Lecture 27

11-22-2021

Parallel Trends Redux

Recall that we our 2x2 DiD design is the ATT + selection bias.

Parallel trends allows us to eliminate the SB term

$$[E[Y_T^0|Post] - E[Y_T^0|Pre] - E[Y_U^0|Post] - E[Y_U^0|Pre]]$$

This requires us to evaluate a term with a counterfactual, which is why we cannot empirically test the parallel trends assumption

Parallel Trends Redux

Parallel trends is obviously violated if the treatment itself is endogenous.

Pretreatment parallel trends are not a reason to assume parallel trends in the future.

It may be helpful to interpret the common trends assumption as a byproduct of a set of underlying variables that differ across states and change over time.

Event Study Plots

While not a direct test of parallel trends, event study plots are often used to show comparability of groups in pre-treatment periods.

Plotting the raw data year by year runs into problems if the number of treatment groups is large or if units may become treated over time.

The current way that researchers evaluate pre-treatment dynamics in a situation with differential treatment timing is to estimate a regression model that includes treatment leads and lags.

$$Y_{its} = \alpha_s + \lambda_t + \sum_{tau=-q}^{-1} \alpha_\tau D_{s\tau} + \sum_{\tau=0}^{m} \delta_\tau D_{s\tau} + x_{ist} + \epsilon_{ist}$$

Event Study Plots

Event Study Plots

Placebo Tests with DiD

1. Balance tests are one form of a placebo test. We've just looked at them in the last slide.
2. Use data on an alternative type of outcome whose value would not be affected by treatment.
3. Consider different treatment timing and see whether it predicts outcomes.

Note that these are all similar to Jakiela's TWFE diagnostics

Estimating TWFE with staggered timing

With two periods and two treatment groups, this is just a basic application of a regression model.

With more than two periods and multiple treatment groups, things get tricky.

Our usual TWFE is a strange and unintuitive combination of weighted averages from units at different times (Goodman-Bacon 2020).

Gardner (2021) Estimator

One possibility is to recognize that under parallel trends, the group and time effects are identified from the subsample of untreated/not-yet-treated observations. Gardner uses this observation to suggest a 2S DiD estimator.

Stage 1: Estimate the model

$$Y_{ist} = \alpha_s + \lambda_t + \epsilon_{ist}$$

using the subsample of untreated/not-yet-treated observations and form adjusted outcomes $\hat{Y}_{ist} = Y_{ist} - \hat{\alpha}_s - \hat{\lambda}_t$

Stage 2: Regress the adjusted outcomes on treatment status D_{st}^k to estimate treatment effects.

The coefficient on this parameter in the second stage is the overall ATT or the average treatment effect for all the treated periods

Gardner (2021) Estimator

Why does this work?

In a dynamic setting (e.g. staggered treatment timing) the longer a group's treatment duration the more that group is weighted up in a TWFE estimator.

2SDiD removes the group and period fixed effects that create these weird weights.

Conceptually, this procedure is simply cleaning out the junk in the TWFE regression that lead it to estimate something other than the ATT.

Manual Code

```
### Run the regression on group and time fixed effects
### in pre-treatment period
s1 <- lm_robust(Y ~ state + time, data = data %>%
                  filter(treat == 0))

## Get transformed outcome
adjusted <- Y - predict(fs, data)

## Run the second stage
s2 <- lm_robust(adjusted ~ treat, data = data, clusters = cluster)
```

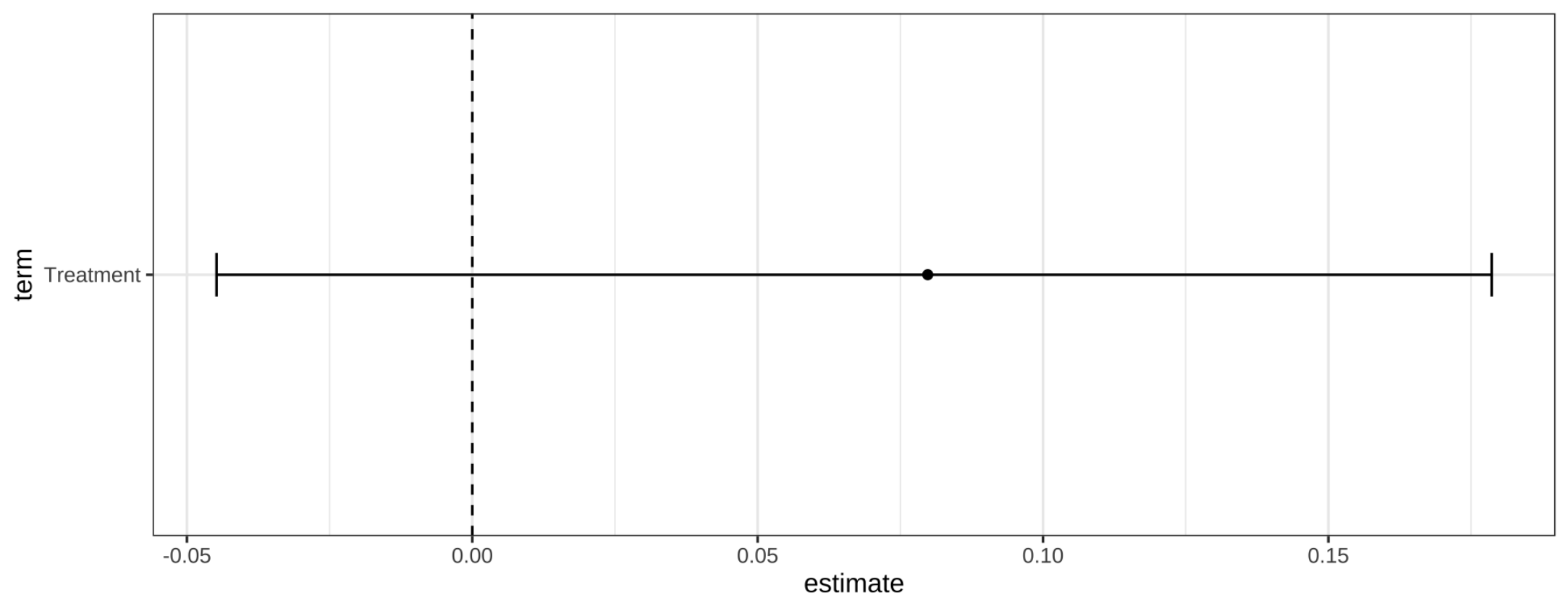
Application Cheng and Hoekstra (2013)

Cheng and Hoekstra are interested in whether "castle doctrine" laws deter crime or escalate violence.

Castle Doctrine laws are often referred to as "Stand Your Ground" laws.

The laws allows a person to use lethal force in places other than the home and removed any civil liability that comes from killing a person. They also gave the presumption of reasonable fear to the shooter.

Cheng and Hoekstra (2013)



Cheng and Hoekstra Event Study Plots