FNCE 926
Empirical Methods in CF

Lecture 7 – Natural Experiment *[P2]*

Professor Todd Gormley

Announcements & Informal Survey

- Exercise #3 is due
- Please fill out informal survey
  - Helps me figure out what changes I can make to improve the course for the second half and for future years
    - For example, what topic should I have spent more time on? What topic did you find the most interesting? Is there too much, or too little work? Etc.

Background readings

- Roberts and Whited
  - *Sections 2.2, 4*
- Angrist and Pischke
  - *Section 5.2*

Outline for Today

- Quick review of last lecture
- Continue to discuss natural experiments
  - How to handle multiple events
  - Triple differences
  - Common robustness tests that can be used to test whether internal validity is likely to hold
- Student presentations of "NE #1" papers

Quick Review *[Part 1]*

- Natural experiment provides random variation in x that allows causal inference
  - Can be used in IV, regression discontinuity, but most often associated with "treatment" effects
- Two types of simple differences
  - Post-treatment comparison of treated & untreated
  - Pre- and post-treatment comparison of treated

Quick Review *[Part 2]*

- Difference-in-differences is estimated with…

$$y_{i,t} = \beta_0 + \beta_1 p_t + \beta_2 d_i + \beta_3 \left(d_i \times p_t\right) + u_{i,t}$$

  - Compares <u>change</u> in *y* pre- versus post-treatment for treated to <u>change</u> in *y* for untreated
  - Requires "parallel trends" assumption

- Let's test your ability to identify a violation of the necessary assumptions for simple diffs and diff-in-diffs…

Quick Review *[Part 3]*

- Suppose Spain exits the Euro. And, Ann compares profitability of Spanish firms after the exit to profitability before…
- **What is necessary for the comparison to have any causal interpretation?**
  - **Answer =** We must assume profitability after Spain's exit would have been same as profitability prior to exit absent exit… Highly implausible

Quick Review *[Part 4]*

- Now, suppose Bob compares profitability of Spanish firms after the exit to profitability of German firms after exit…
- **What is necessary for the comparison to have any causal interpretation?**
  - **Answer =** We must assume profitability of Spanish firm would have been same as profitability of German firms absent exit… Again, this is highly implausible

Quick Review *[Part 5]*

- Lastly, suppose Charlie compares <u>change</u> in profitability of Spanish firms after exit to <u>change</u> in profitability of German firms
- **What is necessary for the comparison to have any causal interpretation?**
  - **Answer =** We must assume change in profitability of Spanish firm would have been same as change for German firms absent exit… I.e. parallel trends assumption

Natural Experiment [P2] – *Outline*

- Difference-in-difference continued…
  - Using group means to get an estimate
  - When additional controls are appropriate
- How to handle multiple events
- Falsification tests
- Triple differences

Standard Regression Format

- Difference-in-differences estimator

$$y_{i,t} = \beta_0 + \beta_1 p_t + \beta_2 d_i + \beta_3 \left(d_i \times p_t \right) + u_{i,t}$$

  - $p_t = 1$ if period t occurs after treatment and equals zero otherwise
  - $d_i = 1$ if unit is in treated group and equals zero otherwise

But, there is another way that just involves comparing four sample means…

Comparing group means approach

- To see how we can get the same estimate, β_3, by just comparing sample means, first calculate expected y under four possible combinations of p and d indicators

Comparing group means approach *[P1]*

- Again, the regression is…

$$y_{i,t} = \beta_0 + \beta_1 p_t + \beta_2 d_i + \beta_3 \left(d_i \times p_t \right) + u_{i,t}$$

  - And, the four possible combinations are:

$$E(y \mid d = 1, p = 1) = \beta_0 + \beta_1 + \beta_2 + \beta_3$$
$$E(y \mid d = 1, p = 0) = \beta_0 + \beta_2$$
$$E(y \mid d = 0, p = 1) = \beta_0 + \beta_1$$
$$E(y \mid d = 0, p = 0) = \beta_0$$

What assumption did I make in doing this?

Answer: $E(u \mid d,p)=0$; i.e. the "experiment" is random

Comparing group means approach *[P2]*

$$E(y \mid d=1, p=1) = \beta_0 + \beta_1 + \beta_2 + \beta_3$$
$$E(y \mid d=1, p=0) = \beta_0 + \beta_2$$
$$E(y \mid d=0, p=1) = \beta_0 + \beta_1$$
$$E(y \mid d=0, p=0) = \beta_0$$

- These can be arranged in two-by-two table

	Post-Treatment, (1)	**Pre-Treatment, (2)**
Treatment, (a)	$\beta_0+\beta_1+\beta_2+\beta_3$	$\beta_0+\beta_2$
Control, (b)	$\beta_0+\beta_1$	β_0

Comparing group means approach *[P3]*

- Now take the simple differences

	Post-Treatment, (1)	Pre-Treatment, (2)	Difference, (1)-(2)
Treatment, (a)	$\beta_0+\beta_1+\beta_2+\beta_3$	$\beta_0+\beta_2$	$\beta_1+\beta_3$
Control, (b)	$\beta_0+\beta_1$	β_0	β_1
Difference, (a)-(b)	$\beta_2+\beta_3$	β_2	

Comparing group means approach *[P4]*

- Then, take difference-in-differences!

	Post-Treatment, (1)	Pre-Treatment, (2)	Difference, (1)-(2)
Treatment, (a)	$\beta_0+\beta_1+\beta_2+\beta_3$	$\beta_0+\beta_2$	$\beta_1+\beta_3$
Control, (b)	$\beta_0+\beta_1$	β_0	β_1
Difference, (a)-(b)	$\beta_2+\beta_3$	β_2	$\boldsymbol{\beta_3}$

This is why they call it the difference-in-differences estimate; regression gives you same estimate as if you took differences in the group averages

Simple difference – Revisited *[Part 1]*

- Useful to look at simple differences

	Post-Treatment, (1)	Pre-Treatment, (2)	Difference, (1)-(2)
Treatment, (a)	$\beta_0+\beta_1+\beta_2+\beta_3$	$\beta_0+\beta_2$	$\beta_1+\beta_3$
Control, (b)	$\beta_0+\beta_1$	β_0	β_1
Difference, (a)-(b)	$\beta_2+\beta_3$	β_2	$\boldsymbol{\beta_3}$

This was cross-sectional simple difference

When does that simple diff give effect of treatment, β_3?

Answer = when β_2 equals zero; i.e. no difference in level of y absent treatment

Simple difference – Revisited *[Part 2]*

- Now, look at time-series simple diff…

	Post-Treatment, (1)	Pre-Treatment, (2)	Difference, (1)-(2)
Treatment, (a)	$\beta_0+\beta_1+\beta_2+\beta_3$	$\beta_0+\beta_2$	$\beta_1+\beta_3$
Control, (b)	$\beta_0+\beta_1$	β_0	β_1
Difference, (a)-(b)	$\beta_2+\beta_3$	β_2	$\boldsymbol{\beta_3}$

This was time-series simple difference

When does that simple diff give effect of treatment, β_3?

Answer = when β_1 equals zero; i.e. no change in *y* absent treatment

Why the regression is helpful

- Some papers will just report this simple two-by-two table as their estimate
- But, there are advantages to the regression
  - Can modify it to test timing of treatment *[we will talk about this in robustness section]*
  - Can add additional controls, X

Natural Experiment [P2] – *Outline*

- Difference-in-difference continued…
  - Using group means to get an estimate
  - When additional controls are appropriate
- How to handle multiple events
- Falsification tests
- Triple differences

Adding controls to diff-in-diff

- Easy to add controls to regression

$$y_{i,t} = \beta_0 + \beta_1 p_t + \beta_2 d_i + \beta_3 \left(d_i \times p_t\right) + \Gamma X_{i,t} + u_{i,t}$$

  - X is some vector of controls
  - Γ is vector of coefficients
- $E[y \mid d,p]$ in prior proofs just becomes $E[y \mid d,p,X]$

From earlier lecture, what type of controls should you NEVER add?

When controls are inappropriate

- Remember! You should never add controls that might themselves be affected by treatment
  - Angrist-Pischke call this a "bad control"
  - You won't be able to get a consistent estimate of β_3 from estimating the equation

A Pet Peeve of TG – *Refined*

- If you have a treatment that is truly random, do not put in controls affected by the treatment!
  - I've had many referees force me to add controls that are likely to be affected by the treatment…
  - If this happens to you, put in both regressions (with and without controls), and at a minimum, add a caveat as to why adding controls is inappropriate

When controls are appropriate

- Two main reasons to add controls
  - Improve precision (i.e. lower standard errors)
  - Restore 'random' assignment of treatment

#1 – To improve precision

- Adding controls can soak up some of residual variation (i.e. noise) allowing you to better isolate the treatment effect
  - Should the controls change the estimate?
    - NO! If treatment is truly random, adding controls shouldn't affect actual estimate; they should only help lower the standard errors!
  - If adding controls changes estimates, you might have 'bad controls' *or* worse, non-random treatment ☹

Example – Improving precision

- Suppose you have firm-level panel data
- Some natural experiment 'treats' some firms but not other firms
  - Could just estimate the standard diff-in-diff

$$y_{i,t} = \beta_0 + \beta_1 p_t + \beta_2 d_i + \beta_3 \left(d_i \times p_t \right) + u_{i,t}$$

  - *Or*, could add fixed effects (like firm and year FE) to get more precise estimate…

Example – Improving precision *[Part 2]*

- So, suppose you estimate…

$$y_{i,t} = \beta_0 + \beta_1 p_t + \beta_2 d_i + \beta_3 \left(d_i \times p_t \right) + \alpha_i + \delta_t + u_{i,t}$$

Firm fixed effects

Year fixed effects

  - **What meaning does $\boldsymbol{\beta}_1$ have now?**
  - **What meaning does $\boldsymbol{\beta}_2$ have now?**

Example – Improving precision *[Part 3]*

- Trick question! They have no meaning!
  - p_t is perfectly collinear with year FE *[because it doesn't vary across firms]*
  - d_i is perfectly collinear with firm FE *[because it doesn't vary across time for each firm]*
- Stata just randomly drops a couple of the FE
  - The estimates on p_t and d_i are just random intercepts with **no** meaning

Example – Improving precision *[Part 4]*

- Instead, you should estimate…

$$y_{i,t} = \beta_0 + \beta_3 \left(d_i \times p_t \right) + \alpha_i + \delta_t + u_{i,t}$$

Firm fixed effects control for treatment

Year fixed effects control for post-treatment

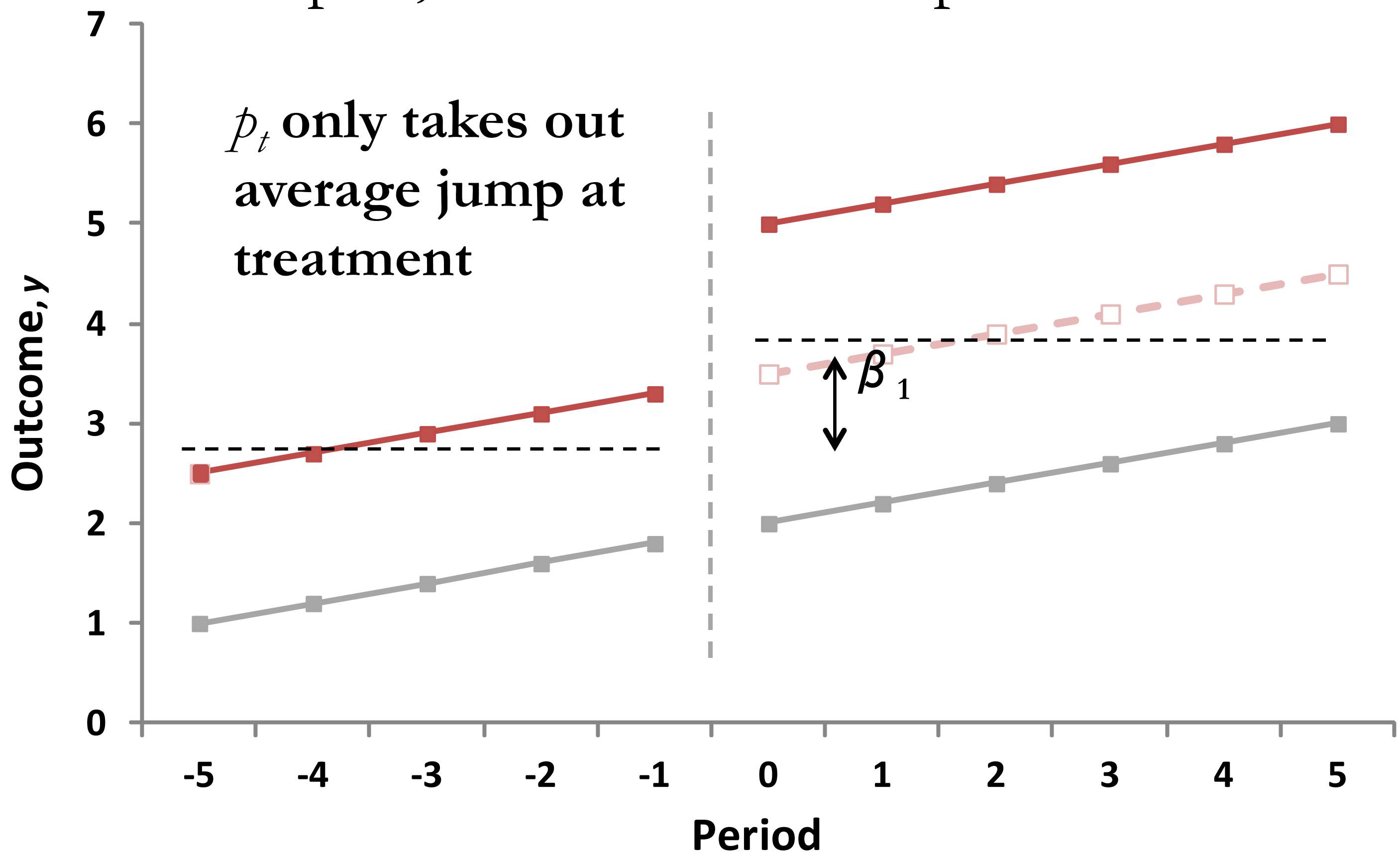
  - **This is what some call the generalized difference-in-differences estimator**

Generalized Difference-in-differences

- Advantage of generalized differences-in-differences is that it can improve precision and provide better fit of model
  - It doesn't assume all firms in treatment (or untreated) group have same average y; it allows intercept to vary for each firm
  - It doesn't assume that common change in y around event is a simple change in level; it allows common change in y to vary by year

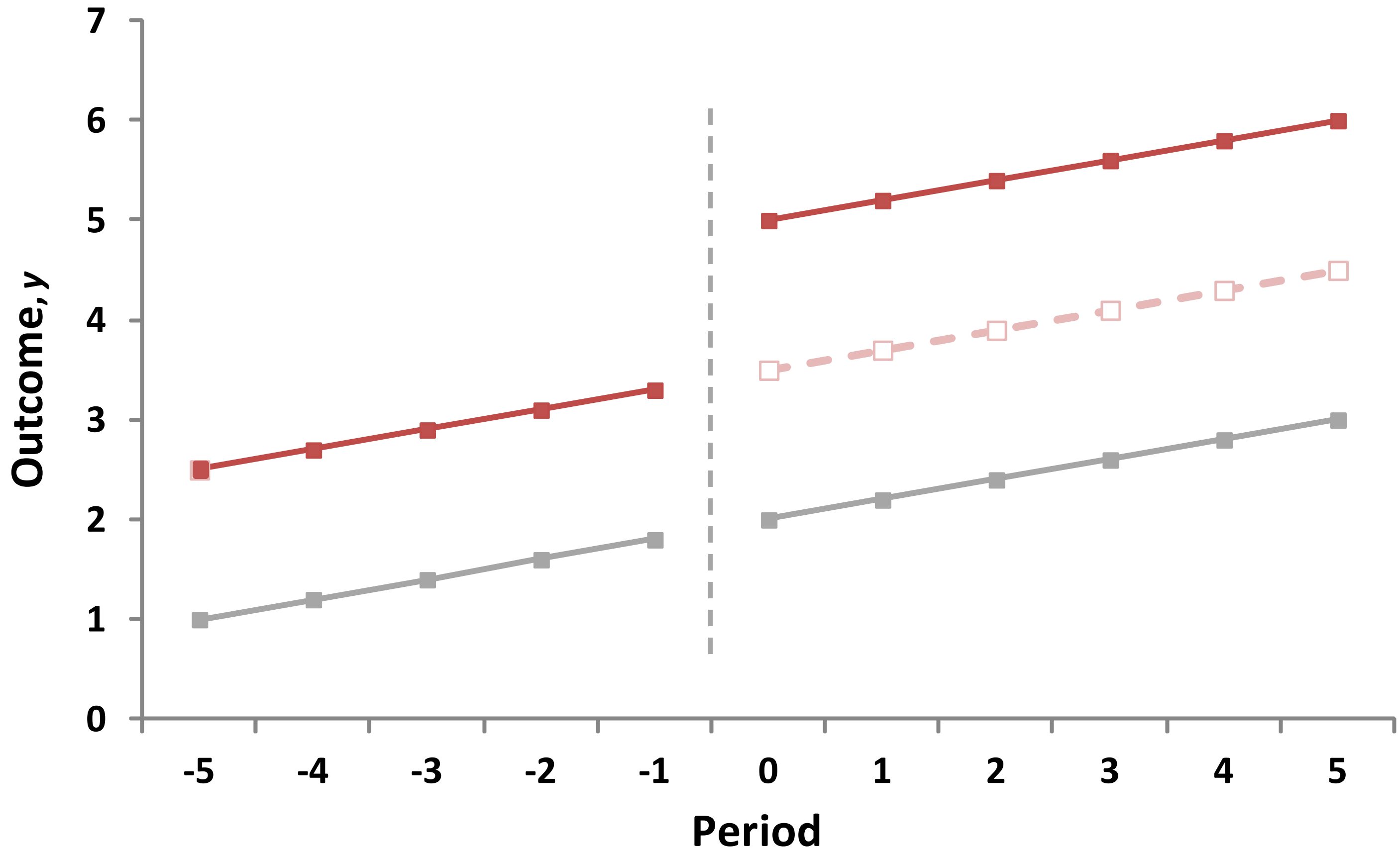
Generalized D-i-D – *Example [Part 1]*

- To see how Generalized D-i-D can be helpful, consider the example from last week



Generalized D-i-D – *Example [Part 2]*

- Year dummies will better fit actual trend



When controls are appropriate

- Two main reasons to add controls
  - Improve precision (i.e. lower standard errors)
  - Restore ‘random’ assignment of treatment

#2 – Restore randomness of treatment

- Suppose the following is true…
  - Observations of certain characteristic, e.g. high x, are more likely to be treated
  - **And,** firms with this characteristic are likely to have differential trend in outcome y
- Adding control for x could restore 'randomness'; i.e. being treated is random after controlling for x!

I.e. treatment isn't random

And, non-randomness is problematic for identification

Restoring randomness – *Example*

- Natural experiment is change in regulation
  - Firms affected by regulation is random, except that it is more likely to hit firms that are larger
  - *And*, we think larger firms might have different trend in outcome y afterwards for other reasons
  - <u>*And*,</u> firm size is not going to be affected by the change in regulation in any way
- If all true, adding size as control would be an appropriate and desirable thing to do

Controls continued…

- In prior example, suppose size is potentially effected by the change in regulation…
  - What would be another approach that won't run afoul of the 'bad control' problem?
    - **Answer:** Use firm size in year <u>prior</u> to treatment <u>and</u> it's interaction with post-treatment dummy
    - This will control for non-random assignment (based on size) and differential trend (based on size)

Restoring randomness – **Caution!**

- In practice, don't often see use of controls to restore randomness
  - Requires assumption that non-random assignment isn't also correlated with <u>unobservable</u> variables…
  - So, not that plausible unless there are very specific reasons for non-randomness
- **But, regression discontinuity is one example of this; we'll see it next week**

One last note… be careful about SEs

- Again, if have multiple pre- and post-treatment periods, need to be careful with standard errors
  - Either cluster SEs at level of each unit
  - Or, collapse data down to one pre- and one post-treatment observation for each cross-section
- We will discuss more about standard errors in lecture on "standard errors"

Natural Experiment [P2] – *Outline*

- Difference-in-difference continued…
- How to handle multiple events
  - Why they are useful
  - Two similar estimation approaches
- Falsification tests
- Triple differences

Motivating example…

- Gormley and Matsa (2011) looked at firms' responses to increased left-tail risk
  - Used discovery that workers were exposed to harmful chemical as exogenous increase in risk
  - One discovery occurred in 2000; a chemical heavily used by firms producing semiconductors was found to be harmful
- Can you think of any concerns about parallel trends assumption of this setting?

Motivating Example – *Answer*

- **Answer:** Yes… This coincides with bursting of technology bubble; technology firms might arguably trend differently after 2000 for this reasons unrelated to chemical
  - How might multiple treatment events, occurring at different times (which is what Gormley and Matsa used), help?

Multiple treatment events

- Sometimes, the natural experiment is repeated a multiple points in times for multiple groups of observations
  - E.g. U.S. states make a particular regulatory change at different points in time
- **These settings are particularly useful in mitigating concerns about violation of parallel trends assumption…**

How multiple events are helpful

- Can show that effect of treatment is similar across different time periods
- Can show effect of treatment isn't driven by a particular set of treated firms
  - I.e. now the "identification police" would need to come up with story as to why parallel trends is violated for <u>each</u> unique event

Natural Experiment [P2] – *Outline*

- Difference-in-difference continued…
- How to handle multiple events
  - Why they are useful
  - Two similar estimation approaches
- Falsification tests
- Triple differences

Estimation with Multiple Events

- Estimating model with multiple events is still relatively easy to do
  - Use approach of Bertrand and Mullainathan (JPE 2003)
  - Or, used "stacked" approach of Gormley and Matsa (RFS 2011)

Multiple Events – Approach #1 *[P1]*

- Just estimate the following estimation

$$y_{ict} = \beta d_{ict} + p_t + m_c + u_{ict}$$

  - y_{ict} is outcome for unit i (e.g. firm) in period t (e.g. year) and cohort c, where "cohort" indexes the different sets of firms <u>treated</u> by each event
    - E.g. different firms might be affected by a change in regulation at different points in time; firms <u>**affected**</u> at one point in time are a 'cohort'

Multiple Events – Approach #1 *[P2]*

$$y_{ict} = \beta d_{ict} + p_t + m_c + u_{ict}$$

d_{ict} = indicator on whether cohort c is affected by time t; this is the interaction between *treatment* & *post*

Time period fixed effects; they will control for *post* dummy in each event

Cohort fixed effects; they are the control for the *treatment* dummy in each event

Multiple Events – Approach #1 *[P3]*

- Intuition of this approach…
  - Every untreated observation at a particular point in time acts as control for treated observations in that time period
    - E.g. a firm treated in 1999 by some event will act as a control for a firm treated in 1994 until itself becomes treated in 1999
  - β will capture <u>average</u> treatment effect across the multiple events

Multiple Events – **Approach #2** *[P1]*

- Now, think of running generalized diff-in-diff for just one of the multiple events…

$$y_{it} = \beta(d_i \times p_t) + \alpha_i + \delta_t + u_{it}$$

  - d_i = indicator for unit i (e.g. firm) being a treated firm in that particular event
  - p_t = indicator for treatment having occurred by period t (e.g. year)
  - Unit i and period t FE control for the independent effects of d_i and p_t

Multiple Events – Approach #2 *[P2]*

- But, contrary to standard difference-in-difference, your sample is…
  - Restricted to a small window around event; e.g. 5 years pre- and post- event
  - And, drops any observations that are treated by <u>another</u> event
    - I.e. your sample starts only with previously untreated observations, and if a 'control' observation later gets treated by a different event, those <u>post</u>-event observations are dropped

Multiple Events – Approach #2 *[P3]*

- Now, create a similar sample for <u>each</u> "event" being analyzed
- Then, "stack" the samples into one dataset and create a variable that identifies the event (i.e. 'cohort') each observation belongs to
  - **Note:** some observation units will appear multiple times in the data [e.g. firm 123 might be a control in event year 1999 but a treated firm in a later event in 2005]

Multiple Events – Approach #2 *[P4]*

- Then, estimate the following on the stacked dataset you've created

$$y_{ict} = \beta d_{ict} + \delta_{tc} + \alpha_{ic} + u_{ict}$$

d_{ict} **= indicator on whether cohort *c* is affected by time *t*;** this is the interaction between *treatment* & *post*

Time-cohort period fixed effects; they control for *post* dummy in each event (i.e. for each 'stack')

Unit-cohort FE; they control for the *treatment* dummy in each cohort (i.e. in each 'stack')

Multiple Events – Approach #2 *[P5]*

- This approach has same intuition of the first approach, but has a couple advantages
  - Can more easily isolate a particular window of interest around each event
    - Prior approach compared all pre- versus post-treatment observations against each other
  - Can more easily extend this into a triple-difference type specification *[more on that later]*

Natural Experiment [P2] – *Outline*

- Difference-in-difference continued…
- How to handle multiple events
- Falsification tests
- Triple differences

Falsification Tests for D-i-D

- Can never directly test underlying identification assumption, but can do some falsification tests to support its validity
  - #1 – Compare pre-treatment observables
  - #2 – Check that timing of observed change in *y* coincides with timing of event *[i.e. no pre-trend]*
  - #3 – Check for treatment reversal
  - #4 – Check variables that shouldn't be affected
  - #5 – Add a triple-difference

#1 – Pre-treatment comparison *[Part 1]*

- Idea is that experiment 'randomly' treats some subset of observations
  - If true, then ex-ante characteristics of 'treated' observations should be similar to ex-ante characteristics of 'untreated' observations
  - Showing treated and untreated observations are comparable in dimensions thought to affect y can help ensure assignment was random

#1 – Pre-treatment comparison *[Part 2]*

- If find ex-ante difference in some variable z, is difference-in-difference is invalid?
  - **Answer** = Not necessarily.
    - We need some story as to why units are expected to have differential trend in y after treatment (for reasons unrelated to treatment) that is correlated with z for this to actually be a problem for identification
    - **And,** even with this story, we could just control for z and it's interaction with time
    - **But, what would be the lingering concern?**

#1 – Pre-treatment comparison *[Part 3]*

- **Answer =** unobservables!
  - If the treated and control differ ex-ante in observable ways, we worry they might differ in unobservable ways that related to some violation of the parallel trends assumption

#2 – Check for pre-trend *[Part 1]*

- Similar to last lecture, can just allow effect of treatment to vary by period to non-parametrically map out the timing
  - "Parallel trends" suggest we shouldn't observe any differential trend prior to treatment for the observations that are eventually treated

#2 – Check for pre-trend *[Part 2]*

- Estimate the following:

$$y_{i,t} = \beta_0 + \beta_1 d_i + \beta_2 p_t + \sum_t \gamma_t (d_i \times \lambda_t) + u_{i,t}$$

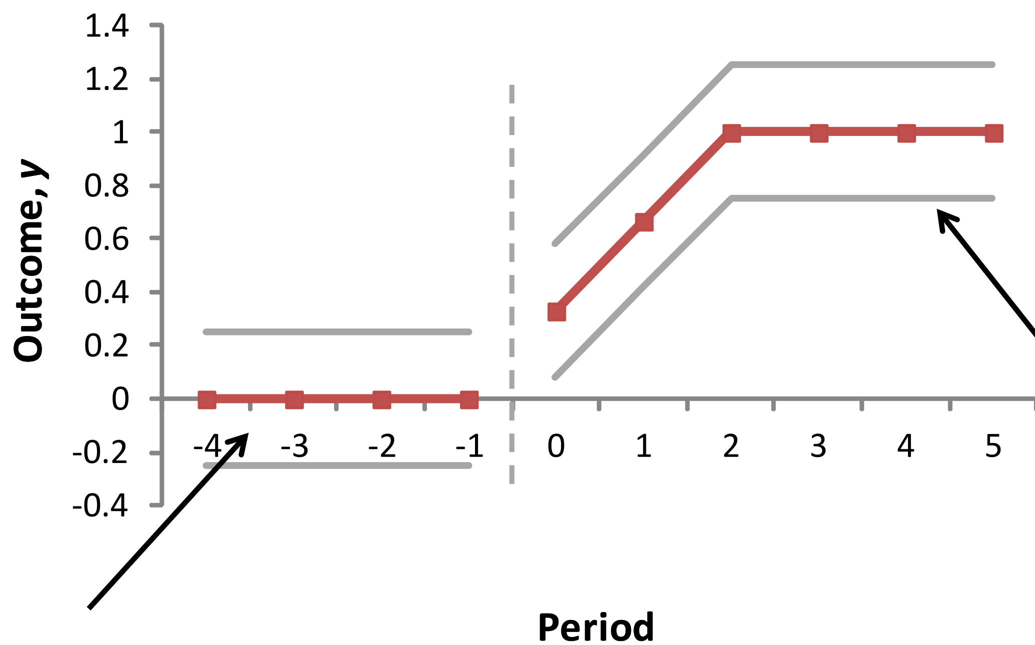
  - d_i and p_t are defined just as before
  - λ_t is indicator that equals 1 if event time = t and zero otherwise, where
    - t = 0 is the period treatment occurs
    - t = -1 is period before treatment

#2 – Check for pre-trend *[Part 3]*

- γ_t estimates change in *y* relative to excluded periods; you then plot these in graph
  - Easiest to **fully saturate** the model (i.e. include λ_t for every period but the very first one); then all estimates γ_t are relative to this period
  - Can also plot confidence interval for each γ_t

#2 – Check for pre-trend *[Part 4]*

- Something like this is ideal…

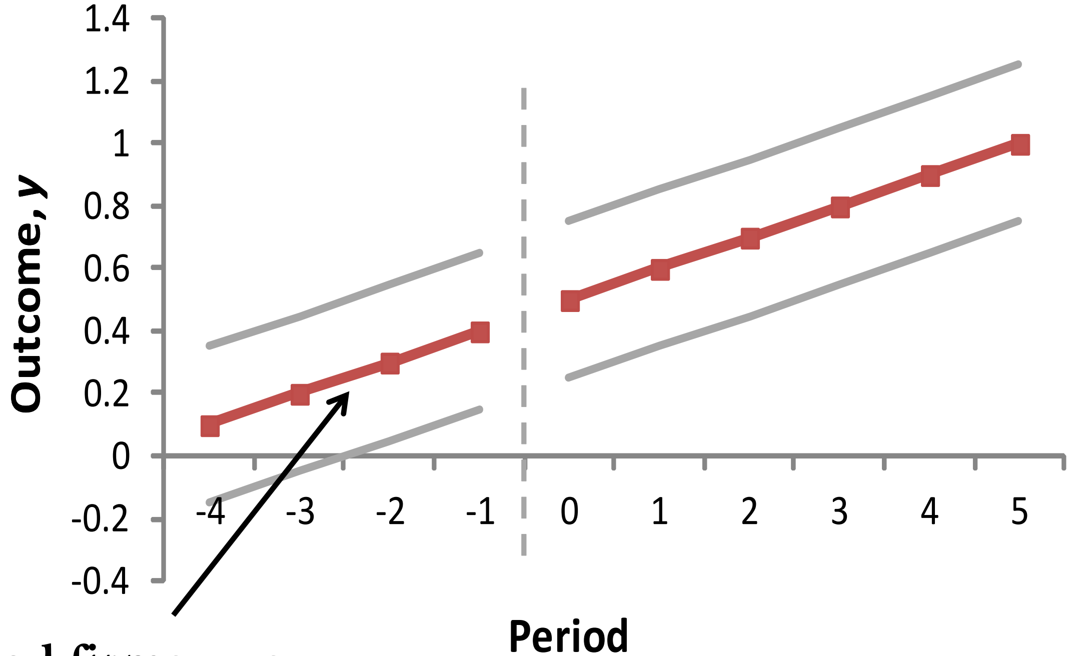


No differential pre-trend

Tight confidence intervals

#2 – Check for pre-trend *[Part 5]*

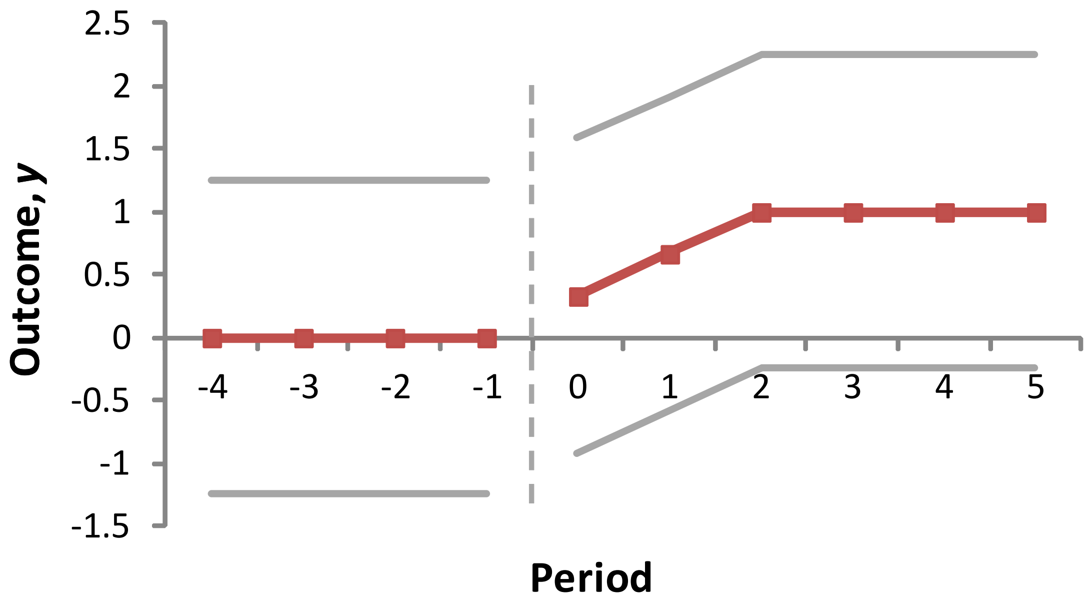
- Something like this is very bad



***y* for treated firms was already going up at faster rate prior to event!**

#2 – Check for pre-trend *[Part 6]*

- Should we make much of wide confidence intervals in these graphs? E.g.



Answer: Not too much… Each period point estimate might be noisy; diff-in-diff will tell us whether post-average y is significantly different then pre-average y

#2 – Check for pre-trend *[Part 7]*

- Another type of pre-trend check done is to do the diff-in-diff in some "random" pre-treatment to show no effect
  - I'm not a big fan of this… Why?
    - **Answer #1** – It is subject to gaming; researcher might choose a particular pre-period to look at that works
    - **Answer #2** – Prior approach allows us to see what the timing was and determine whether it is plausible

#3 – Treatment reversal

- In some cases, the "natural experiment" is subsequently reversed
  - E.g. regulation is subsequently undone
- If we expect the reversal should have the opposite effect, it is good to confirm this

#4 – Unaffected variables

- In some cases, theory provides guidance on what variables should be unaffected by the "natural experiment"
  - If natural experiment is what we think it is, we should see this in the data… so check

#5 – Add Triple difference

- If theory tells us treatment effect should be larger for one subset of observations, we can check this with triple difference
  - Pre- *versus* post-treatment
  - Untreated *versus* treated
  - Less sensitive *versus* More sensitive

This is the third difference

Natural Experiment Outline – *Part 2*

- Difference-in-difference continued…
- How to handle multiple events
- Falsification tests
- Triple differences
  - How to estimate & interpret it
  - Using the popular subsample approach

Diff-in-diff-in-diff – Regression

$$y_{i,t} = \beta_0 + \beta_1 p_t + \beta_2 d_i + \beta_3 h_i + \beta_4 \left(p_t \times h_i\right)$$
$$+\beta_5 \left(d_i \times h_i\right) + \beta_6 \left(p_t \times d_i\right) + \beta_7 \left(p_t \times d_i \times h_i\right) + u_{i,t}$$

- $p_t = 1$ if period t occurs after treatment and equals zero otherwise
- $d_i = 1$ if unit is in treated group and equals zero otherwise
- $h_i = 1$ if unit is group that is expected to be more sensitive to treatment

Diff-in-diff-in-diff – Regression *[Part 2]*

- How to choose and set h_i
  - E.g. If theory says effect is bigger for larger firms; could set $h_i = 1$ if assets of firm in year prior to treatment is above the median size
  - **Note:** Remember to use ex-ante measures to construct indicator if you think underlying variable (that determines sensitivity) might be affected by treatment… Why?
    - **Answer =** To avoid bad controls!

Diff-in-diff-in-diff – Regression *[Part 3]*

$$y_{i,t} = \beta_0 + \beta_1 p_t + \beta_2 d_i + \beta_3 h_i + \beta_4 (p_t \times h_i)$$
$$+ \beta_5 (d_i \times h_i) + \beta_6 (p_t \times d_i) + \beta_7 (p_t \times d_i \times h_i) + u_{i,t}$$

- Easy way to check if done correctly…
  - Should have 8 coefficients (including constant) to capture the 2×2×2=8 different combinations
  - Likewise, a double difference has 4 coefficients (including constant) for the 2×2=4 combinations
- **What do β_6 and β_7 capture?**

Interpreting the estimates *[Part 1]*

- β_6 diff-in-diff estimate for the less-sensitive obs.
  - Captures average differential change in *y* from the pre- to post-treatment period for the less sensitive observations in the treatment group *relative* to the change in *y* for the less sensitive observations in the untreated group

Interpreting the estimates *[Part 2]*

- β_7 is the triple diff estimate; it tells us how much larger effect is for the more sensitive obs.
  - β_7 captures how different the difference-in-difference estimate is for observations considered more sensitive to the treatment
  - **What is total treatment effect for these firms?**
  - **Answer =** $\beta_6+\beta_7$

Tangent – Continuous vs. Indicator?

- Can also do the triple difference replacing h_i with a continuous measure instead of indicator
  - E.g. suppose we expect treatment effect is bigger for larger firms; rather than constructing indicator based on ex-ante size, could just use ex-ante size
  - **What are the advantages, disadvantages of this?**

Tangent – Continuous vs. Indicator?

- Advantages
  - Makes better use of variation available in data
  - Provides estimate on magnitude of sensitivity
- Disadvantages
  - Makes linear functional form assumption; indicator imposes less structure on the data
  - More easily influenced by outliers

Generalized Triple-Difference

- Similar to diff-in-diff, can add in FE to soak up the various terms and improve precision
- E.g. in firm-level panel regression with firm and year fixed effects, you'd estimate

$$
\begin{aligned}
y_{i,t} = {} & \beta_1 \left(p_t \times h_i\right) + \beta_2 \left(p_t \times d_i\right) \\
& + \beta_3 \left(p_t \times d_i \times h_i\right) + \delta_t + \alpha_i + u_{i,t}
\end{aligned}
$$

  - The other terms (including the constant) all drop out; they are collinear with the FE

Natural Experiment [P2] – *Outline*

- Difference-in-difference continued…
- How to handle multiple events
- Falsification tests
- Triple differences
  - How to estimate & interpret it
  - Using the popular subsample approach

Subsample Approach

- Instead of doing full-blown triple-difference, you can also just estimate the double-difference in the two separate subsamples
  - Double-difference for low sensitive obs. (i.e. $h_i = 0$)
  - Double-difference for more sensitive obs. (i.e. $h_i = 1$)
- **Note:** the estimates won't directly match the β_2, $\beta_2 + \beta_3$ effects in prior estimation… **Why?**

Subsample Approach Differences…

- **Answer** = In subsample approach year FE are allowed to differ by sub-sample
  - Therefore, subsample approach is actually controlling for **more** things
  - However, one can easily recover the subsample estimates in one regression (and test the statistical difference) between subsamples by estimating…

Matching Subsample to Combined *[P1]*

$$y_{i,t} = \beta_2 \left(p_t \times d_i \right) + \beta_3 \left(p_t \times d_i \times h_i \right) + \delta_t + \left(\delta_t \times h_i \right) + \alpha_i + u_{i,t}$$

Year FE interacted with sensitivity indicator

- Just add interaction between year FE and indicator for being more sensitivity…
  - This allows for different year FE for each subsample, which is what happened when we estimated the subsamples in two separate regressions

Matching Subsample to Combined *[P2]*

- **In prior regression…**
  - β_2 will equal coefficient from diff-in-diff using just the subsample of less sensitive observations
  - $\beta_2+\beta_3$ will equal coefficient from diff-in-diff using just the subsample of more sensitive observations
  - t-test on β_3 tells you whether effect for more sensitive subsample is statistically different from that of the less sensitive subsample

Triple Diff – Stacked Regression *[Part 1]*

- Another advantage of stacked regression approach to multiple events is ability to more easily incorporate a triple diff
  - Can simply run stacked regression in separate subsamples to create triple-diff or run it in one regression as shown previously

Triple Diff – Stacked Regression *[Part 2]*

- Can’t easily do either of these in approach of Bertrand and Mullainathan (2003)
  - Some observations act as both ‘control’ and ‘treated’ at different points in sample; not clear how create subsamples in such a setting

External Validity – **Final Note**

- While randomization ensures internal validity (i.e. causal inferences), external validity might still be an issue
  - Is the experimental setting representative of other settings of interest to researchers?
    - I.e. can we extrapolate the finding to other settings?
    - A careful argument that the setting isn't unique or that the underlying theory (for why you observe what you observe) is likely to apply elsewhere is <u>necessary</u>

Summary of Today *[Part 1]*

- Diff-in-diff & control variables
  - Don't add controls affected by treatment
  - Controls shouldn't affect estimates, but can help improve precision
- Multiple events are helpful in mitigating concerns about parallel trends assumption

Summary of Today *[Part 2]*

- Many falsification tests one should do to help assess internal validity
  - Ex. #1 – Compare ex-ante characteristics
  - Ex. #2 – Check timing of observed effect
- Triple difference is yet another way to check internal validity and mitigate concerns about identification

In First Half of Next Class

- Regression discontinuity
  - What are they?
  - How are they useful?
  - How do we implement them?
- Related readings… see syllabus

Assign papers for next week…

- Gormley and Matsa (RFS 2011)
  - Risk & CEO agency conflicts
- Becker and Stromberg (RFS 2012)
  - Agency conflicts between equity & debt
- Ashwini (JFE 2012)
  - Investor protection laws & corporate policies

Break Time

- Let's take our 10 minute break
- We'll do presentations when we get back