Chapter 7. Difference in Differences

Joan Llull

Quantitative & Statistical Methods II
Master in Economics of Public Policy
Barcelona School of Economics

I. Difference in Differences Setup

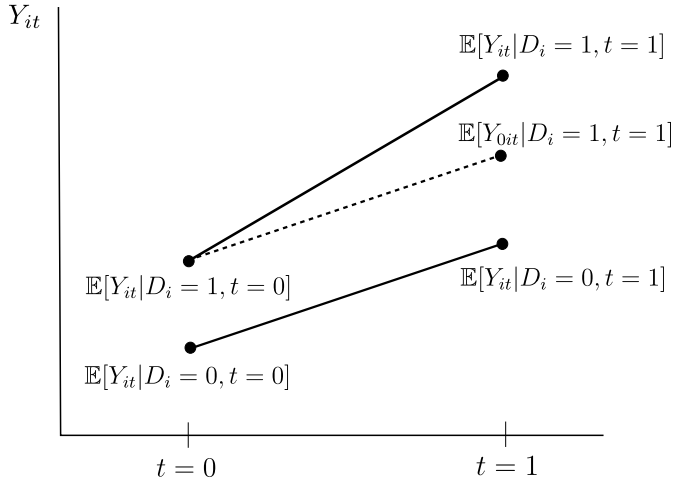
With data from a randomized experiment, the simple comparison of the mean outcome in treatment and control groups (which we can define here as the "difference" estimator) provides an unbiased and consistent estimate of the average treatment effect, as discussed in Chapter 2. This is so because the randomization ensures there are no systematic differences in any "pre-treatment" variables, and, hence, confounding factors are balanced.

In subsequent chapters we have dealt with deviations from the independence assumption. In Chapter 3, and in sharp RD designs in Chapter 5, we proposed different techniques that balance out systematic differences among treated and control units, creating comparable groups, and, thus, ruling out confounders. In Chapter 4 and fuzzy RD designs in Chapter 5 we tried to get causal effects by using instrumental variables. However, good instruments are hard to find, and we would like to have other techniques to rule out unobserved confounders.

The approach in this chapter, which builds on the toolkit developed in Chapter 6, follows an approach that is closer to the first of the two broad approaches described in the previous paragraph. Linking Chapter 6 to treatment effects approaches, we propose an alternative method to eliminate confounders that are fixed over time (like a fixed effect), using repeated observations over time. We assume that, even though treated and control groups are not comparable, the evolution of the outcome pre- and post-treatment would be the same in the absence of treatment. In other words, we assume that treated and control groups have the same counterfactual *trends*, even if the levels differ. In this case, we use data on treatment and control groups before the treatment to estimate the pre-treatment difference between these groups and then compare this difference with the difference in average outcomes after the treatment group received the treatment. Intuitively, the two differences would be equal in the absence of treatment, and the extra difference is imputed to the treatment effect.

The figure below illustrates this discussion. Let Y_{it} denote the observed outcome for individual i in period $t \in \{0, 1\}$, and let $D_i = 1$ if the individual is in the treated

group, with $D_i = 0$ otherwise. Note that we did not subscript D_i by time in this notation, as $D_{it} = 0$ when $t = 0$ for both treated and untreated individuals. For treated individuals we observe $\mathbb{E}[Y_{it}|D_i = 1, t = 0] = \mathbb{E}[Y_{0it}|D_i = 1, t = 0]$, because at $t = 0$ no observation is treated, and $\mathbb{E}[Y_{it}|D_i = 1, t = 1] = \mathbb{E}[Y_{1it}|D_i = 1, t = 1]$, because these individuals are treated at $t = 1$. Likewise, for controls, we observe $\mathbb{E}[Y_{it}|D_i = 0, t = 0] = \mathbb{E}[Y_{0it}|D_i = 0, t = 0]$ as well, but, in this case, the mean observed in the second period is $\mathbb{E}[Y_{it}|D_i = 0, t = 1] = \mathbb{E}[Y_{0it}|D_i = 0, t = 1]$. What we do not observe is $\mathbb{E}[Y_{0it}|D_i = 1, t = 1]$, which we need to compute the average treatment effect on the treated:



What the figure suggests is to use the same trend observed for untreated individuals to predict the counterfactual trend for treated individuals in the absence of treatment. Thus, our prediction of the counterfactual value $\mathbb{E}[Y_{0it}|D_i = 1, t = 1]$ is:

$$\mathbb{E}[Y_{0it}|D_i = 1, t = 1] = \underbrace{\mathbb{E}[Y_{it}|D_i = 0, t = 1]}_{\text{level for controls at } t=1} + \underbrace{\{\mathbb{E}[Y_{it}|D_i = 1, t = 0] - \mathbb{E}[Y_{it}|D_i = 0, t = 0]\}}_{\text{difference in levels at } t=0 \text{ difference}}, \quad (1)$$

which builds on the fundamental assumption that $\mathbb{E}[Y_{0i1} - Y_{0i0}|D_i = 1] = \mathbb{E}[Y_{0i1} - Y_{0i0}|D_i = 0]$. This assumption is known as *the common trend assumption*, and, where there are multiple periods before treatment, it is typically checked by showing that trends before treatmend coincided. Hence, the difference in differences coefficient (which is an average treatment effect on the treated) is:

$$\begin{aligned}\beta &= \mathbb{E}[Y_{1it}|D_i = 1, t = 1] - \mathbb{E}[Y_{0it}|D_i = 1, t = 1] \\ &= \{\mathbb{E}[Y_{it}|D_i = 1, t = 1] - \mathbb{E}[Y_{it}|D_i = 1, t = 0]\} \\ &- \{\mathbb{E}[Y_{it}|D_i = 0, t = 1] - \mathbb{E}[Y_{it}|D_i = 0, t = 0]\}. \quad (2)\end{aligned}$$

Intuitively, β measures the difference between the increase in average observed outcomes for treated and the increase in average observed outcomes for controls.

II. Difference in Differences in the Regression Context

The difference in differences coefficient can be obtained as the β coefficient in the following regression:

$$Y_{it} = \beta_0 + \beta_D D_i + \beta_T T_{it} + \beta D_i T_{it} + U_{it}, \tag{3}$$

where $T_{it} = 1$ if individual i is treatment period $t = 1$, and $T_{it} = 0$ otherwise. With a proof that is very similar than those done in previous chapters, one can prove that β_0 is $\mathbb{E}[Y_{it}|D_i = 0, t = 0]$, $\beta_0 + \beta_D = \mathbb{E}[Y_{it}|D_i = 1, t = 0]$, $\beta_0 + \beta_T = \mathbb{E}[Y_{it}|D_i = 0, t = 1]$, and β is the difference in differences coefficient.

This regression model can be expanded in several ways. First, by including further periods, both before, and after the treatment. In such case, T_{it} is not a time dummy but, instead, a dummy that equals one in the post-treatment period. One could additionally include time effects, but the interaction term should be with the "post" dummy only. Second, the regression allows for controls, X_{it}. In this context, the difference between the regression coefficient and the difference in differences coefficient (obtained nonparametrically from differences in means) is analogous to the difference between matching and regression coefficients discussed in Chapter 3. Third, actually there is no need for panel data to estimate (3): repeated cross sections should suffice. However, in the repeated cross-section context, the researcher needs to sustain the assumption that the sample composition does not vary over time, which is satisfied by construction with panel data. Furthermore, panel data would allow to control for individual fixed effects in the same way we discussed in Chapter 6. Finally, some authors use the same regression setup to build *placebo exercises*. A placebo regression is a regression that simulates the difference in differences analysis but for a point in time or group of individuals that resemble the treatment period or group but that was actually not treated. It is a "placebo" in the sense that it looks as if treatment was administered, but it actually was not.

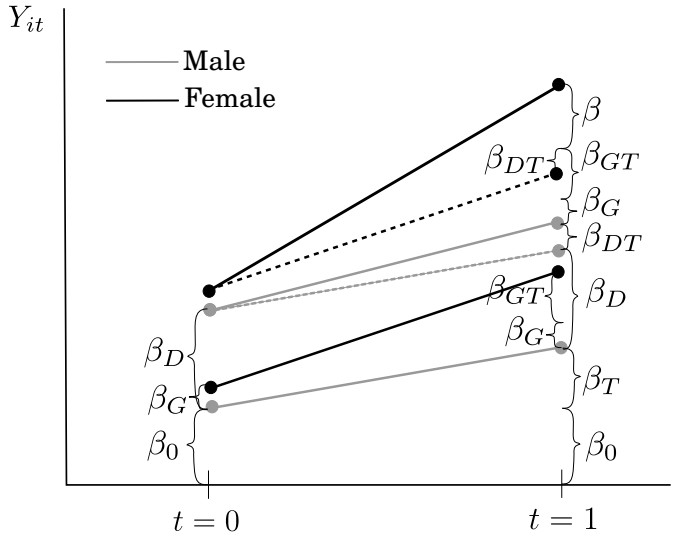
III. Triple Differences Model

Some authors pose *triple-differences* models, in which the difference in differences assumption does not hold, but the change in trends is assumed to be the same across sub-groups, some of which should be more affected than others. For example, let G_i denote the (say sociodemographic) group to which individual i

belongs. Then, the triple-differences model is:

$$Y_{it} = \beta_0 + \beta_D D_i + \beta_T T_{it} + \beta_G G_i + \beta_{GD} G_i D_i + \beta_{GT} G_i T_{it} + \beta_{DT} D_i T_{it} + \beta G_i D_i T_{it} + U_{it}. \quad (4)$$

For example, consider the analysis of maternity leave policies on labor supply. These policies affect young women but do not affect old women. In this context, even though the labor supply of old women is systematically different than that of young women (level difference), this systematic difference persists before and after the policy, and, therefore, we can use old women as a control group in a differences in differences setting. Now imagine that, at the same time that the maternity leave policy is introduced, a tax reform occurs that particularly affects the labor supply of young workers relative to old workers. This additional policy would constitute a counfounder that would break the common trend assumption, because it affects the treated group only in the "post-reform" period, as the maternity leave policy change. However, we have a different group of people, males, that are equally affected by the tax reform, but not affected by the maternity leave policy. In this context, we can use a difference-in-difference estimation for male to "remove" the effect of taxes from the composite effect on female (taxes plus maternal leave policy). In this case, the key assumption is that taxes affect male of different ages in the same way that they affect female. The triple difference coefficients are easily interpreted in the following figure:



IV. Synthetic Control Methods

Consider the case in which we have several periods before treatment is implemented, and, thus, we can check the common trends assumption. For example,

consider the case where one state implements a policy and other states do not. With enough data, we could define as the control the state that has the most similar pre-trend compared to the treated group (or alternatively, all non-treated states). However, often no state is the perfect counterfactual for another.

Synthetic control methods use longitudinal data to build the weighted average of non-treated units that best reproduces the characteristics of the treated unit over time prior to the treatment. Thus, we build an artificial control that has the best possible pre-trend possible, and then we compute the difference in differences estimate using such synthetic control group.