No	No	Yes	No
Zip code fixed effects	No	No	No	Yes
Mean outcome	327.19	327.19	327.19	327.19
Mean effect in percent	15.32	15.7	32.96	39.13
Observations	5,055,761	5,055,761	5,055,761	5,055,761
R^2	0.002	0.010	0.089	0.225

E2: Change in Ownership, Raw Data



E2: Change in Ownership, Regression Form

Estimate either

$$Y_{i,t} = \beta_0 D_{it}^{\text{PRE}} + \beta_1 D_{it}^{\text{POST}} + \beta_C \text{CHAIN}_i + \mu_t + \epsilon_{i,t}$$

or

$$Y_{i,t} = \beta_1 D_{it}^{\text{POST}} + \alpha_i + \mu_t + \epsilon_{i,t}$$

- D_{it}^{PRE}: 1 for indep's that change to chain, before change
- D_{it}^{POST}: 1 for indep's that change to chain, after change
- CHAIN_i: 1 for always chains
- α_i: pharmacy FE

E2: Change in Ownership, Regression Form

Estimate either

$$Y_{i,t} = \beta_0 D_{it}^{\text{PRE}} + \beta_1 D_{it}^{\text{POST}} + \beta_C \text{CHAIN}_i + \mu_t + \epsilon_{i,t}$$

or

$$Y_{i,t} = \beta_1 D_{it}^{\text{POST}} + \alpha_i + \mu_t + \epsilon_{i,t}$$

- D_{it}^{PRE}: 1 for indep's that change to chain, before change
- D_{it}^{POST}: 1 for indep's that change to chain, after change
- CHAIN_i: 1 for always chains
- α_i: pharmacy FE

- how do we interpret β_0?

E2: Change in Ownership, Regression Form

Estimate either

$$Y_{i,t} = \beta_0 D_{it}^{\text{PRE}} + \beta_1 D_{it}^{\text{POST}} + \beta_C \text{CHAIN}_i + \mu_t + \epsilon_{i,t}$$

or

$$Y_{i,t} = \beta_1 D_{it}^{\text{POST}} + \alpha_i + \mu_t + \epsilon_{i,t}$$

- D_{it}^{PRE}: 1 for indep's that change to chain, before change
- D_{it}^{POST}: 1 for indep's that change to chain, after change
- CHAIN_i: 1 for always chains
- α_i: pharmacy FE

- how do we interpret β_0?
- and β_1?

E2: Change in Ownership, Regression Form

Estimate either

$$Y_{i,t} = \beta_0 D_{it}^{\text{PRE}} + \beta_1 D_{it}^{\text{POST}} + \beta_C \text{CHAIN}_i + \mu_t + \epsilon_{i,t}$$

or

$$Y_{i,t} = \beta_1 D_{it}^{\text{POST}} + \alpha_i + \mu_t + \epsilon_{i,t}$$

- D_{it}^{PRE}: 1 for indep's that change to chain, before change
- D_{it}^{POST}: 1 for indep's that change to chain, after change
- CHAIN_i: 1 for always chains
- α_i: pharmacy FE

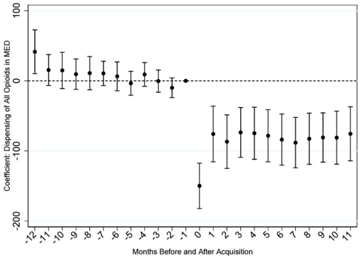
- how do we interpret β_0?
- and β_1?
- why not both equations together?

E2: Change in Ownership, Results

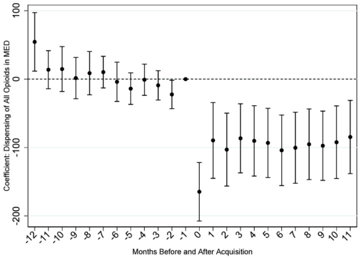
	All			
	OLS	OLS	OLS	OLS
	(1)	(2)	(3)	(4)
D^{PRE}	1.516	32.777	−1.226	
	(33.915)	(33.655)	(32.747)	
D^{POST}	−102.89	−130.867	−153.215	−110.507
	(19.755)	(19.61)	(20.439)	(16.65)
CHAIN	−49.933	−50.89	−127.879	
	(4.931)	(4.934)	(5.912)	
Constant	356.624			
	(4.883)			
Year-month fixed effects	No	Yes	Yes	Yes
Zip code fixed effects	No	No	Yes	No
Facility fixed effects	No	No	No	Yes
Mean outcome	327.19	327.19	327.19	327.19
Mean effect in percent	−31.45	−40	−46.83	−33.77
Observations	5,055,761	5,055,761	5,055,761	5,055,761
R^2	0.002	0.01	0.225	0.809

E2: Change in Ownership, Event Study Estimates

From Online Appendix, Figure E.1



(a) Dispensing of all opioids in MED, facility and year-month fixed effects



(b) Dispensing of all opioids in MED, facility and ZIP code × year-month fixed effects

E3: Reformulation

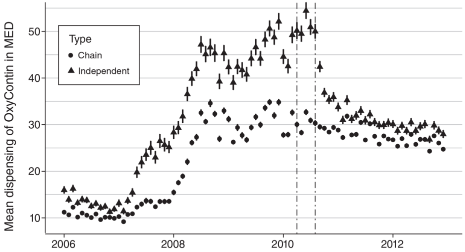


Figure 2

E3: Reformulation

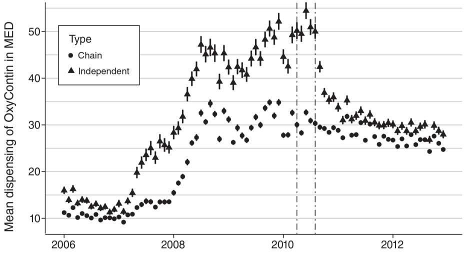


Figure 2

- why should reformulation matter?
- what should we be comparing in this figure to see the double diff?
- what should we be comparing to look for validity?

E3: Specification

What regression should we use to test impact of reformulation at independent pharmacies vs chains?

E3: Specification

What regression should we use to test impact of reformulation at independent pharmacies vs chains?

$$Y_{it} = \beta \text{Indep}_i * \text{Post}_t + \alpha_i + \mu_t + \epsilon_{it}$$

E3: Specification

What regression should we use to test impact of reformulation at independent pharmacies vs chains?

$$Y_{it} = \beta \text{Indep}_i * \text{Post}_t + \alpha_i + \mu_t + \epsilon_{it}$$

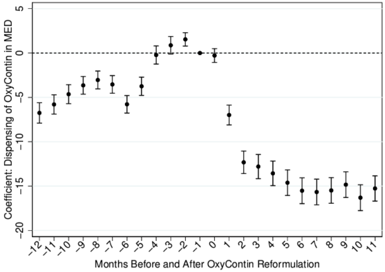
- Why no γ_{FE}?
- How do we interpret β?

E3: Reformulation, Results

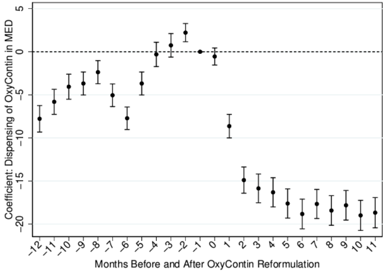
	Full sample: 2006–2012			
	(1)	(2)	(3)	(4)
Independent × Post	−6.097	−6.436	−6.996	−5.339
	(0.529)	(0.529)	(0.565)	(0.484)
Independent	10.569	10.912	18.886	
	(0.681)	(0.683)	(0.832)	
Post	6.095			
	(0.154)			
Constant	21.495			
	(0.281)			
Year-month fixed effects	No	Yes	Yes	Yes
Zip code fixed effects	No	No	Yes	No
Pharmacy fixed effects	No	No	No	Yes
Mean outcome	27.14	27.14	27.14	27.14
Mean effect in percent	−22.47	−23.72	−25.78	−19.67
Observations	5,055,761	5,055,761	5,055,761	5,054,885
R^2	0.004	0.019	0.159	0.650

E3: Reformulation Event Study Results

Online Appendix Figure E.5



(a) Dispensing of OxyContin in MED, pharmacy and year-month fixed effects



(b) Dispensing of OxyContin in MED, pharmacy and ZIP code × year-month fixed effects

Next Lecture

- Read
  - *Mastering Metrics* Chapter 3
  - an oldie but goodie: Angrist and Kreuger, 1991
  - skim 2c
- Turn in PS 2
- Summary due next week if you're on the list