Lecture 6 – Differences in Differences

Intuition

The basic intuition for Differences in Differences is that there are two groups are similar. Importantly, they would follow the same trend absent treatment. Then, treatment occurs and one of the groups sees a jump relative to the trend.

Two by Two

In MM, they use the example of banks in the 6th federal reserve district versus the 8th. The chair of these districts reacted differently to the beginning of the great recession. The 6th district lent money to the struggling banks. The table below shows the number of banks open in the different districts in each year.

	'30	'31
District 6	135	121
District 8	165	132

One estimate would be to just compare the number of banks from 1931 to 1930 or 135-121=-14. What would be the problem with this comparison?

Other things changed from 1930 to 1931, namely a global economic crisis, and so the change in banks is due to lending to banks and other things.

The with differences in differences is to find a comparison group that was not treated. If we assume that the control group acts as a good counterfactual we can get an estimate. We call this the parallel trends assumption. That is, the treatment group would have followed the same trend as the control group if there had not been treatment.

The way we implement this is to see what happened in the 8th district 132-165=-33.

We get our total estimate of the effect of lending to banks by taking the difference of these two differences (hence the name). -14-(-33)=19.

More than 1 period

What does this look like for more than one period? See MM Figures 5.1, 5.2, and 5.3

Fixed Effects

We give a special name to certain kinds of controls in regression, fixed effects. Fixed effects can be thought of as a series of indicator variables for the different categories of a variable. For instance, sex fixed effects would be an indicator for male and female.

Let's say you have m categories, you only include m-1 categories in the regression? Why? Including the last category doesn't include any new information–you can infer that it is in the last category because it was 0 in the other m-1 categories.

So let's say you want to account for sex in your regression, you would include an indicator for male (or female, but not both).

Let's say you want to control for state characteristics (e.g. differences in governance, economy, demographics, etc.) Your data set has a variable for state and it goes from 1-50. You want to turn this into fixed effects. Otherwise you are saying that going from state 1 to state 2 has the same effect as state 45 to 46. That is very unlikely to be true since the order doesn't mean anything.

Fixed effects control for a lot of things–more than you'd think at first! They control for all things that are fixed within that group. Take states, state fixed effects control for all characteristics of the state that do not change. For example, they control for the political party of the governor in 2020, for the deaths from Covid in 2023, etc. *Anything* fixed is controlled for.

Some things are not fixed, for example, the unemployment rate in the state.

You also need at least two observations from that group to control for fixed effects from that group. I have two kids and so if you had both of them in your data set, you could control for Denning family fixed effects and control for all fixed characteristics of our family.

Could you control for where the father got his bachelor's degree and Denning family fixed effects? No! That is fixed within the Denning family.

Regression Version

We can implement the 2x2 DD using regression. In the above example on banks it would be

$$Y_{it} = \beta_0 + \beta_1 TREAT_i + \beta_2 POST_t + \beta_3 TREAT_i * POST + \varepsilon_{it} \tag{1}$$

$TREAT$ is an indicator for the treatment group which would be district 6. $POST$ is an indicator for being after treatment. The coefficient β_3 represents the treatment effect.

You can recover the four means in the above table using these estimates. (How?)

Advantages of using regression–you can control for other important covariates that might affect the outcome (similar to controlling for covariates in an RCT). In the most basic set up you don't *need* to do this unless you want to assume that trends would be parallel only after accounting for differences in covariates.

How strong is parallel trends? Pretty strong–you don't know what the trend would be without treatment because it is never observed.

What supporting evidence can you provide? Trends actually did follow the same path before treatment. Nothing else changed afterwards (showing this is true for covariates is easy, harder to do for unobservables)

However, let's suppose that in addition to lending to banks, the federal reserve also changed reserve requirements in the 6th district. Would you be able to interpret the coefficient as the effect of lending to banks? (No) One way this is often described is that there were no other contemporaneous shocks.

Another thing to worry about is if the trends would change absent the treatment. For instance, there is a shock to the control group. You need to think about something like an exclusion restriction here too–the only thing changing in treated groups is treatment

Another thing to worry about is if treatment is endogenous. That is, treatment occurs in places that have some sort of underlying, unobserved difference. For instance, if you have treatment that is voted on–the places that pass the treatment and the places that don't might be on different trends.

Multiple Periods, Multiple Treatment Groups

The basic idea can be extended beyond the two groups, treatment and control. Equation 1 above is all you need. When you do this, you lose the simple intuition of the 2 group case, but you gain more "experiments."

You swap out the $POST$ indicator in equation 1 with fixed effects for years, η_t. You also swap out the $TREAT$ indicator with indicators for each group (for instance state), γ_i. You still keep $TREAT * POST$ which is still the treatment effect.

$$Y_{it} = \beta_0 + \gamma_i + \eta_t + \beta_3 TREAT_i * POST_{it} + \varepsilon_{it} \tag{2}$$

The identifying assumption can be stated as $E[\varepsilon_{it}|\gamma_i, \eta_t] = 0$. That is, the error term is expected to be 0. A violation of this would occur is some other policy changed that affects the outcome at the same time. Also violations of parallel trends would have a non-zero mean expectation.

What are these fixed effects doing? The group fixed effect is accounting for level differences in each of the groups. For instance, one state may have more banks than another. The time fixed effect accounts for any shared trends in the outcome.

You can also relax the parallel trends assumption a bit by including linear time trends for groups, $t * \gamma_i$. However, this is kind of weird (see MM for details)

Event study graphs are a good way to check for parallel trends in the pre period. This happens by estimating the following equation:

$$Y_{it} = \beta_0 + \gamma_i + \eta_t + \sum_{s=-T, s\neq -1}^{T} \beta_3^t TREAT_i * 1(t = s) + \varepsilon_{it} \tag{3}$$

The coefficients β_3^t show the difference between treatment and control in the pre period and the post period. In the pre-period, you would want these to be flat and indistinguishable from the year prior to treatment (which is omitted, $t = -1$)

Starting with "Difference-in-differences with variation in treatment timing" by Andrew Goodman-Bacon in 2021 people realized that the two way fixed effects estimators of DD had some serious flaws.

If treatment timing was staggered and effects were heterogeneous and/or grew over time, the two way fixed effects estimators will give you the wrong answer.

Thankfully, there are several solutions. The most popular is described in "Difference-in-Differences with multiple time periods" 2021 by Callaway and Sant'Anna