Week 4: Difference-in-differences

Marcelo Coca Perraillon

University of Colorado
Anschutz Medical Campus

Health Services Research Methods I
HSMP 7607
2020

Outline

- Introduction to difference-in-differences
- The role of time and policy changes or interventions
- Two periods: John Snow and cholera
- Differencing to cancel out constant additive effects
- Multiple periods
- Parametric estimation
- Examples and formal assumptions

Big picture

- So far we have covered methods in which we need to assume ignorability (selection on observables, no unmeasured confounders, exchangeability)
- To use regression adjustment, we also needed to assume overlap. The propensity score helped us diagnose the problem and also find some solutions: matching, IPW, stratification
- We will see that matching precedes the propensity score, but matching estimators have received renewed attention and are being more used. They solve the overlap problem by restricting estimation to the region where there is common support, which makes treatment effects a form of LATE
- In today's lecture and next week we will move away from ignorability: situation in which we can't argue anymore that we have ignorability or are even close to it
- In these situations, we need to get creative to find causal effects. These designs are often called **quasi-experimental designs** or **natural experiments**: difference-in-difference (DiD), regression discontinuity (RD), and instrumental variables (IV)

Big picture II

- Just to avoid confusion: these designs can be combined, and there are many connections among them
- DiD and propensity score (IPW) are a common combo
- You can do some DiD restricting samples to regions where there is comparability, invoking insights from RDD
- One type of RDD ("fuzzy RDD") follows the same logic and estimation of IV

John Snow in the times of cholera (circa 1850)

- Sometimes it helps to back in time to the earliest application of a method in its simplest form, so we'll talk about John Snow (not the GoT one)
- Snow was interested in understanding if cholera was transmitted by water rather than air
- One district in London had changed water source. After the change, cholera deaths decreased, which would suggest that water was the source of infection
- Let's call the houses that changed water source the "treated" group
- One form of evidence would be to do a before and after comparison: $E[Y_i|D_i = 1, T_i = 1] - E[Y_i|D_i = 1, T_i = 0]$, where T is 1 if after the change of supplier, Y is an indicator for death and D is our treatment indicator
- The key question, of course, is whether $E[Y_i|D_i = 1, T_i = 0] = E[Y_{0i}|D_i = 1, T_i = 1]$
- Is the observed outcome before the change ($T = 0$) a good prediction of the counterfactual after the change ($T = 1$)? Dubious

John Snow in 1850s and cholera

- Remember, if the same people are measured before and after, they are different units in our causal inference framework
- Intuitively, before and after comparisons are valid if nothing else changed at the same time as the treatment; that is, now we need to assume that no other factor X is correlated with T (contemporaneous factors, trends)
- In Snow's case a simple before and after comparison did not solve the problem because the suspicion was that air was a source of contamination. What if air changed at the same time as the water supply was changed? How could he "hold air constant"?
- He came up with a clever solution: use control areas and hold air constant by using as controls places where water supply did not change with the catch that these control houses shared the same air
- Snow actually used a salt test to verify water source. See details here `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7150208/`

Comparisons

- Let's simplify the notation using only realized outcomes for now: A pre and post comparison of outcome Y for the treated is: $E[Y_{tpost}] - E[Y_{tpre}]$
- We want to compare that difference with the difference in the control: $E[Y_{cpost}] - E[Y_{cpre}]$
- So the estimate of interest is:

  $\Delta_{DiD} = E[Y_{tpost}] - E[Y_{tpre}] - \{[Y_{cpost}] - E[Y_{cpre}]\}$
- If that difference of differences is zero, that would suggest that water is not causing infections
- If that difference is not zero, then there is some evidence that air is not the source of transmission – it doesn't **prove** that water causes cholera – there could still be other factor, such as physical contact, that could explain transmission
- The **estimator** is a difference-in-differences, hence the name. We need four expected values (four "cells")

Differencing within each group over time

- The Snow example provides some intuition on why this approach works but we need to elaborate to make it clearer
- Comparing before and after the houses (**within** the treated group) that changed water supplier helps us with things that that did not change over time
- For example cleanliness in each house or genetic factors or the role of sex and race
- All those fixed or constant factors could not explain the before and after change in the outcome since these factors did not change before and after
- Same within the control group

Differencing between groups

- The same logic applies to comparisons **between** groups
- A comparison between groups is valid if one group is a good conterfactual of the other
- But we can relax that assumption and argue that we do not need to assume that houses that changed water supplier are the same as those who did not
- We could assume instead that they are in fact different, but the factors that affect cholera mortality between the groups did not change when we do the before and after comparison
- For Snow, the key factor was air. We can think of other important factors in these COVID times: one group was more prone to parties than the other
- As long as we can argue that "party proneness" was the same before and after, we are fine

Two imperfect solutions make a good one

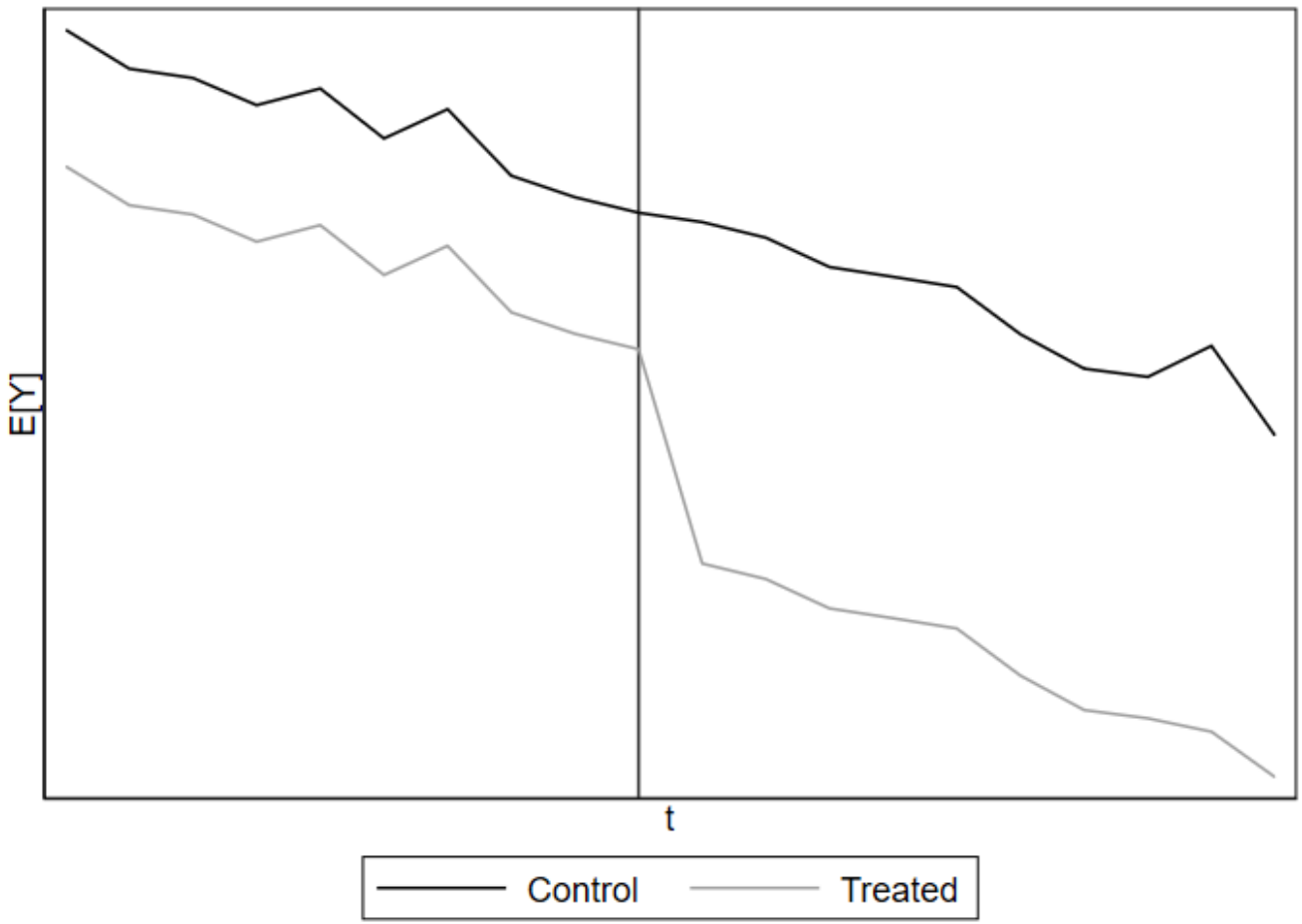
- Note that differences-in-differences is one interesting case in which two imperfect solutions combined make something useful
- A before and after analysis of outcomes is not ideal since it could be hard to argue that nothing else changed
- A comparison between groups that are not comparable is not ideal because other factors could account for differences between the groups
- But when we combine them, a difference-in-differences estimator may give you something right provided the other factors are **additive and constant**
- It's not a mild assumption, but it can be relaxed by controlling for other covariates (more in a bit)

Key elements

- In the examples that follow the key elements are:
  1. **Time**. In DiD we always have a time: a before and after period. So far we have talked about only two periods, but we could have multiple periods before and after (more information)
  2. **Comparison groups**: In DiD, one group receives the intervention or is subjected to the policy change *only in the post-period*. *These groups do not need to be comparable*
  3. **Fixed factors**: We assume that important factors that explain the outcome Y are fixed during the pre and post periods. If observed, we can control for those factors that could affect trends
- Trends are key in DiD

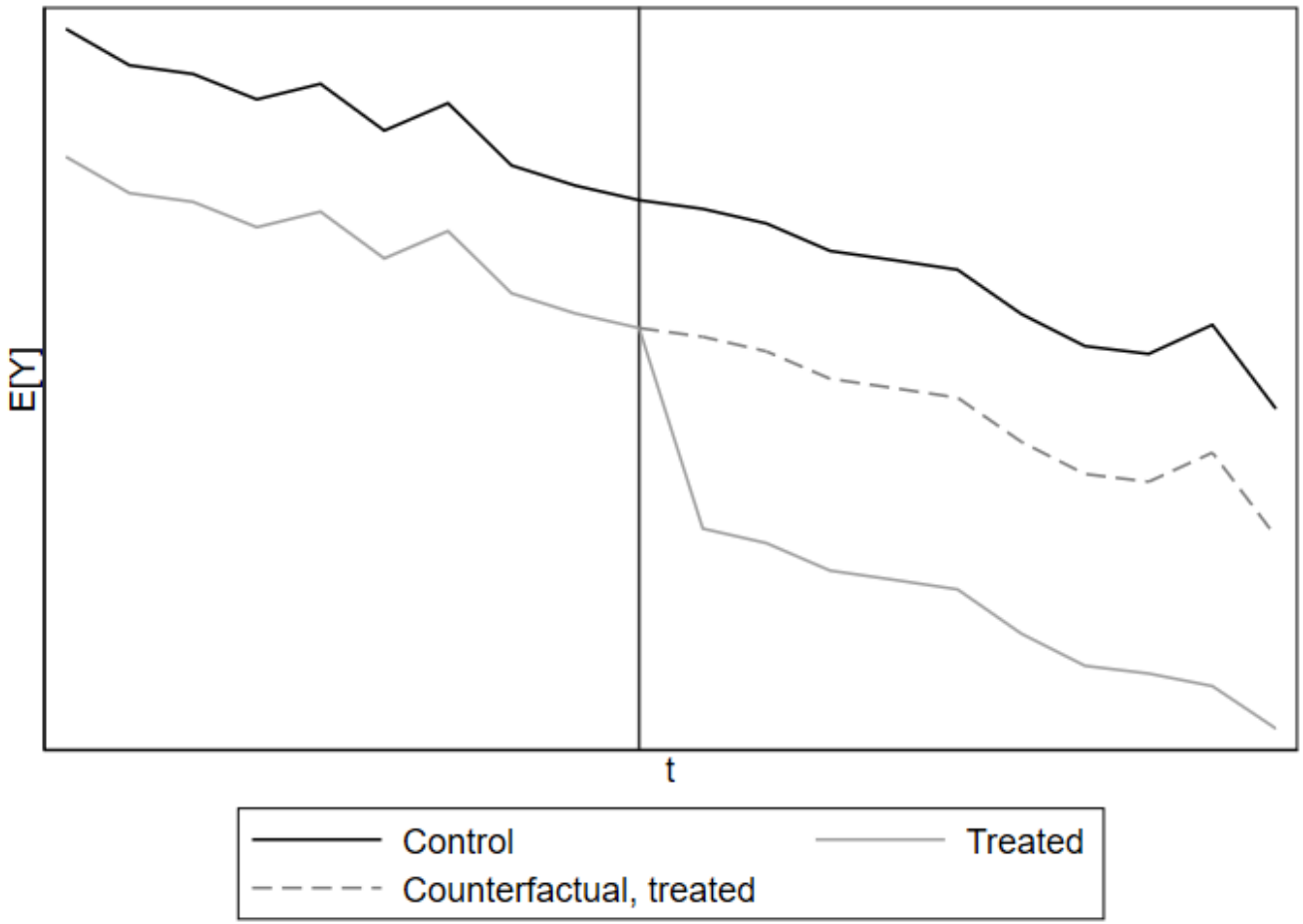
More periods

- This would be in ideal DiD scenario with constant (parallel) trends before the intervention at the time marked by the vertical line (I simulated data)



More periods

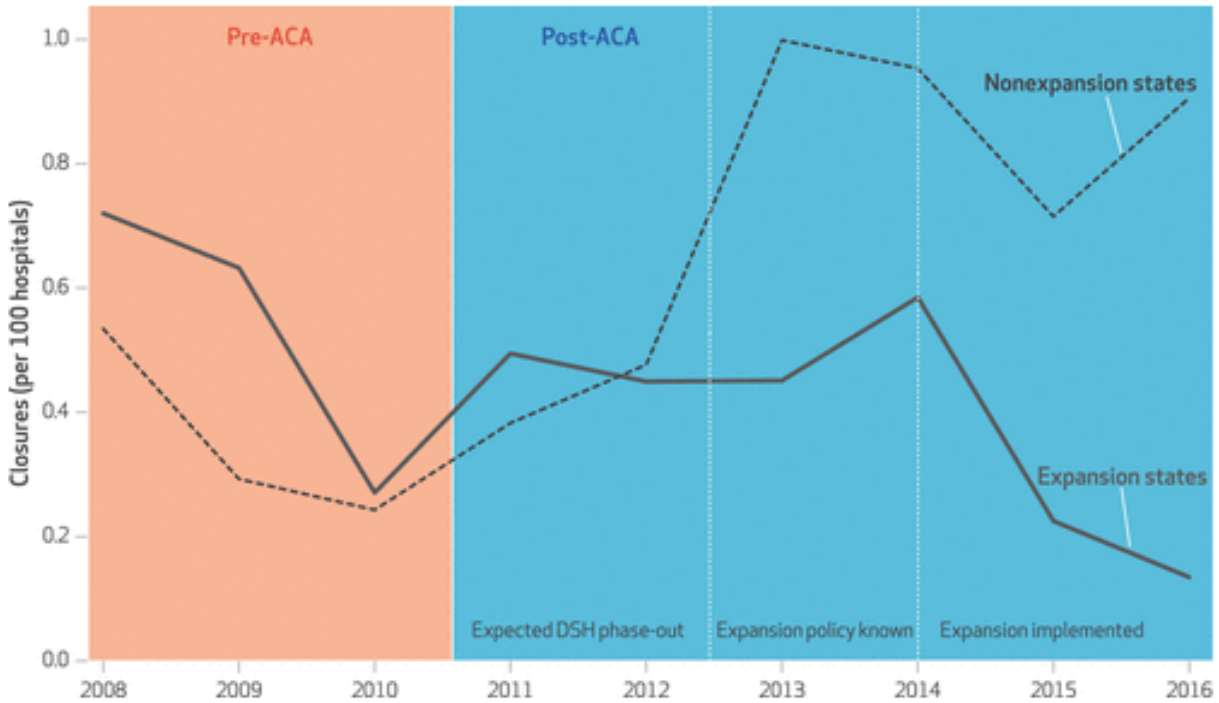
- The counterfactual for the treated is the dashed line: trends would have remained constant over time with no treatment, which is the underlying key assumption



Actual example

- Medicaid expansion and hospital closures (Lindrooth et al., 2018)

Exhibit 2 Unadjusted hospital closure rates by state Medicaid expansion status, 2008–16



SOURCE Authors' analysis of data from the Centers for Medicare and Medicaid Services. NOTE Closures were independently validated from multiple sources.

- Available at: `https://www.healthaffairs.org/doi/full/10.1377/hlthaff.2017.0976`

More formal

- We can formalize the previous discussion by assuming that the outcome follows this structure:

  $Y_{it} = c_i + d_t + \delta D_{it} + \eta_{it}$
- This is a causal structural assumption. c are d are variables, not coefficients. η_{it} is an unexplained random cause of variation, where i indexes the unit of observation and t indexes time
- The outcome depends on **constant** (**fixed**, **time-invariant**) factors at the unit of observation level (c_i) and factors that depend on time (d_t) but not on unit of observation i
- Think of c and d as confounding variables with a coefficient of 1. They are **unobserved effects**. We could put a coefficient next to them, but we won't estimate them. We could add more of both, say: $\sum_{j=1}^{w} c_{ij}$ and $\sum_{k=1}^{m} d_{kt}$ so think of c and d representing more than one factor
- δ is the *difference* between groups and is constant, so we assume **homogeneous** treatment effects

Potential outcomes

- Since $Y_{it} = c_i + d_t + \delta_{it} D_{it} + \eta_{it}$ represents a causal relationship, they also define potential outcomes:

  $Y_{1it} = c_i + d_t + \delta + \eta_{it}$

  $Y_{0it} = c_i + d_t + \eta_{it}$
- So $Y_{1it} - Y_{0it} = \delta_{it}$. Note that δ has an index because we could define the difference between potential outcomes at different times
- With two periods, $t \in \{0, 1\}$. $t = 0$ is before the intervention

  (Notation here gets messy with potential outcomes. This notation for potential outcomes would be better: $Y(1)_{it}$ if treated, $Y(0)_{it}$ if not treated)

Differencing

- We could do this with potential outcomes or realized outcomes. I'll do it with realized outcomes

  (1) Treated group after and before: $E[Y_{i1}|D_i = 1] - E[Y_{i0}|D_i = 1] = c_i + d_1 + \delta_1 - (c_i + d_0 + \delta_0) = d_1 - d_0 + (\delta_1 - \delta_0) = d_1 - d_0 + \delta$
- The above shows the problem of a simple before and after comparison: it doesn't get rid of contemporaneous factors that depend on time (trends) and would affect an estimate of treatment effects on the treated, measured by the variable d: $d_1 - d_0$
- That could be a measure of air in both periods that changed between $t = 0$ and $t = 1$

  (2) Control group after and before: $E[Y_{i1}|D_i = 0] - E[Y_{i0}|D_i = 0] = c_i + d_1 - (c_i + d_0) = d_1 - d_0$
- The difference of the differences **(1)-(2)** $= d_1 - d_0 + \delta - (d_1 - d_0) = \delta$
- The within group differencing got rid of factor c, the between group differencing got rid of d. It would be the same if we had aded more factors c's nd d's

Notes

- We assumed $E[\eta_{it}|D_{it} =] = E[\eta_{it}]$, that is mean independence
- That assumption could be relaxed by saying that the unobserved error component is mean independent respect to the change before and after ($E[\Delta\eta_i|D_{it}] = E[\Delta\eta_i]$), which is what you would get using "fixed effects" in longitudinal data
- Separating the justification of the difference-in-difference estimator from estimation is more helpful. We could estimate the four expected values separately, following the logic of estimating treatment effects in previous lectures (we will see a version that combines propensity score with kernel weighting; Heckman, Ichimura, Todd, 1998)
- The constant trend is important, and one that we can't evaluate if we don't have more observations before the intervention took place
- The two-groups, two-periods example is **helpful for the intuition but not for evaluating the assumption of constant trends**

What about other factors (covariates)?

- If we follow the logic of differencing, then we do not need to account for any other observed or unobserved constant (fixed, **time-invariant**) factors
- But we could take into account factors that vary at the unit of observation and by time (**time-varying covariates**)
- This means that we can extend our notation to condition for a vector of covariates $\mathbf{X}_{it}$, although we will imposed some assumptions when we use regression analysis (exogeneity)
- So now we have: $Y_{it} = c_i + d_t + \delta D_{it} + \mathbf{X}'_{it}\beta + \eta_{it}$

Assumptions

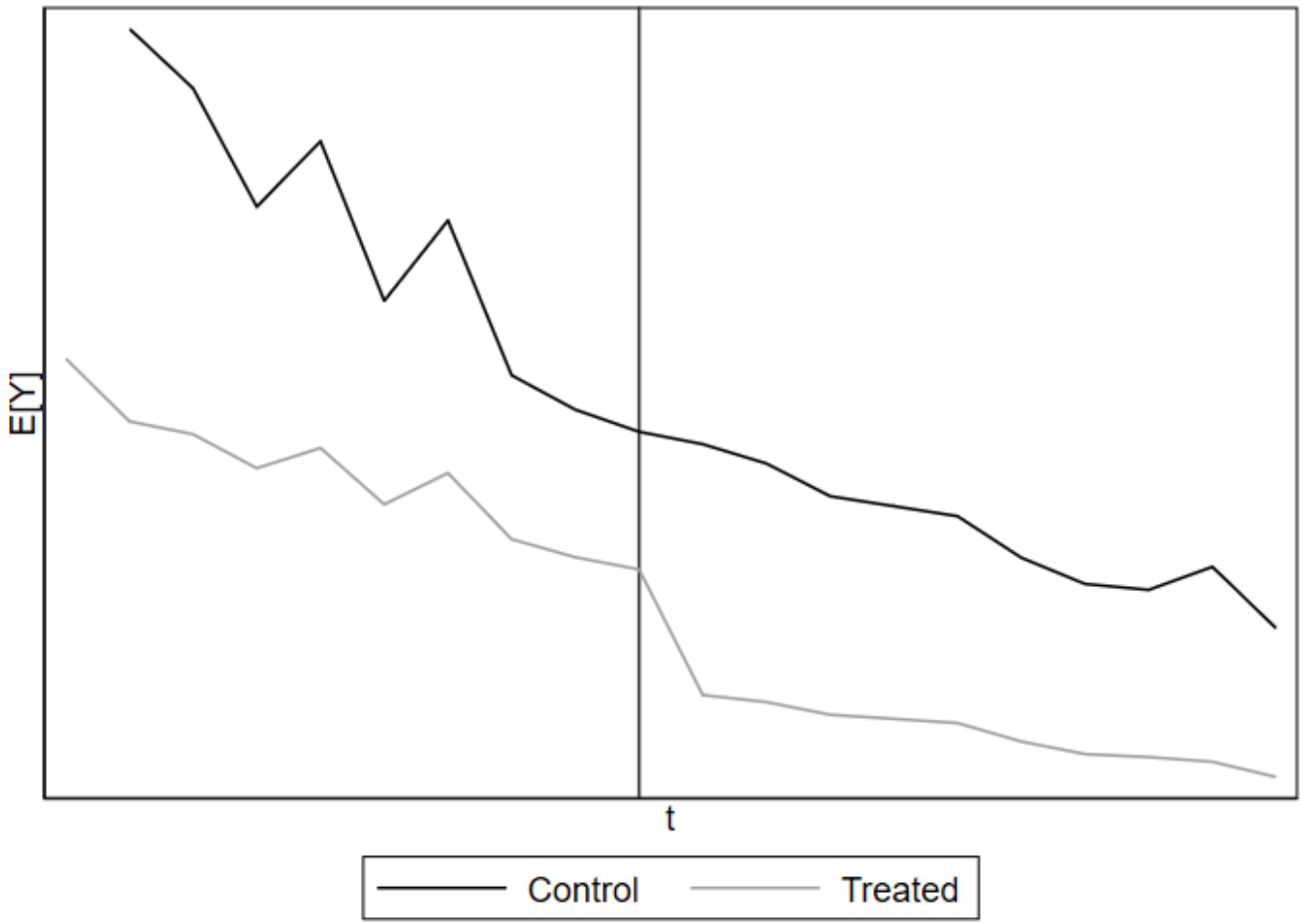
- We can now state the assumptions following Lechner (2010)
  1. **SUTVA**. We always need SUTVA but in DiD it tends to be more relevant. No interference (spillovers) and variation in treatment. Think about this in the context of cholera and Medicaid expansions
  2. **Exogeneity**: The covariates **X** are not influenced by the treatment, so we can condition on them
  3. **Common trends** or constant bias. If the treated had not been treated, both treatment and control groups would have the same trends over time (after controlling for other factors). Constant bias is the same assumption. Treated and control groups are different, but that difference remains constant over time
- The last assumption could be divided into an assumption about observed parallel trends before the intervention and the idea that "shocks" have a common effect in both groups
- With two or mode pre-periods, we can **test** the parallel trend assumption

Testing parallel trends

- We will see two ways of testing parallel trends. We need at least two pre-period time points:
  1. Using the pre-period data only: test if changes in trends are the same in both groups
  2. Placebo tests: using lags and leads, past treatment predicts future outcome, but future treatment should not predict present changes in the outcome
- The first approach is much more intuitive. With two pre-periods only, say, $t-1$ and $t=0$ the test is whether the change $E[Y_{t=0} - Y_{t=-1}]$ is the same in the treatment and control groups. The relevant test is an interaction, so it follows the parametric structure of DiD models
- The second one is less intuitive and it estimates more parameters. We need to use lag and lead variables for the treatment indicator (Stata of course make this easy)
- One aspect of the parallel trend assumption is annoying: if it holds in Y then it doesn't hold in retransformations like $log(y)$ (a monotonic transformation)

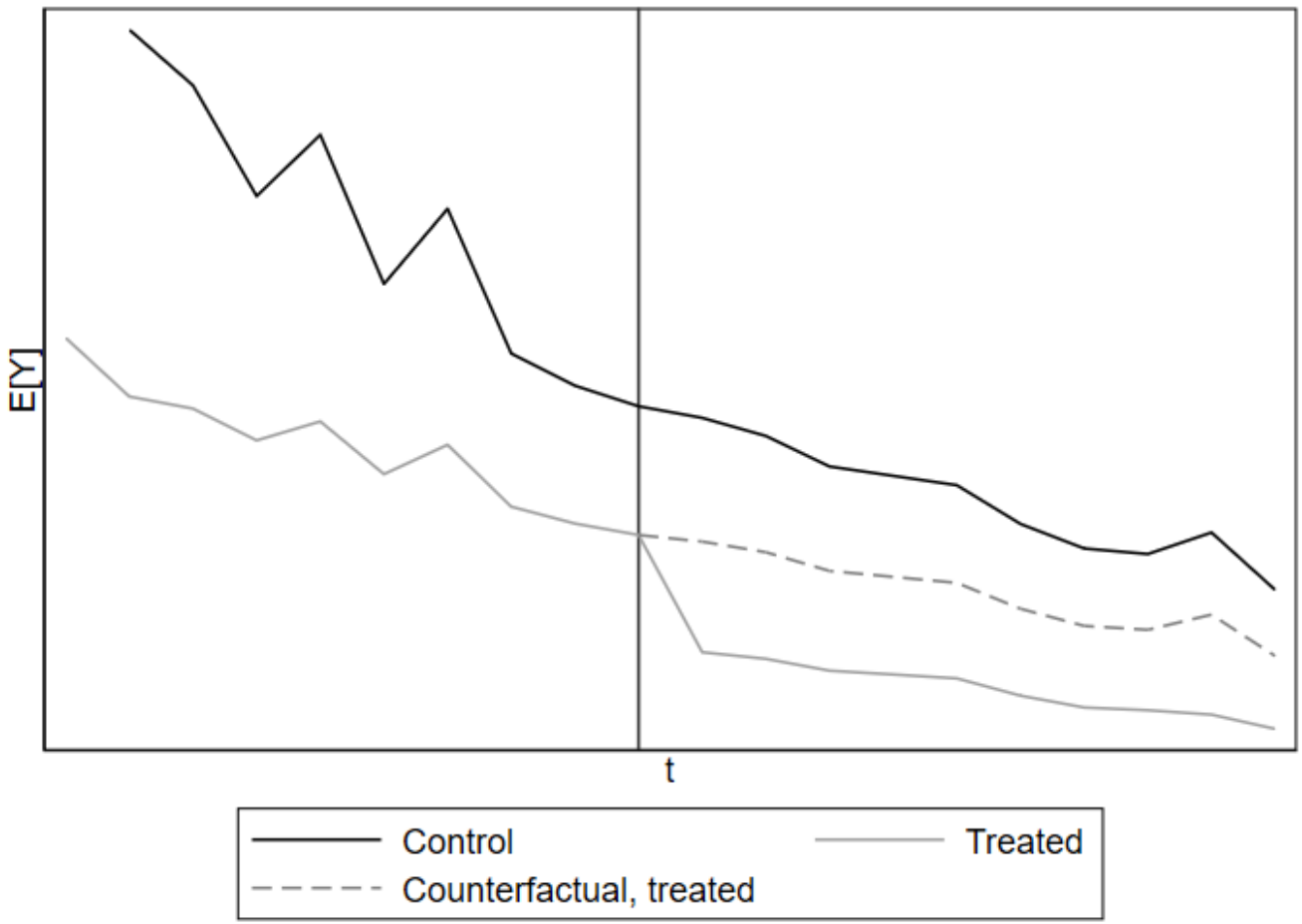
Yikes?

- The perfect example of parallel trends doesn't hold in the log scale. I simply took the log of Y in the simulated data



Yikes? II

- Look at the counterfactual



Parametric regression

- We can estimate models like:

  $$Y_{it} = \beta_0 + \beta_1 D_i + \beta_2 P_t + \beta_3 (D_i \times P_t) + \epsilon_{it}$$

  $$Y_{it} = \beta_0 + \beta_1 D_i + \beta_2 P_t + \beta_3 (D_i \times P_t) + \mathbf{X}'_{it}\beta + \epsilon_{it}$$

- The unit of analysis i could be a person or state. P is an indicator that equals 1 after the change. Note the index it carefully. Treatment D doesn't depend on time (a person i is treated or control in both periods), while P depends on time but it's the same by person
- The first model is a saturated model; we will get four predicted means
- We are going to come back to the structure of the data, which is reflected in the index, because it's important for inference and can be confusing

Person id	D	P	Y
101	1	0	100
101	1	1	120
104	0	0	90
104	0	1	92
.	.	.	.
.	.	.	.

Model interpretation

- Why is that model a difference-in-difference? (Using the version without covariates; with covariates we need to hold them constant)

  $E[Y]$ Treated in post period: $\beta_0 + \beta_1 + \beta_2 + \beta_3$

  $E[Y]$ Treated in pre period: $\beta_0 + \beta_1$

  (1) Difference treated post - pre: $\beta_2 + \beta_3$

  $E[Y]$ Control in post period: $\beta_0 + \beta_2$

  $E[Y]$ Control in pre period: β_0

  (2) Difference control post - pre: β_2

  Difference of differences (1)-(2): $\Delta_{DiD} = \beta_3$

- **Caution**: Interacted models are not difference-in-differences **research designs**, but interactions with dummy variables are difference-in-differences

This is not a DiD research design

- Suppose you are interested in evaluating the effects of a new preventive care benefit offered at no cost to patients (e.g. Medicare's annual wellness visit, AWV) on race disparities
- The outcome is a measure of preventive care. The AWV became available in 2011. For simplicity, let's say that you want to compare Whites vs other races. You could estimate the following model:

  $PC_{it} = \beta_0 + \beta_1 White_i + \beta_2 Post2011_t + \beta_3 Post2011_t * White_i + \eta_{it}$
- β_3 is a difference of differences. If not zero, then disparities after the AWV are different that disparities before
- But this is a before and after analysis. **There is no comparison group; all races were exposed to the treatment –the AWV**. The *outcome* is a comparison. It just happens that the parametric model structure is the same as a simple DiD design
- See a related model structure in Lind et al. (2018): `https://onlinelibrary.wiley.com/doi/full/10.1111/jgs.15494`

More details

- Note that with DiD designs we are exploiting longitudinal (aka panel in econometrics) data: could be cross-sectional data (difference units at difference times) or the same unit measured at different times
- Each type of longitudinal data has implication for inference
- We need to revise our assumption of i.i.d (no longer independent) errors $\epsilon_i \sim N(0, \sigma_2)$
- We will address this issue later but the "fixed" factors that we control for help us solve this problem (in some cases)
- We will review "fixed" effects regression and the connection between **demeaning** and **first-differencing**