Overview of Difference-in-Differences

Kevin Lingfeng Li

Table of contents

1 Classical DID

1.1 Intuition of Identification

We have a treatment group g = 1, and a control group g = 0. We have two time periods t = 0 and t = 1.

- In time period t = 0, both treatment group g = 1 and control group g = 0 both do not receive treatment.
- In time period t = 1, only treatment group g = 1 receives treatment. Control group g = 0 is still untreated.

To estimate the average causal effect of receiving treatment, it might seem natural to compare the treatment group's average outcome after and before receiving treatment.

$$\tau_{\text{naive}} = \overline{\text{Y}}_{\text{g=1,t=1}} - \overline{\text{Y}}_{\text{g=1,t=0}} \tag{1}$$

Issue: what if in between time periods t = 0 and t =1, something happened that affected the average outcome of Y (some trend in Y without treatment)? Our estimates will be biased because the difference contains both the causal effect, and the change in the trend of Y.

Solution: Use an control group to estimate the trend in Y had no treatment occurred.

$$\text{trend in Y with no treatment} = \overline{\text{Y}}_{\text{g=0,t=1}} - \overline{\text{Y}}_{\text{g=0,t=0}}$$

Then, subtract this trend from the difference to obtain an unbiased estimate.

$$\tau_{\text{DID}} = \underbrace{\left(\overline{\text{Y}}_{\text{g=1,t=1}} - \overline{\text{Y}}_{\text{g=1,t=0}}\right)}_{\text{change in treatment group}} - \underbrace{\left(\overline{\text{Y}}_{\text{g=0,t=1}} - \overline{\text{Y}}_{\text{g=0,t=0}}\right)}_{\text{trend in control group}} \tag{2}$$

1.2 Regression Estimator

We have established that the difference-in-differences estimate is

$$\tau_{\text{DID}} = \underbrace{\left(\overline{\text{Y}}_{\text{g=1,t=1}} - \overline{\text{Y}}_{\text{g=1,t=0}}\right)}_{\text{change in treatment group}} - \underbrace{\left(\overline{\text{Y}}_{\text{g=0,t=1}} - \overline{\text{Y}}_{\text{g=0,t=0}}\right)}_{\text{trend in control group}} \tag{3}$$

We can estimate the same effect more simply (with standard errors) with a simple regression model:

$$\text{Y}_{\text{it}} = \alpha + \gamma \text{G}_{\text{i}} + \delta \text{T}_{\text{t}} + \tau \text{G}_{\text{i}} \text{T}_{\text{t}} + \varepsilon_{\text{it}} \tag{4}$$

Where $\text{G}_{\text{i}} \in \{0, 1\}$ represented which group (control/treatment) unit i is in, and $\text{T}_{\text{t}} \in \{0, 1\}$ representing which time period t the observation is in.

Proof of equivalency:

Using conditional expectation definition of regression, we can see:

$$\begin{aligned}
\mathbb{E}[\text{Y}_{\text{it}} | \text{G}_{\text{i}} = 1, \text{T}_{\text{t}} = 1] &= \alpha + \gamma + \delta + \tau \\
\mathbb{E}[\text{Y}_{\text{it}} | \text{G}_{\text{i}} = 1, \text{T}_{\text{t}} = 0] &= \alpha + \gamma \\
\mathbb{E}[\text{Y}_{\text{it}} | \text{G}_{\text{i}} = 0, \text{T}_{\text{t}} = 1] &= \alpha + \delta \\
\mathbb{E}[\text{Y}_{\text{it}} | \text{G}_{\text{i}} = 0, \text{T}_{\text{t}} = 0] &= \alpha
\end{aligned} \tag{5}$$

Plugging values from Equation 5 into Equation 3, we get:

$$\begin{aligned}
\tau_{\text{DID}} &= \left(\overline{\text{Y}}_{\text{g=1,t=1}} - \overline{\text{Y}}_{\text{g=1,t=0}}\right) - \left(\overline{\text{Y}}_{\text{g=0,t=1}} - \overline{\text{Y}}_{\text{g=0,t=0}}\right) \\
&= \left([\alpha + \gamma + \delta + \tau] - [\alpha + \gamma]\right) - \left([\alpha + \delta] - [\alpha]\right) \\
&= (\delta + \tau) - (\delta) \\
&= \tau
\end{aligned} \tag{6}$$

Use standard errors clustered by group for accurate statistical inference.

1.3 Identification Assumptions

The whole idea of difference-in-differences is to use the control group trend to adjust for a possible trend in the treatment group.

But for this to be accurate, we are assuming that the trend in outcome in the control group is equal to the trend in the treatment group had no treatment occured. This is called the **parallel trends** assumption.

> Had the treatment group not received treatment, it would have followed the same average trend in outcome as the control group.

If parallel trends is not met, we can use **conditional parallel trends**: where the parallel trends assumption is met subject to conditioning on a set of covariates. This set of covariates X might be correlated with Y, and the trends of X in treatment vs. control groups may differ - thus controlling/conditioning for X solves the issue of parallel trends.

Another assumption is **no anticipation**: that the individuals in the treatment group are not anticipating treatment and so respond in t = 0. This can be considered a measurement problem: event though the treatment officially begins in t = 1, its effects are being felt in t = 0, so the actual official treatment begin date is mis-measuring the true treatment begin date.

Finally, the **stable unit treatment value assumption (SUTVA)** states that if one unit i is treated, that does not affect the potential outcomes of another unit j. Or in other words, ones outcomes under treatment and control are only caused by their own treatment status, not other individual's treatment status.

For repeated cross-section designs where the sample of individuals is different between time periods t = 0 and t = 1, we also require **stable group composition**. This means that for key covariates that affect Y, the average values in each sample for each time period must be equivalent. If they are not equivalent, it is possible differences in samples are explaining our effect, and not the treatment.

2 Two-Way Fixed Effects

2.1 Introduction

Recall the regression estimator for two-period estimates from Equation 4:

$$Y_{it} = \alpha + \gamma G_i + \delta T_t + \tau G_i T_t + \varepsilon_{it} \tag{7}$$

We can rewrite this regression equation using a fixed effect approach:

$$Y_{itg} = \alpha_g + \delta_t + \tau D_{it} + \varepsilon_{it} \tag{8}$$

Where $D_{it} \in \{0, 1\} = G_i T_t$ is a treatment variable that indicates if unit i is treated at time period t. α_g is a group-level fixed effect, and δ_t is a time-fixed effect.

Recall regression is only unbiased if exogeneity is met, or in other words, no omitted variables/confounders. Group fixed effects α_g accounts for between-group differences. Time fixed effects δ_t accounts for between-time differences. The only remaining potential confounders are differential trends, which are assumed to not be present based on parallel trends.

If we have panel data, we can further-ensure exogeneity is met by using unit, rather than group fixed effects. With unit fixed effects, we are not only accounting for between-group differences, but also between-unit differences within and between groups:

$$Y_{it} = \alpha_i + \delta_t + \tau D_{it} + \varepsilon_{it} \tag{9}$$

This is the standard TWFE estimator for panel data. With all regression estimators, standard errors should be clustered by unit.

We can also add covariates $\mathbf{X}_{it}$ to condition for parallel trends in our model:

$$Y_{it} = \alpha_i + \delta_t + \tau D_{it} + \mathbf{X}_{it}^{\top}\boldsymbol{\beta} + \varepsilon_{it} \tag{10}$$

2.2 Generalised DID

So far, we have considered DID with only two time periods, t = 0 before treatment, and t = 1 after treatment. But we do not need this limitation. We can have many pre-treatment periods, and many post-treatment periods.

We still use TWFE for causal effects, since if the parallel trends assumption is met, and all units in the treatment group begin treatment at the same time, TWFE is still exogenous.

Since we have multiple time periods, we might be interested in not just an overall treatment effect τ, but also treatment effects for each post-treatment period t = 1, t =2, etc. This is called an **event-study** or **dynamic treatment effects**.

To estimate these effects, we define a new variable called relative time R. This variable has a value of 0 for the first treatment period, positive values for subsequent post-treatment periods, and negative values for pre-treatment periods.

- Ex. If treatment begins in 2007, 2007 would get a value R = 0 since it is the initial treatment period. 2005 would be R = -2, 2009 would be R = 2.

Using the relative time variable, we can use the TWFE estimator to estimate an event study (you can include covariates as well to condition for parallel trends):

$$Y_{itr} = \alpha_i + \delta_t + \sum_{r \neq -1} \tau_r \cdot 1\{R_{itr} = r\} + \varepsilon_{it} \tag{11}$$

$1\{R_{itr} = r\}$ is an indicator function that takes value 1 if the observation itr has a relative time R value of r. The $\sum$ basically tells us to calculate an effect τ for every relative time period r expect for r = -1 (our reference category).

The τ_r estimates post-treatment (so $R \geq 0$) are dynamic treatment effects. We can plot τ_r to see how treatment effects change over time-periods.

The τ_r estimates pre-treatment (so $R \leq -2$ since R = -1 is always 0) are pre-treatment estimates. They are not causal estimates since this is before treatment. They can be used to check for parallel trends - if they are insignificant, parallel trends is likely to be met. We generally want to see a few pre-treatment periods in a row right before treatment (so r = -2, r = -3, r = -4) to be insignificant if we want evidence for parallel trends.

2.3 TWFE Bias in Staggered Treatment

So far, we have assumed all units in the treatment group begin treatment at the same time. But frequently, this is not the case. For example, many states begin a programme at different times (staggered rollout), such as different states adopting the Common Core curriculum in different years.

Issue: TWFE is biased in

3 Matching and Reweigthing

4 Imputation