

# 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

So far, we have used panel data - when we observe the same units  $i$  over multiple time periods  $t$ .

Sometimes, we will have observations over many time periods  $t$ , but not of the same units.

For example, let's say we are using a survey of Americans, that happens each year. Each year  $t$  will have a different sample of individuals. This is repeated cross-section.

We can use DID in repeated cross-section, if units are sampled from some control group and treatment group (like let's say state gets a policy - sample from individuals within the treated state, and the control state).

# Stable Composition Group Assumption

**Issue:** If our sample from different time periods  $t$  differs on key confounders for outcome  $Y$ , then we will not know if our causal estimates are accurate, since the changes in confounders may cause changes in  $Y$ .

**Solution:** we must ensure that in terms of key confounders for  $Y$ , the sampled individuals from each time period  $t$  is consistent.

- ▶ If it is not consistent, we can control for observable confounders in our regression, just like we would for conditional parallel trends.

Note: this is a relatively new area of research - very few papers combine conditional parallel trends with controls for stable group composition.

## 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, extension of DIDmultiple).

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  $\Delta$  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  $\tau_t$  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.

## 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:

$$\hat{Y}_{it}(0) = \hat{\alpha}_i + \hat{\delta}_t + \hat{\lambda}_i^\top \hat{\mathbf{f}}_t$$

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

**Summary:** This additional  $\hat{\lambda}_i^\top \hat{\mathbf{f}}_t$  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. Imagine a matrix  $\mathbf{Y}(\mathbf{0})$  of potential outcomes, where  $Y_{it}(\mathbf{0})$  is in row  $i$  and column  $t$ . This matrix  $\mathbf{Y}(\mathbf{0})$  has missing values for the units that are treated (that we need to impute to estimate causal effects).

Idea: assume there are patterns in  $\mathbf{Y}(\mathbf{0})$ , or in other words, assume it is a low-rank matrix. Low rank matrix can be estimated as product of two matrices  $\mathbf{Y}(\mathbf{0}) = \mathbf{\Lambda}\mathbf{F}$ , where  $\mathbf{F}$  is a matrix of multiple  $\mathbf{f}_t$  from IFect (last slide), and  $\mathbf{\Lambda}$  is a matrix of  $\lambda_i$ .

Using iterative algorithm for matrix completion methods, and regularisation to prevent overfitting, we can estimate missing entries in the matrix.

# Sensitivity Analysis

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

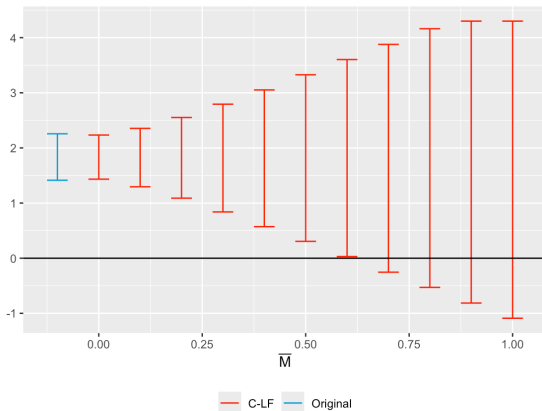
Instead of re-estimating with a new model, why not consider how a parallel trends violation would affect our causal estimates with our original model?

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

Then, it creates confidence intervals for our  $\tau_{ATT}$  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.)



$\bar{M}$  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  $\bar{M} > 0.7$ .