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 4 different extensions to DiD:

1. Non-Absorbing Treatment.
2. Weakening Parallel Trends
3. Abandoning Parallel Trends (Synthetic Controls)
4. Sensitivity Analysis

Non-Absorbing 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. For example, maybe a State repeals a law. If this happens, we have **non-absorbing** treatment.

Only some estimators can deal with this:

- Imputation estimators (see last class)
- DIDmultiple (see last class)
- PanelMatch (similar to DIDmultiple, we will not disucss this).

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.

We will first discuss weakening the parallel trends assumption, before completely dropping it.

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.

Synthetic Controls Method

Both IFEct and MC, while they can account for some violations in parallel trends, they still require that after accounting for those violations, parallel trends is met.

What if we want to abandon the entire parallel trends assumption? The solution is **synthetic controls**.

- Setup: We have one treated unit, and a lot of untreated units with tons of pre-treatment data.
- Idea: Let us use the untreated units to create a plausible counterfactual for the treated unit.
- Execution: Find a weighted average of untreated units that approximate the pre-treatment Y values of the treated unit. Then, use the same weighted average of untreated units to estimate the counterfactual in post-treatment periods.

Generalised Synthetic Controls

Xu (2017) proposes to sort of combine DiD and synthetic controls into Generalised synthetic controls.

- Instead of a weighted average of untreated units, let us create a regression model to combine untreated units to approximate the counterfactual.
- This is more flexible than a weighted average, so we can create a more plausible counterfactual.

Idea: estimate a IFEct model with only the untreated units (do not include the pre-treatment period of the treated units).

- Also can accommodate more than just 1 treated unit, when synthetic controls only accounts for one treated unit.

Sensitivity Analysis

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

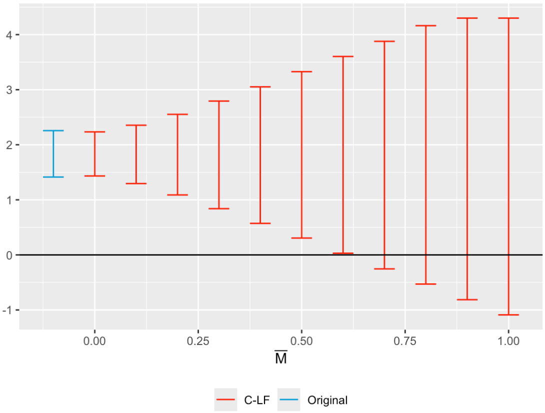
Instead of re-estimating with a new model, why not consider how a parallel trends violation would affect our causal estimates with our original model?

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 $\hat{\tau}_{\text{ATT}}$. The causal effect becomes insignificant around $\overline{M} > 0.7$.

Other Extensions

There has been increasing work on a few extensions in DiD:

1. Continuous treatment variables. DIDmultiple (last lecture) can already accommodate continuous treatment. Callaway et al (2024) also have a new estimator for continuous treatment.
2. **Triple Differences (DDD)**: basically difference-in-differences but with an additional difference. Has been used for a while to adjust for concerns over confounders.
3. Other areas of research include IV-DiD, Spatial DiD, Quantile DiD, and Changes-in-Changes (CiC).