Week 6 TA Session

Anna-Elise Smith

Panel Data

- So far, we've focused on data across units, at one point in time
- With **panel data**, we'll be adding in observations at different points in time
- This is helpful because
- 1) we can let unit i serve as a 'control' for itself - i is more similar to i than any given j, after all! So, when we estimate τ^{ATE}:
- $\tau^{ATE} = E[Y_t(D_t = 1) - Y_t(D_t = 0)$
- Supposing that unit i gets treated in only period 2 (t = 1). Any time-invariant unobservable characteristics
- $U_i$of the individual get washed out:
  $\hat{\tau}^{TS} = Y_{t=1} - Y_{t=0} = \tau(D_{i,t=1} - D_{i,t=0}) + \beta(U_i - U_i) = \tau(D_{i,t=1} - D_{i,t=0})$
- Basically, we can get an unbiased estimator without matching nonsense, since i can *be* a control of itself.

Time Series

- Backing up one step, time series data is one unit at many points in time
- *So, I* yesterday can be a control for I today
- Assumption: *The counterfactual trend must be zero, or in english, we don't have* time-variant* characteristics affecting our treatment unit.
- Meaning, something can't effect the outcome variable differently through time, that is unaccounted for in the analysis.
- This is untestable. But, we can combine this time-varied control technique to get one of the 'bread and butters' of program evaulation.

DiD

- **When we have panel data, that is time series across many units we can do a difference in difference estimation**
- Let's combine this *with-unit, across time* comparison with an *across-unit, within time* comparison we used for our naive estimator. By doing this, we can estimate our *time-variant* characteristics might vary in our treatment groups. In other words, we can solve for the missing counterfactual (!!!). This is our **difference-in-difference** estimator.
- To present why this happens, consider that Y_{it} outcome may be broken down into a treatment effect, as well as *individual-variant* characteristics U_i and *time-variant* characteristics S_t. Observe the DiD estimator fix our problems:
- $Yit = \tau Dit + \beta Ui + \delta St \; Yjt = \beta Uj + \delta St$
- $\hat{\tau}^{DD} = (Yi,t=1 - Yi,t=0) - (Yj,t=1 - Yj,t=0)$
  $\hat{\tau}^{DD} = [\tau(Di,t=1 - Di,t=0) + \beta(Ui - Ui) + \delta(St=1 - St=0)] - [\beta(Uj - Uj) + \delta(St=1 - St=0)] \; \hat{\tau}^{DD} = \tau(Di,t=1 - Di,t=0)$

- Outcome = b + treat_i + post_t + ATE(treat*post)_it + u_it

- ATE =(Post_treated - Pre_treated) - (Post_untreated - Pre_untreated)

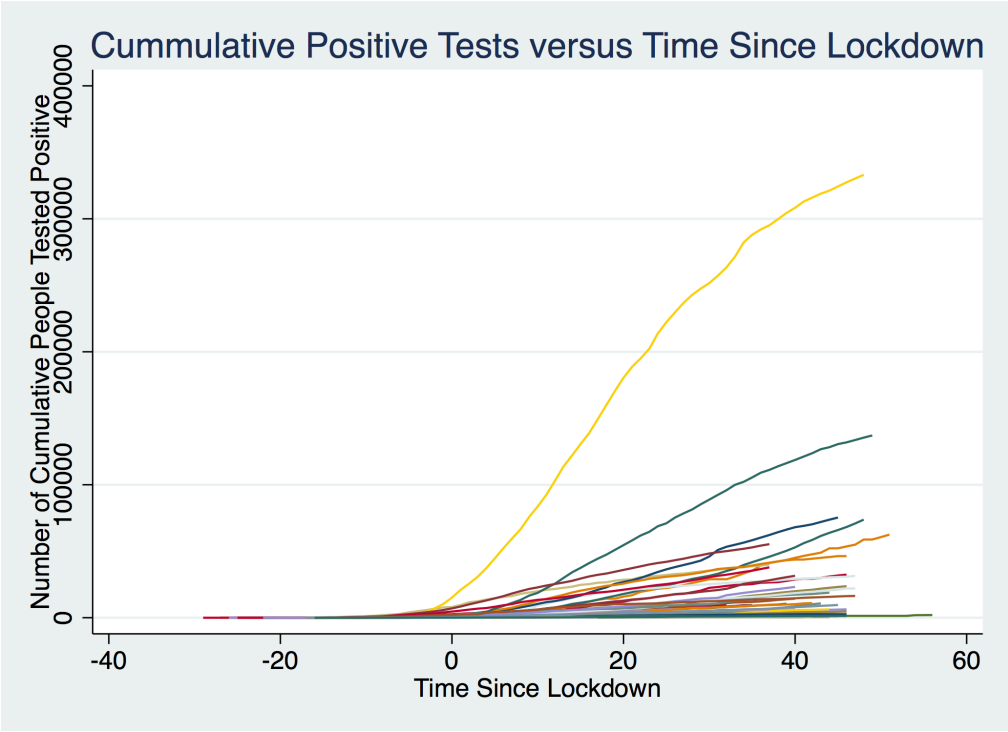
DiD Assumptions

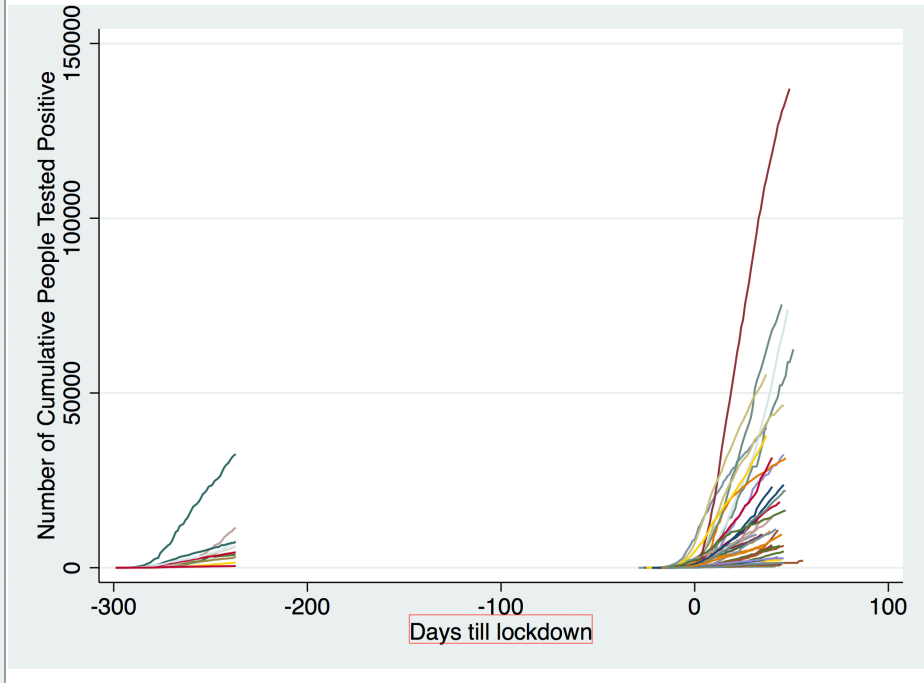
- Parallel trends: time- variant trends in treat and control must be similar, or else untreated Y is not a good control for treated Y
- We measure the difference in slope of the pre treatment to the post in the treated group to the pre treatment and the post of the control group (these are the two differences)
- This is superior to time series because we can test the assumption graphically

CoVid example

- Positive_it = a + State_i + Date_t + D(lockdown)_it + e_it
- Does common trends assumption hold?

Covid Example Cont.





Covid Cont

- Maybe states are on different trajectories
- Control for state*time effects!
- Found lockdown decreased positive tests on average by ~1,800/ state
- Issues here: common trends, small pre-treatment trends, does positive tests on a day tell you the infection rate?, lag in infections/tests, functional forms/exponential growth