Natural Experiments - Part 1

Professor Ji-Woong Chung
Korea University

Outline

Outline

Recall... CMI assumption is key

- A violation of conditional mean independence (CMI), such that $E(u|x) \neq E(u)$ precludes our ability to make causal inferences

$$y = \beta_0 + \beta_1 x + u$$

- $Cov(x, u) \neq 0$ implies CMI is violated

CMI violation implies non-randomness

- Another way to think about violation is that it indicates that our x is non-random
  - I.e., the distribution of x (or the distribution of x after controlling for other observable covariates) isn't random
- E.g., firms with high x might have higher y (beyond just the effect of x on y) because high x is more likely for firms with some omitted variable contained in u...

Randomized experiments are great. . .

- In many of the "hard" sciences, the researcher can simply design experiment to achieve the necessary randomness
  - Ex. #1 – To determine effect of new drug, you randomly give it to certain patients
  - Ex. #2 – To determine effect of certain gene, you modify it in a random sample of mice

But we simply can't do them

- We can't do this in finance!
  - E.g., we can't randomly assign a firm's leverage to determine its effect on investment
  - And we can't randomly assign CEOs' # of options to determine their effect on risk-taking
- Therefore, we need to rely on what we call "Natural experiments"

Defining a Natural Experiment

- Natural experiment is basically when some event causes a random assignment of (or change in) a variable of interest, x
  - Ex. #1 – Some weather event increases leverage for a **random** subset of firms
  - Ex. #2 – Some change in regulation reduces usage of options at a **random** subset of firms

NEs Provide Randomness

- We can use such "natural" experiments to ensure that randomness (i.e., CMI) holds and make causal inferences!
  - E.g., we use the randomness introduced into x by the natural experiment to uncover the causal effect of x on y

NEs can be used in many ways

- Technically, natural experiments can be used in many ways
  - Use them to construct IV
    - E.g., gender of first child being a boy used in Bennedsen, et al. (2007) is an example NE
  - Use them to construct regression discontinuity
    - E.g., cutoff for securitizing loans at credit score of 620 used in Keys et al. (2010) is a NE

And the Difference-in-Differences...

- But admittedly, when most people refer to natural experiment, they are talking about a difference-in-differences (DiD) estimator
  - Basically, compares outcome y for a "treated" group to outcome y for "untreated" group where treatment is randomly assigned by the natural experiment

Outline

Outline

Notation and Framework

- Let d equal a treatment indicator from the experiment we will study
  - $d = 0$ untreated by experiment (i.e., control group)
  - $d = 1$ treated by experiment (i.e., treated group)
- Let y be the **potential** outcome of interest
  - $y = y(0)$ for untreated group
  - $y = y(1)$ for treated group
  - Therefore, $y = y(0) + d[y(1) - y(0)]$

Example treatments in corporate finance

- Ex. #1 – Treatment might be that your firm's state passed anti-takeover law
  - $d = 1$ for firms incorporated in those states
  - y could be several things, e.g., ROA
- Ex. #2 – Treatment is that your firm discovers workers exposed to carcinogen
  - $d = 1$ if have exposed workers
  - y could be several things, like M&A

Average Treatment Effect (ATE)

- Can now define some useful things
  - Average Treatment Effect (ATE) is given by $E[y(1) - y(0)]$
- What does this mean in words?
  - Answer: The expected change in y from being treated by the experiment; this is the causal effect we are typically interested in uncovering.

But ATE is unobservable

- Why can't we directly observe ATE?
  - Answer: We only observe one outcome...
  - If treated, we observe $y(1)$; if untreated, we observe $y(0)$. We never observe both.
  - E.g., we cannot observe the counterfactual of what your y would have been absent treatment

Defining ATT

- Average Treatment Effect if Treated (ATT) is given by $E[y(1) - y(0)|d = 1]$
  - This is the effect of treatment on those that are treated; i.e., change in y we'd expect to find in treated random sample from a population of observations that are treated
  - What don't we observe here?
  - Answer: $y(0)|d = 1$

Defining ATU

- Average Treatment Effect if Untreated (ATU) is given by $E[y(1) - y(0)|d = 0]$
  - This is what the effect of treatment would have been on those that are not treated by the experiment
  - We don't observe $y(1)|d = 0$

Uncovering ATE [Part 1]

- How do we estimate ATE, $E[y(1) - y(0)]$?
  - Answer: We instead rely on $E[y(1)|d = 1] - E[y(0)|d = 0]$ as our way to infer the ATE
- In words, what are we doing & what are we assuming?

Uncovering ATE [Part 2]

- In words, we compare average y of treated to average y of untreated observations
  - If we interpret this as the ATE, we are assuming that absent the treatment, the treated group would, on average, have had same outcome y as the untreated group
  - We can show this formally by simply working out $E[y(1)|d = 1] - E[y(0)|d = 0]$...

Uncovering ATE [Part 3]

$$\underbrace{\{E[y(1)|d=1] - E[y(0)|d=1]}_{ATT} + \underbrace{E[y(0)|d=1] - E[y(0)|d=0]}_{Selection\ bias}$$

- Simple comparison doesn't give us the ATE!
- What is the "selection bias" in words?

Outline

Selection bias defined

- Selection bias: $E[y(0)|d = 1] - E[y(0)|d = 0]$
  - Definition: What the difference in average y would have been for treated and untreated observations absent any treatment
  - We do not observe this counterfactual!
- Now let's see why randomness is key!

Introducing random treatment

- A random treatment, d, implies that d is independent of potential outcomes; i.e.,
  - $E[y(0)|d = 1] = E[y(0)|d = 0] = E[y(0)]$
  - $E[y(1)|d = 1] = E[y(1)|d = 0] = E[y(1)]$
- In words, the expected value of y is the same for treated and untreated absent treatment
- With this, easy to see that selection bias $= 0$
- And remaining ATT is equal to ATE!

Random treatment makes life easy

- I.e., with random assignment of treatment, our simple comparison gives us the ATE!
  - This is why we like randomness!
  - But, absent randomness, we must worry that any observed difference is driven by selection bias

Outline

ATE in Regression Format [Part 1]

- Remember $y = y(0) + d[y(1) - y(0)]$.
- Can re-express everything in regression format

$$y = \underbrace{\beta_0}_{E[y(0)]} + \underbrace{\beta_1}_{y(1)-y(0)} d + \underbrace{u}_{y(0)-E[y(0)]}$$

$$E[y|d=1] = \beta_0 + \beta_1 + E[u|d=1]$$
$$E[y|d=0] = \beta_0 + E[u|d=0]$$
$$\Rightarrow \; E[y|d=1] - E[y|d=0] = \beta_1 + (E[u|d=1] - E[u|d=0])$$

ATE in Regression Format [Part 2]

- We are interested in $E[y|d=1] - E[y|d=0]$
  - $\beta_1 = E[y|d=1] - E[y|d=0]$ if $E[u|d=1] - E[u|d=0] = 0$, i.e, no correlation between u and d.
  - $E[u|d=1] - E[u|d=0] = E[y(0)|d=1] - E[y(0)|d=0]$, implying the difference in (no-treatment) potential outcomes between those who get treated and those who don't.
  - We know that this regression gives consistent estimate of β_1 if $cov(d,u) = 0$, i.e., $(E[u|d=1] - E[u|d=0]) = 0$.
  - Hence, selection bias term occurs only if CMI isn't true!

Adding additional controls [Part 1]

- Regression format also allows us to easily put in additional controls, X
  - Intuitively, comparison of treated and untreated just becomes $E[y(1)|d = 1, X] - E[y(0)|d = 0, X]$
  - Same selection bias term will appear if treatment, d, isn't random after conditioning on X
  - Regression version just becomes

$$y = \beta_0 + \beta_1 d + X\Gamma + u$$

  - **Question:** If we had truly randomized experiment, are controls necessary?

Adding additional controls [Part 2]

- **Answer:** No, controls are not necessary in truly randomized experiment
  - But they can be helpful in making the estimates more precise by absorbing residual variation... we'll talk more about this later

Treatment effect – Example

- We want to compare leverage of firms with and without a credit rating [or equivalently, regress leverage on indicator for rating]
  - Treatment is having a credit rating
  - Outcome of interest is leverage

- Why might our estimate not equal ATE of rating?
- Why might controls not help us much?

Treatment effect – Example

- Answer #1: Having a rating isn't random
  - Firms with rating likely would have had higher leverage anyway because they are larger, more profitable, etc.; selection bias will be positive
  - Selection bias is basically an omitted variable!
- Answer #2: Even adding controls might not help if firms also differ in unobservable ways, like investment opportunities

Outline

Outline

Cross-sectional Simple Difference

- Very intuitive idea
  - Compare post-treatment outcome, y, for treated group to the untreated group
  - I.e., just run cross-section simple difference

$$y_{i,t} = \beta_0 + \beta_1 d_i + u_{i,t}$$

    - $d = 1$ if observation i is in treatment group and equals zero otherwise
    - Regression only contains post-treatment time periods
  - What is needed for β_1 to capture the true (i.e., causal) treatment effect?

Identification Assumption

- Answer: $E(u|d) = 0$; i.e., treatment, d, is uncorrelated with the error
  - In words... after accounting for effect of treatment, the expected level of y in post-treatment period isn't related to whether you're in the treated or untreated group
  - I.e., expected y of treated group would have been same as untreated group absent treatment

Multiple time periods & SEs

- If have multiple post-treatment periods, need to be careful with standard errors
  - Errors $u_{i,t}$ and $u_{i,t+1}$ likely correlated if dependent variable exhibits serial correlation
  - E.g., we observe each firm (treated and untreated) for five years after treatment (e.g., regulatory change), and our post-treatment observations are not independent
- Should do one of two things
  - Collapse data to one post-treatment per unit; e.g., for each firm, use average of the firm's post-treatment observations
  - Or cluster standard errors at firm level [We will come back to clustering in later lecture]

Outline

Time-series Simple Difference

- Very intuitive idea
  - Compare pre- and post-treatment outcomes, y, for just the treated group [i.e., pre-treatment period acts as 'control' group]
  - I.e., run time-series simple difference

$$y_{i,t} = \beta_0 + \beta_1 p_t + u_{i,t}$$

    - $p_t = 1$ if period t occurs after treatment and equals zero otherwise
    - Regression contains only observations that are treated by "experiment"
  - What is needed for β_1 to capture the true (i.e., causal) treatment effect?

Identification Assumption

- Answer: $E(u|p) = 0$; i.e., post-treatment indicator, p, is uncorrelated with the error
  - I.e., after accounting for effect of treatment, p, the expected level of y in post-treatment period wouldn't have been any different than expected y in pre-treatment period

Again, be careful about SEs

- Again, if you have multiple pre- and post-treatment periods, you need to be careful with estimating your 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

Using a First-Difference (FD) Approach

- Could also run regression using first-differences specification

$$y_{i,t} - y_{i,t-1} = \beta_1(p_t - p_{t-1}) + (u_{i,t} - u_{i,t-1})$$

  - If just one pre- and one post-treatment period (i.e., $t-1$ and t), then will get identical results
  - But, if more than one pre- and post-treatment period, the results will differ...

FD versus Standard Approach [Part 1]

- Why might these two models give different estimates of β_1 when there are more than one pre- and post-treatment periods?

$$y_{i,t} = \beta_0 + \beta_1 p_t + u_{i,t}$$

versus

$$y_{i,t} - y_{i,t-1} = \beta_1 (p_t - p_{t-1}) + (u_{i,t} - u_{i,t-1})$$

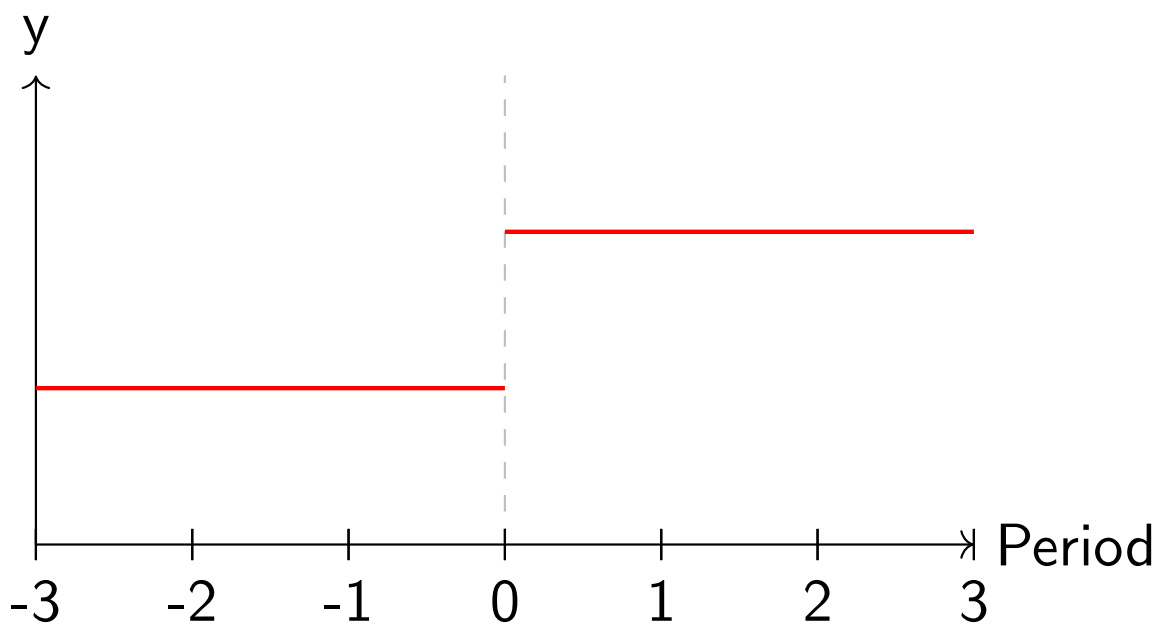
FD versus Standard Approach [Part 2]

Answer:

- In the first regression, β_1 captures the difference between average y pre-treatment versus average y post-treatment
- In the second regression, β_1 captures the difference in Δy immediately after treatment versus Δy in all other pre- and post-treatment periods
  - I.e., the Δp variable equals 1 only in immediate post-treatment period, and 0 for all other periods.

FD versus Standard Approach [Part 3]

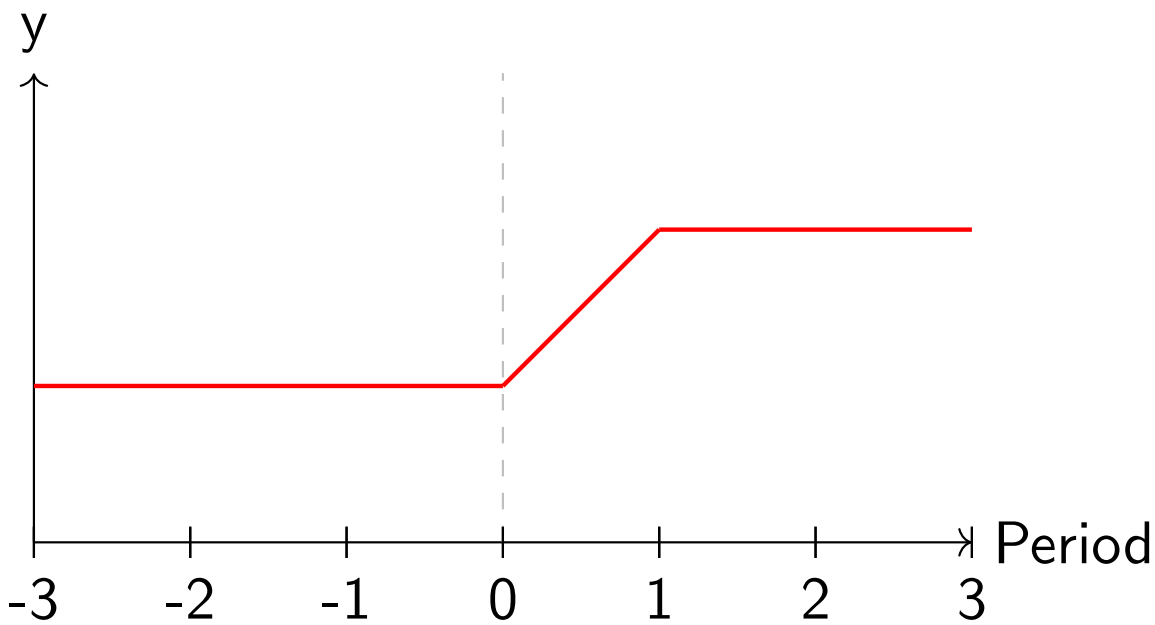
- Both approaches assume the effect of treatment is immediate and persistent.



  - In this scenario, both approaches give the same estimate

FD versus Standard Approach [Part 4]

- But suppose the following is true...



  - In this scenario, the FD approach gives a much smaller estimate
  - The first approach compares average pre- versus post-treatment
  - FD compares Δy from $t = 0$ to $t = -1$ against Δy elsewhere (which isn't always zero!)

Correct way to do difference

- Correct way to get a 'differencing' approach to match up with the more standard simple difference specification in multi-period setting is to instead use

$$\bar{y}_{i,\text{post}} - \bar{y}_{i,\text{pre}} = \beta + \bar{u}_{i,\text{post}} - \bar{u}_{i,\text{pre}}$$

  - This is exactly the same as simple difference

Outline

Treatment effect isn't always immediate

- In prior example, the specification is wrong because the treatment effect only slowly shows up over time
  - Why might such a scenario be plausible?
  - Answer = Many reasons. E.g., firms might only slowly respond to change in regulation, or CEO might only slowly change policy in response to compensation shock

Accounting for a delay...

- Simple-difference misses this subtlety; it assumes effect was immediate
- For this reason, it is always helpful to run regression that allows effect to vary by period
  - How can you do this?
  - Answer = Insert indicators for each year relative to the treatment year

Non-parametric approach

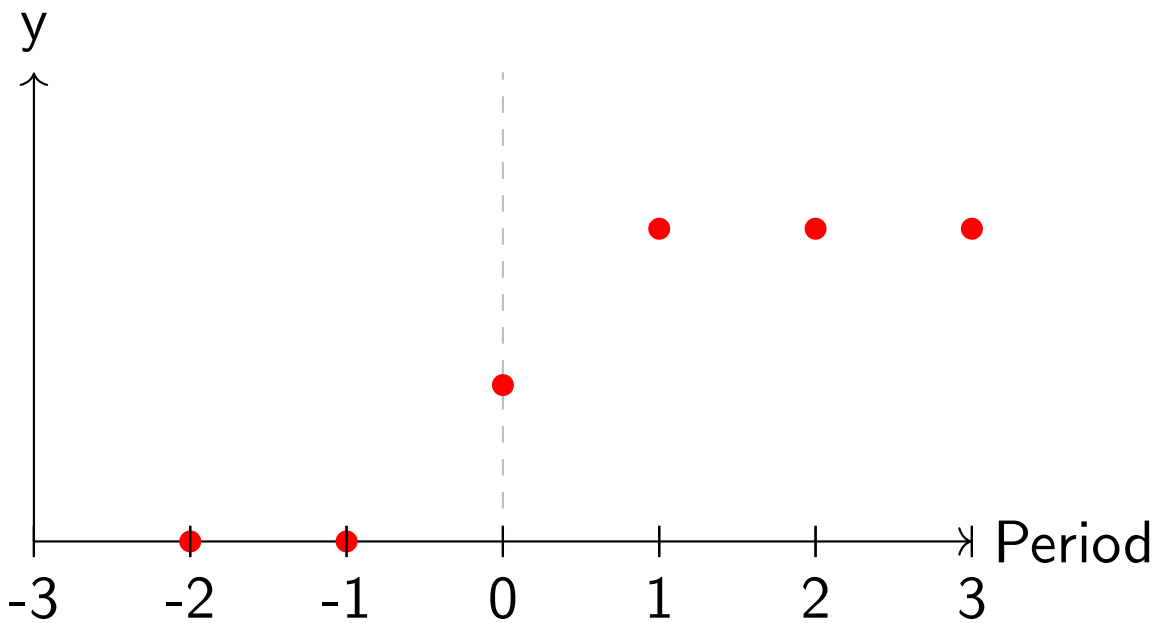
- If have 5 pre- and 5 post-treatment observations; could estimate:

$$y_{i,t} = \beta_0 + \sum_{t=-4}^{t=5} \beta_t p_t + u_{i,t}$$

  - p_t is now an indicator that equals 1 if year $= t$ and zero otherwise; e.g. $t = 0$ is the period treatment occurs. $t = -1$ is period before treatment
  - β_t estimates change in y relative to excluded periods; you then plot these in graph

Non-parametric approach – Graph

- Plot estimates to trace out effect of treatment



  - It allows effect of treatment to vary by year!
  - Pre-period y was same as y in excluded period ($t = -3$)
  - Post-period estimates capture change relative to excluded period ($t = -3$)
  - Could easily plot confidence intervals as well

Simple Differences – Advice

- In general, simple differences are not that convincing in practice. . .
  - Cross-sectional difference requires us to assume the average y of treated and untreated would have been same absent treatment
  - Time-series difference requires us to assume the average y would have been same in post- and pre-treatment periods absent treatment
- Is there a better way?

Outline

Outline

Difference-in-differences

- Yes, we can do better!
- We can do a difference-in-differences that combines the two simple differences
  - Intuition = compare change in y pre- versus post-treatment for treated group [1st difference] to change in y pre- versus post-treatment for untreated group [2nd difference]

Implementing diff-in-diff

- Difference-in-differences estimator

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

  - $p_t = 1$ if period t occurs after treatment and equals zero otherwise
  - $d_i = 1$ if unit is in treated group and equals zero otherwise
  - What do β_1, β_2, and β_3 capture?

Interpreting the estimates

- Here is how to interpret everything...
  - β_1 captures the average change in y from the pre- to post-treatment periods for the untreated groups
  - β_2 captures the average difference in level of y of the treated group in the pre-treatment period
  - β_3 captures the average differential change in y from the pre- to post-treatment period for the treatment group relative to the change in y for the untreated group
  - β_3 is what we call the diff-in-diffs estimate

Comparing group means [Part 1]

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

- Four possible combinations:

$$\begin{aligned}
E(y|d=1, p=1) &= \beta_0 + \beta_1 + \beta_2 + \beta_3 \\
E(y|d=1, p=0) &= \beta_0 + \beta_2 \\
E(y|d=0, p=1) &= \beta_0 + \beta_1 \\
E(y|d=0, p=0) &= \beta_0
\end{aligned}$$

- Assumption: $E(u|d, p) = 0$, i.e., the "experiment" is random.

Comparing group means [Part 2]

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

These can be arranged in two-by-two table

	Post (1)	Pre (2)	Diff (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
Diff. $(a) - (b)$	$\beta_2 + \beta_3$	β_2	β_3

This is why it's called the difference-in-differences estimate; regression gives you same estimate as if you took differences in the group averages. Again, β_3 has a causal interpretation when $E(u|d, p) = 0$.

Simple difference –Revisited [Part 1]

Useful to look at simple differences

	Post (1)	Pre (2)	Diff (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
Diff. $(a) - (b)$	$\beta_2 + \beta_3$	β_2	β_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 (1)	Pre (2)	Diff (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
Diff. $(a) - (b)$	$\beta_2 + \beta_3$	β_2	β_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

Outline

"Parallel trends" assumption

- Identification assumption is what we call the parallel trends assumption
  - Absent treatment, the change in y for treated would not have been different than the change in y for the untreated observations
  - But we cannot test this!
  - Typically, we examine the "pre-trend" and hope that the trend would continue after treatment.
- Looking at what difference-in-differences estimate is doing in graphs will also help you see why the parallel trends assumption is key

Why we like diff-in-diff [Part 1]

- With simple difference, any of the below arguments would prevent causal inference
  - Cross-sectional diff – "Treatment and untreated avg. y could be different for reasons a, b, and c, that just happen to be correlated with whether you are treated or not"
  - Time-series diff – "Treatment group's avg. y could change post-treatment for reasons a, b, and c, that just happen to be correlated with the timing of treatment"

Why we like diff-in-diff [Part 2]

- But now the required argument to suggest the estimate isn't causal is...
  - "The change in y for treated observations after treatment would have been different than the change in y for untreated observations for reasons a, b, and c, that just happen to be correlated with **both** whether you are treated and when the treatment occurs"
- This is (usually) a harder story to tell

Example...

- Bertrand & Mullainathan (JPE 2003) uses state-by-state changes in regulations that made it harder for firms to do M&A
  - They compare wages at firms pre- versus post-regulation in treated versus untreated states

Are these concerns for internal validity?

- The regulations were passed during a time period of rapid growth of wages nationally...
  - Answer = No. Indicator for post-treatment accounts for common growth in wages
- States that implement regulation are more likely have unions, and hence, higher wages...
  - Answer = No. Indicator for treatment accounts for this average difference in wages

Example continued...

- However, ex-ante average differences is troublesome in some regard...
  - Suggests treatment wasn't random
  - And ex-ante differences can be problematic if we think that their effect may vary with time...
    - Time-varying omitted variables **are** problematic because they can cause violation of "parallel trends"
    - E.g., states with more unions were trending differently at that time because of changes in union power

Summary of Today [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

Summary of Today [Part 2]

- Simple differences require strong assumptions; typically, not plausible
- Difference-in-differences helps with this
  - Compares change in y pre- versus post-treatment for treated to change in y for untreated
  - Requires "parallel trends" assumption