How much should we trust differences-in-differences estimates?

Carlos Lezama
Advanced Microeconometrics
ITAM

Spring 2022

Introduction

Introduction

Differences-in-Differences (DD) estimation consists of identifying a specific intervention or treatment. One then compares the difference in outcomes after and before the intervention for groups affected by the intervention to the same difference for unaffected groups.

The great appeal of DD estimation comes from its simplicity as well as its potential to circumvent many of the endogeneity problems that typically arise when making comparisons between heterogeneous individuals.

Obviously, DD estimation also has its limitations. It is appropriate when the interventions are as good as random, conditional on time and group fixed effects. Therefore, much of the debate around the validity of a DD estimate typically revolves around the possible endogeneity of the interventions themselves.

Introduction

Focus on issues relating to the *standard error* of the estimate.

DD estimates and their standard errors most often derive from using Ordinary Least Squares (OLS) in repeated cross sections (or a panel) of data on individuals in treatment and control groups for several years before and after a specific intervention.

One then typically estimates the following regression using OLS:

$$Y_{ist} = A_s + B_t + cX_{ist} + \beta I_{st} + \varepsilon_{ist},$$

where the subscripts i, s, and t stand for the individual, group (such as a state), and time (such as a year), respectively. Furthermore, let Y be our outcome of interest, I be a dummy, A and B be fixed effects, X be relevant individual controls, and ε be an error term.

Introduction

The estimated impact of the intervention is then the OLS estimate $\hat{\beta}$. Standard errors used to form confidence interval for $\hat{\beta}$ are usually OLS standard errors.

Remark

Note that this is valid only under the very restrictive assumption that changes in the outcome variable over time would have been exactly the same in both treatment and control groups in the absence of the intervention.

Introduction

Three factors make serial correlation an especially important issue in the DD context:

1. DD estimation usually relies on fairly long time series.
2. The most commonly used dependent variables in DD estimation are typically highly positively serially correlated.
3. The treatment variable I_{st} changes itself very little within a state over time.

These three factors reinforce each other so that the standard error for $\hat{\beta}$ could severely understate the standard deviation of $\hat{\beta}$.

A Survey of DD Papers

A Survey of DD Papers

Text.

Overrejection in DD Estimation

Overrejection in DD Estimation

Text.

Solutions

Parametric Methods

Text.

Block Bootstrap

Text.

Ignoring Time Series Information

Text.

Empirical Variance-Covariance Matrix

Text.

Arbitrary Variance-Covariance Matrix

Text.

Summary

Text.

Conclusion

Conclusion

Text.