

Difference-in-Differences IV: Extensions

Lecture 6 - Introduction to Causal Inference

Kevin Li

Extensions to DiD

We have covered most of the basics of DiD. However, there are still some extensions to DiD that might deal with issues we face.

In this lecture, we will cover 5 different extensions to DiD:

1. Repeated Cross-Sections
2. Non-Absorbing Treatment
3. Continuous Treatments
4. Weakening Parallel Trends with IFect and MC
5. Sensitivity Analysis

Repeated Cross-Section

Stable Composition Group Assumption

Repeated Cross-Section in R

We can run a TWFE model (see lecture 2):

```
library(fixest)
feols(
  fml = Y ~ D + X1 + X2 | group + time
  data = df,
  vcov = ~group
)
```

For Callaway Sant'Anna Doubly-Robust (see lecture 3), we can implement by adding this argument to the `att_gt()` function:

```
panel = F
```

- This argument gets rid of the *idname* argument.

Non-Absorbing and Continuous Treatment

So far in DiD, we have assumed that once a unit gets treated, they remain treated. This is called **absorbing treatment**.

But sometimes, units will become untreated later. If this happens, we have **non-absorbing** treatment.

Only some estimators can deal with this:

- ▶ Imputation estimators (see last class)
- ▶ DIDmultiple (later)
- ▶ PanelMatch (we will not discuss here.).

Sometimes, we will also have treatment that has more than the classic treat/control dynamic (**continuous** treatments). DIDmultiple can also deal with this.

DIDmultiple (De Chaisemartin and D'Haultfœuille 2024)

DIDmultiple is an estimator that focuses on **switchers** - those units who change their treatment status between two time periods.

The estimator compares the change Δ in Y between switchers and non-switchers in that specific two-time period window.

$$\tau_t = \mathbb{E}[\Delta Y | \text{switchers}] - \mathbb{E}[\Delta Y | \text{non-switchers}]$$

These τ_t are the properly weighted together for a singular ATT or dynamic treatment effects.

The advantage of focusing on switchers: switchers can be generalised to **continuous** treatment variables D_{it} , making this estimator very versatile.

DIDmultiple in R

Parallel Trends Violations

We discussed in lecture 3, the importance of the parallel trends assumption. If it is violated, our DiD estimates can be wildly off.

Parallel trends is often not met - so in lecture 4, we discussed conditional parallel trends by conditioning for other variables X that might cause violations in parallel trends.

What if even with conditioning for parallel trends, we still cannot meet the parallel trends assumption?

- ▶ We will need strategies that can still estimate with violations of parallel trends.

One way to do this is to estimate the violations of parallel trends, and then control for them.

Interactive Fixed Effects (IFect)

Liu, Xu, and Wang propose a modified imputation estimator:

$$Y_{it}(0) = \underbrace{\hat{\alpha}_i + \hat{\gamma}_t + \mathbf{X}_{it}^\top \hat{\boldsymbol{\beta}}}_{\text{TWFE imputation}} + \hat{\boldsymbol{\lambda}}_i^\top \hat{\boldsymbol{\xi}}_t$$

- ▶ $\hat{\boldsymbol{\xi}}_t$ is a vector of time-varying latent (unobserved) variables that IFect estimates.
- ▶ $\hat{\boldsymbol{\lambda}}_i$ is a vector of how each unit i is affected by each of the time-varying latent variables.

Summary: This additional $\hat{\boldsymbol{\lambda}}_i^\top \hat{\boldsymbol{\xi}}_t$ allows IFect to account for some differential trends between units (parallel trends violations). We still need parallel trends for IFect - but the additional part makes it more likely we meet parallel trends.

Matrix Completion (MC)

Athey et al (2021) propose another extension on imputation estimators:

$$\mathbf{Y}(\mathbf{0}) = \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\mathbf{L}}$$

The way this estimator works is kind of complicated - and uses computer science/machine learning principles.

But it can account for some differential trends between units (parallel trends violations).

- ▶ We still need parallel trends for MC (like IFECT) - but the additional part makes it more likely we meet parallel trends.

IFEct and MC in R

```
library(fect)

model <- fect(
  method   = "ife", # change to mc or ife
  formula  = Y ~ D + X1 + X2,
  data     = df,
  index    = c("unit", "time"),
  CV       = T, # don't change
  se       = T,
  seed     = 2342, # any number is ok
)

print(model) # ATT
plot(model) # dynamic
```

Sensitivity Analysis

We can also take an alternative approach to potential violations to parallel trends.

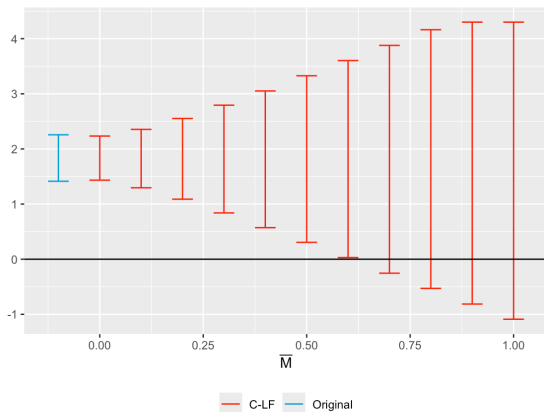
Instead of re-estimating with a new model, why not consider how a parallel trends violation would affect our causal estimates with our original model?

Sensitivity analysis considers different magnitudes of parallel trends violations (how large the differential trend is).

Then, it creates confidence intervals for our τ_{ATT} based on each magnitude of parallel trends violation. We can then see how “robust” our significant estimates are to parallel trends violations.

- ▶ If even with a potentially large parallel trends violation, our causal effect remains significant, we can be confident that our significant causal effect is true.

Sensitivity Analysis (Cont.)



\overline{M} indicates the magnitude of parallel trend violations. For each magnitude, we have a confidence interval for the $\hat{\tau}_{ATT}$. The causal effect becomes insignificant around $\overline{M} > 0.7$.