Lecture 26

11-22-2021

Announcements

There is no class or section on 11/24 (Happy Holidays)

There is no class on 11/29

WP10 and Checkpoint 13 redos are due on 11/24

PS5 Makeups are due today

Extended lecture slides for Wednesday will be up on bCourses

Differences in Differences

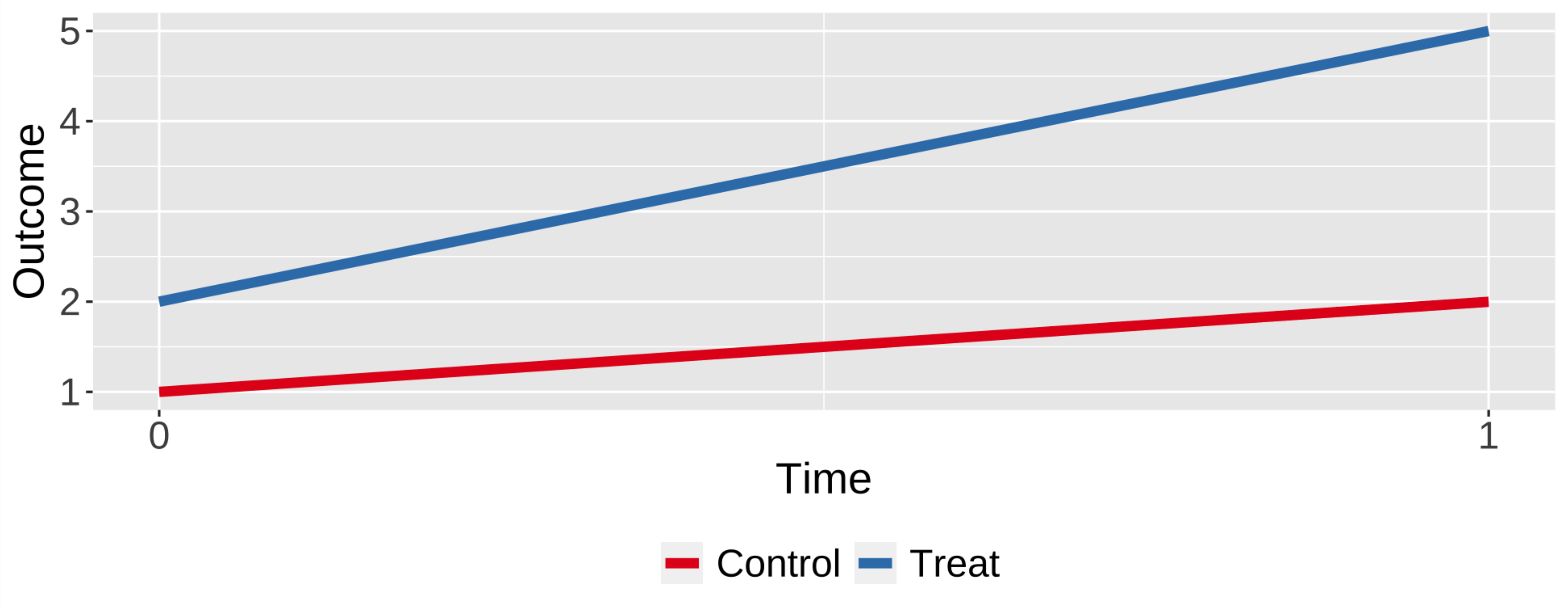
We have been thinking in class about situations where we have random assignment.

Sometimes though treatment and control outcomes move in parallel in the absence of treatment.

The divergence of a post-treatment path may signal a treatment effect.

This argument is the heart of Differences in differences (DiD)

Differences in Differences



Differences in Differences

Consider the simplest case where we have two units and two periods.

We have some intervention D and our goal is to estimate the effect of treatment.

Unit	Outcome
A	Y = A + D
B	Y = B

Is it reasonable to just say that the difference as stated is causal?

Differences in Differences

Suppose we just compare a unit to itself

Unit	Outcome
Before	Y = A
After	Y = A + (Time + D)

This procedure eliminates the fixed effects of unit A but doesn't give an unbiased estimate of the treatment effect.

There are natural changes that are happening just due to time!

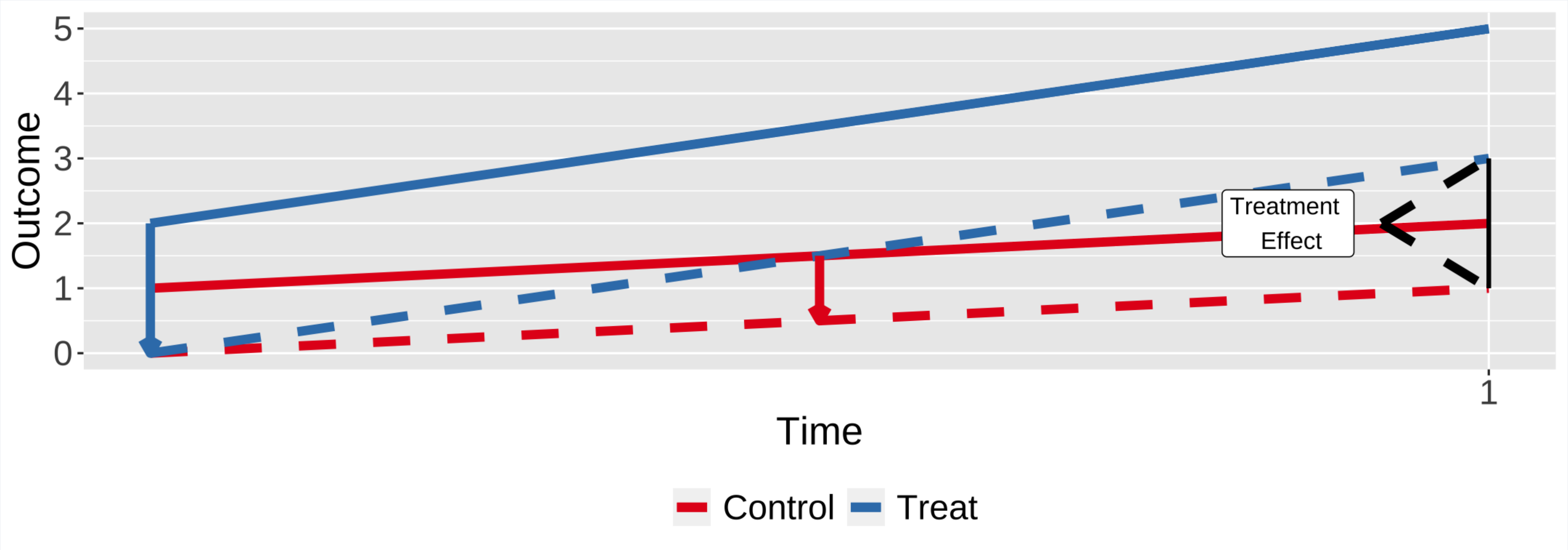
Differences in Differences

Let's combine the two simple strategies together

Unit	Time Period	Outcome	Diff 1
A	Before	Y = A	
A	After	Y = A + T + D	T+D
B	Before	Y = B	
B	After	Y = B + T	T

Once we difference out the before and after for each group, and then take the differences we get back an unbiased estimator of *D*

Differences in Differences



Parallel Trends

The parallel trends assumptions is the key assumption in any DiD design.

Informally, there is no time varying differences that exist between the two groups *other* than the treatment assignment.

This is fundamentally untestable!

DiD Decomposition

Our estimator δ is

$$\delta = (E[Y_t|Post] - E[Y_t|Pre]) - (E[Y_u|Post] - E[Y_u|Pre])$$

This equation yields the ATT for the treatment group.

DiD Decomposition

We can use a trick where we just add and subtract a term to this equation.

$$\delta = (E[Y_t(1)|Post] - E[Y_t(0)|Pre]) - (E[Y_u(0)|Post] - E[Y_u(0)|Pre]) + (E[Y_t(0)|Post] - E[Y_t(0)|Post]$$

Rearrange terms:

$$\delta = (E[Y_t(1)|Post] - E[Y_t(0)|Post]) + [(E[Y_u(0)|Post] - E[Y_u(0)|Pre]) + (E[Y_t(0)|Post] - E[Y_t(0)|Post]$$

This yields the ATT + SB for non-parallel trends!

Regression and DiD

We can estimate the DiD equation from the last slide with a linear regression.

$$Y_{it} = \alpha + \beta_1 D_i + \beta_2 POST_t + \delta(D_i * POST_t)$$

The advantage of this form is that we get estimate of our treatment effect and standard errors.

We can also easily add covariates to this model, which we practically will always want to do because we are in a selection on observables world

Assumptions of DiD

Parellel Trends: $E[Y0(2) - Y0(1) \mid A = 1] = E[Y0(2) - Y0(1) \mid A = 0]$

Consistency: The treatment status of a unit can vary over time but once a unit is treated we observes the potential outcome under treatment. Additionally future treatment does not affect past outcomes.

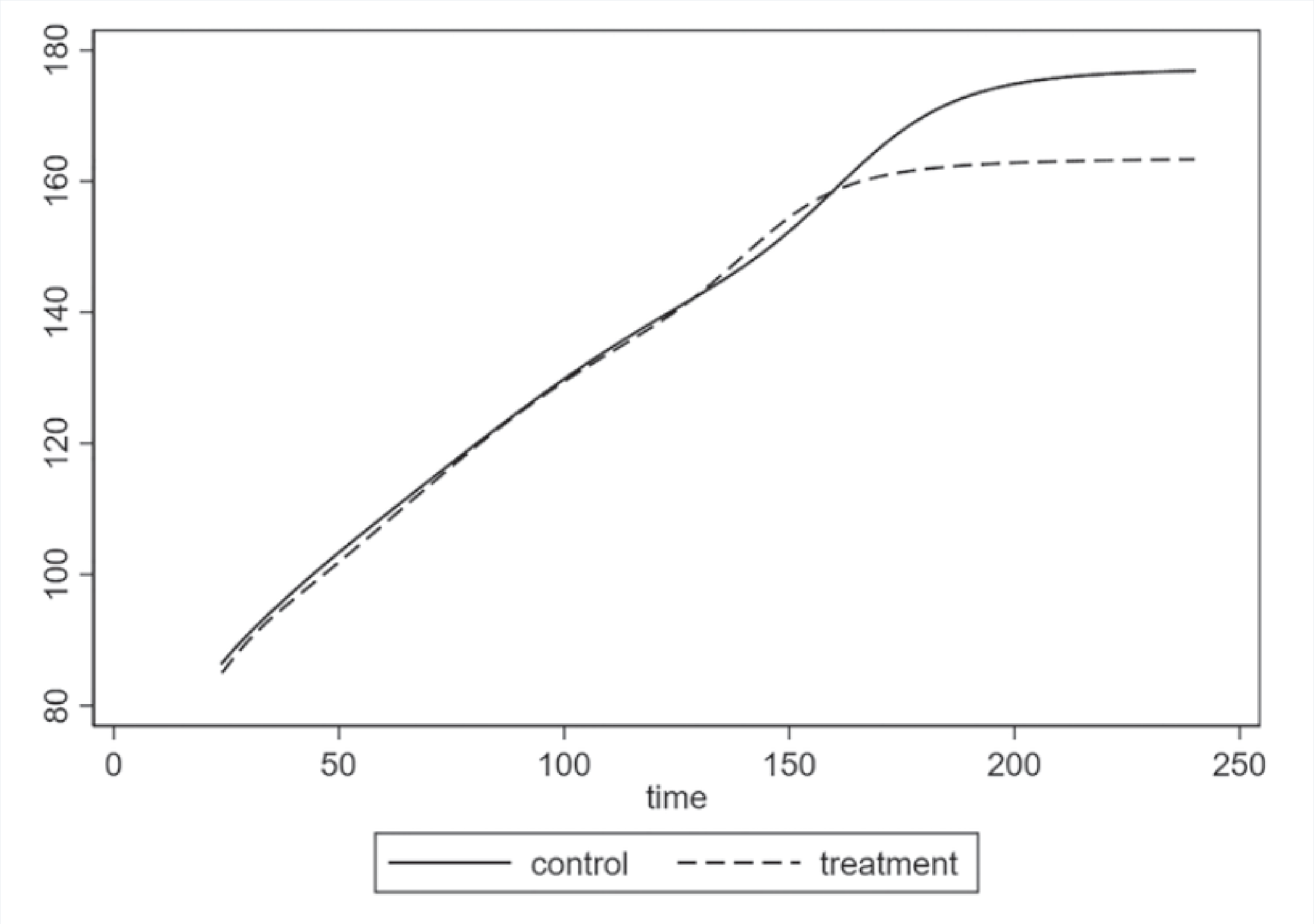
Positivity: $0 < P(A = 1|X) < 1$ for all X

How to assess parallel trends

"The squiggly line test" (McKenzie 2021)

Plot the raw treatment and control series and not just their differences. The longer and squigglier the pre-treatment trends, the more plausible parallel trends should be.

Informal Line test is not fool proof



Inference in DiD

In the two period case, standard errors work like we expect them to.

Often though we have lots of periods. e.g (Paglayan 2018)

Methods for getting appropriate errors (Bertrand, Duflo, Mullainathan 2004):

1) Block bootstrapping

2) Aggregation

3) Clustering