

Difference-in-Differences

i. Motivation

We have discussed how unit fixed effects in panel data remove any possible omitted variable bias from time-invariant variables. But what about other possible sources of OVB? If unit characteristics change over time in a way that is correlated with both the independent and outcome variables, this could cause bias. This could be solved if we randomize the allocation of the independent variable (the treatment), but what if we aren't able to randomize?

In section 11 we talked about regression discontinuity as one method to recover causal impacts without randomization. To run a regression discontinuity design, we need a clear threshold for treatment assignment along a continuous running variable. Of course we will not always have this, and hence will not be able to run an RD for many treatments of interest. But if we have access to panel data (observations of the same units over time) where some units received a treatment in between periods and others did not, **Difference-in-Differences (DD)** is another way of getting around a non-random assignment of a treatment. In this method, we exploit the timing of treatment to look at how our outcome of interest *changes* between the treatment and control groups before and after the treatment is applied (over time).

ii. Causal Effect

Pischke and Angrist: "sometimes treatment and control outcomes move in parallel in the absence of treatment. When they do, the divergence of a post-treatment path from the trend established by a comparison group may signal a treatment effect."

Implementation of the method is as follows:

1. Compute the difference in the outcome variable Y after (period 1) and before (period 0) for the control group (C):

$$\bar{Y}_{C1} - \bar{Y}_{C0} = \Delta \bar{Y}_C$$

2. Compute the difference in the outcome variable Y after (period 1) and before (period 0) for the treatment group (T):

$$\bar{Y}_{T1} - \bar{Y}_{T0} = \Delta \bar{Y}_T$$

3. The impact of the program is measured by the difference in differences as:

$$(\bar{Y}_{T1} - \bar{Y}_{T0}) - (\bar{Y}_{C1} - \bar{Y}_{C0}) = (\Delta \bar{Y}_T - \Delta \bar{Y}_C)$$

In a regression framework, we estimate the following:

$$Y_{it} = \beta_0 + \beta_1 Post_t + \beta_2 Treat_i + \beta_3 Post_t \times Treat_i + u_{it}$$

The regression framework has the added benefit of providing us with standard errors and t-statistics/p-values so we can test for significance of our estimators. We call β_3 the difference-in-difference estimator as it illustrates the differential impact of being in the post period (after treatment takes effect) if you're treated (relative to control).

$$\begin{aligned} E[Y_{T1}] - E[Y_{T0}] &= (\beta_0 + \beta_1 + \beta_2 + \beta_3) - (\beta_0 + \beta_2) = \beta_1 + \beta_3 \\ E[Y_{C1}] - E[Y_{C0}] &= (\beta_0 + \beta_1) - (\beta_0) = \beta_1 \end{aligned}$$

Then

$$(E[Y_{T1}] - E[Y_{T0}]) - (E[C_{T1}] - E[C_{T0}]) = \beta_1 + \beta_3 - \beta_1 = \beta_3$$

In this specification, β_0 gives you the baseline (pre-treatment) mean of Y in the Control group. β_2 is the difference in the baseline mean for the Treatment group, relative to Control. β_1 gives you the average change in Y from the pre- to the post-treatment period for the Control group. This captures “secular change” that we allow to occur in the outcome due to factors other than the treatment. A key assumption is that this change is the same for everyone. β_3 is the difference in change over time in Y for the Treatment group relative to the control.

See subsection v below for a graph showing these parameters.

iii. Key Assumption

The key assumption for the validity of the method: the difference between before and after the treatment in the comparison group is a good counterfactual for the treatment group. In other words, the trend in outcomes of the comparison group over time is what we would have observed in the treatment group absent the policy/intervention/reform. Here we allow for the possibility that the outcome variable could be changing over time, even absent the treatment. What we assume is that the outcome variable would be changing in the same way for both the treatment and control groups. We call this the “**parallel trends**” assumption.

In short, the only thing that can change over time between the treatment and control groups is exposure to the treatment. We therefore have a similar assumption for recovering a causal estimate as with fixed effects regression: changes in u_i over time cannot be correlated with changes in x_i - in particular, they cannot be correlated with the treatment variable.

iv. Tests for Validity of Assumption

How can we provide evidence that our control group is a good counterfactual for the treatment group? If we have data on multiple periods before the treatment, we can verify that before the treatment, the Control and Treatment groups had the same trend (“parallel trends”) for the outcome variable.

Implementation of the method is as follows:

1. Compute the difference in the outcome variable Y before the treatment (pre- period) and one period even before that (pre-pre period if you will) for the control group (C):

$$\bar{Y}_{C_{pre}} - \bar{Y}_{C_{pre-pre}} = \Delta \bar{Y}_C$$

2. Compute the difference in the outcome variable Y before the treatment pre-period and pre-pre period for the treatment group (T):

$$\bar{Y}_{T_{pre}} - \bar{Y}_{T_{pre-pre}} = \Delta \bar{Y}_T$$

3. The test for parallel trends is:

$$(\bar{Y}_{T_{pre}} - \bar{Y}_{T_{pre-pre}}) - (\bar{Y}_{C_{pre}} - \bar{Y}_{C_{pre-pre}}) = (\Delta \bar{Y}_T - \Delta \bar{Y}_C)_{\text{in the period before treatment}}$$

In a regression framework, we can run the following estimation (using data only from BEFORE treatment):

$$Y_{it} = \beta_0 + \beta_1 Pre_t + \beta_2 Treat_i + \beta_3 Pre_t \times Treat_i + u_{it}$$

Where for example Pre is a dummy variable equal to 1 if we are in the pre-period, and 0 if we are in the pre-pre period. Now β_3 the differential impact of being in the pre period (relative to the pre-pre period) if you're treated. We expect the coefficient on this term to be statistically insignificant: in the periods leading up to treatment the trends in the control group should be the same as the trends in the treatment group.

We can extend this comparison for multiple pre-periods. We will also generally plot the means by time period for treatment and control groups, to check visually whether they seem parallel before the treatment and see how they change after treatment.

Thinking back to our fixed effects, including the $Treat$ variable similarly works like a fixed effect, in that it controls for all time-invariant differences between treatment and control groups. So we don't need to worry about OVB from those factors. But we might be concerned about OVB from time-varying differences between the groups. Seeing whether other variables that we observe are changing differently over time between treatment and control groups before treatment will determine whether it seems reasonable that the two groups would change similarly over time during the treatment period, were it not for the treatment, i.e., whether we should be concerned about OVB from time-varying differences between the groups.

If we lack data for multiple periods before the treatment, we cannot test for parallel pre-treatment trends. In this case we will usually at least check for balance at baseline before the treatment, using the same method used to test for balance in the context of randomized treatment. If treatment and control groups look similar on observable characteristics before the treatment we might expect they would also experience similar trends over time. But balance in a single pre-treatment period is less strong evidence for the key parallel trends assumption than actual being able to show parallel pre-trends.

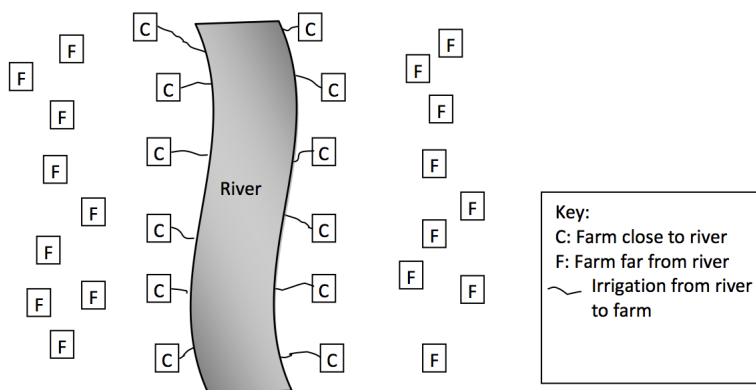
v. Example

The World Bank used to think that big infrastructure projects were the key to development in poor countries. For example, building irrigation passages to divert water from a river to nearby farms, to increase yields. Suppose you were asked to evaluate whether a particular irrigation project successfully increased farmers' yields. (*Solutions to questions at the end of the notes.*)

Attempt 1: Cross sectional regression

Suppose that the World Bank does the project and then collects one season's worth of data on crop yields (metric tons per hectare) for farms in the area, both those close enough to the river to get irrigation and those too far away to be irrigated¹. So the snapshot of the treatment assignment looks something like this:

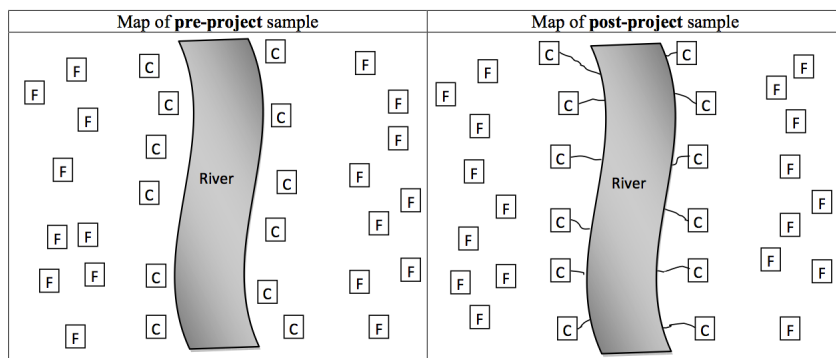
¹Note that if there is threshold distance that determines who is close enough to get irrigation, this would lend itself nicely to an RD analysis.



1. What regression would you “naively” estimate here to find the effect of irrigation, with one season of data on yields?
2. What is the counterfactual here, i.e., the comparison group? What’s wrong with this strategy?

Attempt 2: Difference-in-differences regression

Instead, suppose that the World Bank collects data in two growing seasons, one before the project was started and one after it was completed. Here’s what you now have, along with yield data in each season:



How do we identify the causal find the effect of irrigation?

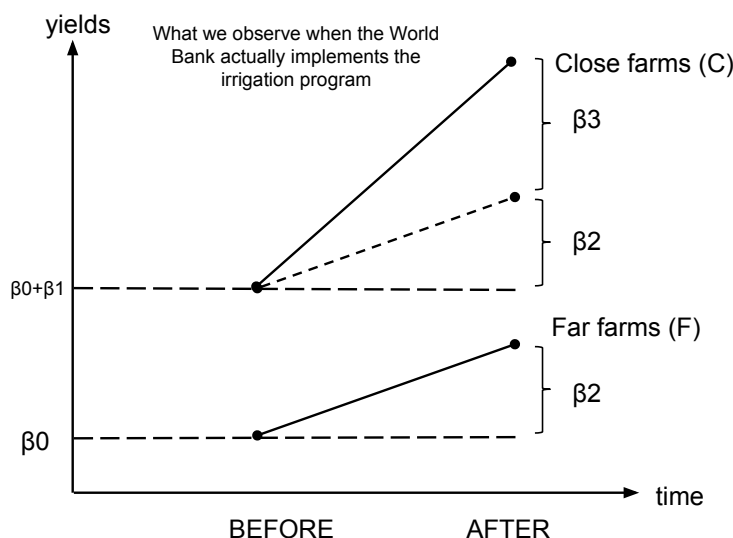
- We run the following regression

$$yield_{it} = \beta_0 + \beta_1 irrigation_i + \beta_2 post_t + \beta_3 (irrigation_i * post_t) + u_{it}$$

- This is better than Attempt 1 because it accounts for differences (some of which we can’t observe) between the C and F farms, getting around the fact that the irrigation was not randomly assigned across farms.

3. What is the counterfactual here?
4. What is the identifying assumption?

We can also draw a picture to understand the diff in diff assumptions and strategy.



The short-dashed line for close farms on this picture demonstrates the key assumption: if not for the irrigation program, the close farms would have had the same *change* in yields over time as the far farms. The solid line for close farms shows what actually happened (the irrigation passages were dug, and the close farms were treated by the program), and you should see how the difference-in-differences strategy finds the treatment effect—given that the key assumption holds.

Basically, the diff-in-diff strategy is to conclude that any difference in the slope of these two lines is due to the treatment (because we are assuming that the slopes *would* have been the same without the program).

So, how do we interpret β_3 ? It's the average treatment effect!² It's the additional difference between irrigated farms and non-irrigated farms after the irrigation passages have been dug—in other words, the estimated effect of improved irrigation on farmer yields under our key assumption.

5. How we can provide some validation for the assumption? What might we still be worried about?

²If not all close farms took up and used the irrigation, this would instead be an intent to treat effect. But we expect all farms would use irrigation once the passages have been dug.

Solutions

1. Attempt 1: What regression would you “naively” estimate here to find the effect of irrigation, with one season of data on yields?

$$yield_i = \beta_0 + \beta_1 irrigation_i + u_i$$

where *irrigation* is a dummy variable equal to 1 if the farmer got an irrigation passage and 0 if not.

2. Attempt 1: What is the counterfactual here, i.e., the comparison group? What’s wrong with this strategy?

Because variable *irrigation* is a dummy indicating whether a farm got irrigation from the river, we’re comparing irrigated farms to non-irrigated farms. Our assumption is that the non-irrigated farms are a good counterfactual for the farms that got the irrigation, i.e., the only real difference between them is the actual irrigation. This shouldn’t feel right: the C farms that got the irrigation are all farms that are close to the river! There are probably a lot of things that vary between C and F farms besides irrigation, and the *irrigation* variable is going to pick *all* of those up, not just the project effect!

Another way to think about it is that distance to the river is an omitted variable in this regression that’s correlated with both irrigation and yields (maybe the soil or other land characteristics vary by distance to the river), so there will be OVB for β_1 .

3. Attempt 2: What is the counterfactual here?

The counterfactual here is much trickier, but you can think of it this way: instead of there being a control *group* of farms that we can point to as what would’ve happened to the treatment farms had the irrigation program not occurred, we have a control *change*. We have an idea of what the *change* in yield for treatment farmers between the pre-project sample and the post-project sample should be—about the same as the *change* in yield for control farmers over the same period.

4. Attempt 2: What is the identifying assumption?

The difference between before and after in the comparison group is a good counterfactual for the treatment group. In other words the trend in outcomes of the comparison group is what we would have observed in the treatment group absent the policy/intervention/reform. You might have some reasons to think this might not be the case (for example, maybe weather was different in the two seasons in a way that differently affected farms near or far from the river with different soils and elevations).

5. Attempt 2: How we can provide some validation for the assumption? What might we still be worried about?

We want to show parallel trends hold. To do this, we can look at yields from farms near the irrigation and for farms far from the irrigation for many periods (days, months, years, etc.) before this data. With this data, we could see whether the slopes, or trends, in yields were the same for both groups leading up to the introduction of the irrigation. If they were pretty similar before, then it sounds more reasonable to assume they would have *continued* to have similar slopes.

Even if this is the case, we might be worry about anything else happening at the time of treatment that affects the closer farm differently from the farther farms. For example, if the river happens to flood in the year of the treatment or if there are other public investments that accompany the irrigation scheme, we might expect closer farmers to be more affected than farther farms, and we would not be able to distinguish how much of the difference between them in the post period is actually due to irrigation. In other words, anything changing differently between treatment and control groups can only be affecting outcomes through the treatment.