**Samuel Snelson**

**SOC 723 Spring 2023**

**Homework for Chapter 18: Difference-in-Differences**

*How Does It Work?*

1. In the Event Studies chapter we estimated the effect of something that occurs at a specific time by just comparing before-event to after-event, without really using a control group. What assumption is made by no-control-group event studies that we *don’t* have to make with difference-in-differences?
   1. I am not super confident on this one because we didn’t go through the event studies chapter. I don’t think this is quite what NHK is looking for but one important difference between event studies and DID seems to be that there are ways of indirectly assessing whether the treatment group is unique in the change of the outcome over time.
2. Which of the following potential back doors is controlled for by comparing the treated group to a control group?
   1. The treated group may be following a trend, unique to the group, that would make the outcome change from before-treatment to after-treatment anyway.
      1. If there is a time-varying confounder unique to the treatment group, then this is an issue that violates the parallel trends assumption and would bias estimates.
   2. There may be events affecting everyone that would change the outcome from before-treatment to after-treatment anyway.
      1. Comparing the within-variability of the treated group to the control group does account for time-constant differences between them. So, if the “events affecting everyone” are time-constant, then this backdoor would be closed by comparing within-variability.
   3. There may be differences in typical outcome levels between the treated group and the untreated group
      1. Different baseline levels do not present an issue to DID estimation because between-variability is subtracted out.
   4. The decision to treat the treated group, rather than some other group, may be based on factors that are related to the outcome
      1. If the factors being referred to are time-constant, then they will be subtracted out by comparing within-variability. However, if the factors are time-varying, then they present an unresolved issue.
3. Consider a treatment and control group. Looking only at the pre-treatment period, they have exactly the same outcomes (zero gap between them in each period).
   1. Despite having exactly the same outcomes pre-treatment, it happens to be the case that parallel trends is violated for these two groups. How is this possible? Explain what it means for parallel trends to be violated in this case, or give an example of how it could be violated.
      1. Well, if the treatment and control trends are equivalent (and therefore parallel) during the pre-treatment period and they still violate the parallel trends assumption, then this means that the control group must diverge from its path post-treatment. The following plot shows this idea – where the control group is revealed post-treatment to be a potentially poor counterfactual. To make this point clear, our issue with these simulated groups is that the control group appears to respond to the treatment which shouldn’t be happening. This means that there is some unobserved time-varying confounding that makes these groups potentially incomparable.

Chart, line chart

Description automatically generated

* 1. If we estimate the causal effect in this case using difference-in-differences, even though parallel trends is violated, how much would our effect be off by? (note you won’t be able to give a specific number)
     1. Using the above plot as an example of this situation, our estimate of the causal effect would be an overestimate because the difference between the control and treatment groups increases post-treatment. Compared to the DID estimate if the control group’s rate of change had remained constant on average, the apparent DID estimate is larger. This would give the erroneous impression that the treatment has a larger effect than it may actually.
     2. To put this generally, the error of the DID estimates would be in proportion to the extent that parallel trends is violated.

1. Consider the below graph showing the average outcome for treated and control groups in the leadup to treatment (indicated by the dashed line), and also after treatment.  
   Chart, line chart

   Description automatically generated
   1. Based on the prior trend, does it seem likely that parallel trends holds in this instance?
      1. We can think about this plot in the frame of the placebo test that NHK discusses in the chapter. If we just focus on the pre-treatment period, it looks relatively clear that there is a significant interaction occurring. This means that trends are *not* likely to be parallel, this parallel trends seems to be violated based on the pre-treatment trends.
   2. If we estimate difference-in-differences anyway, are we likely to overestimate the actual causal effect, underestimate it, or get it right on average?
      1. Because the pre-treatment trends shows a convergence of the treatment and control groups, the estimate of the causal effect of the treatment would likely be an underestimate. Looking closely at the post-treatment slopes for the treatment and control group, it looks like the control group has a larger slope than the treatment group, indicating that the treatment would be estimated to have a negative effect. If the pre-treatment trend of the control group were parallel with the treatment group (assuming its own existing intercept), the difference in their post-treatment slopes doesn't look like it would be very large, but simply larger than the current estimate.
2. In mid-2020, during the COVID-19 pandemic, different countries pursued different courses of action. Some locked down fully, imposing harsh penalties to most people for leaving the house outside certain proscribed times. Some were looser and only suggested staying at home, and some had hardly any restrictions at all. You notice that COVID rates tend to spike dramatically in different countries at seemingly-random times, and want to know if certain restrictions helped.  
     
   From March through May 2020, US and Canada COVID case rates followed similar trends (US rates were higher, but the trends were similar). You want to look at the effect of COVID restrictions enacted in Canada in late May 2020 on case rates. Is DID, with the US as a control group, a good way to estimate this effect? If not, what concerns would you have about this research design?
   1. In general, it would be a more involved undertaking to analyze the full extent of nations’ responses to COVID as more information came out beginning in 2019. However, for our purposes here, our question really is whether or not the US is a plausible counterfactual to Canada for the 3-month period from March to May 2020. We are particularly concerned about time-varying differences between them because DID tolerated time-constant differences between them. Taking a brief look at the US’ and Canada’s responses to COVID-19 in March 2020, there seems to be some time-varying differences that should not be ignored (see <https://globalhealthsciences.ucsf.edu/sites/globalhealthsciences.ucsf.edu/files/covid-us-case-study.pdf>; <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8297373/#:~:text=On%20March%2018%2C%20the%20Government,%2D19%20Economic%20Response%20Plan>). This means that the time-varying differences between the US and Canada make them potentially unsuitable as counterfactuals for estimates treatment effects of COVID-19 responses.
3. Consider the below table of mean outcomes, and calculate the difference-in-difference effect of treatment. Write out the equation you used to calculate it (i.e. show how the four numbers in the table are combined to get the estimate)

|  |  |  |
| --- | --- | --- |
|  | Before | After |
| Treated | 5 | 9 |
| Untreated | 6 | 7.5 |

* 1. The DID estimate of the effect of the treatment is 2.5 (units).

*How is it Performed?*

1. You are planning to estimate whether voter-protection laws increase voter turnout. You note that, in 2015, a lot of new voter-protection laws were enacted in some provinces but not in others. Conveniently, no new laws were enacted in 2012, 2014, or 2016, so you decide to use 2012 and 2014 as your “before” periods and 2016 as “after”.
   1. Which of the following best describes what you’d want to regress state-and-year level “voter turnout” measures on?
   2. Here are some ways we can do this:

Interaction representing DID term (difference of slopes by group post-treatment) and fixed effects for state-groups and pre- vs post-treatment. Note that Protect refers to the set of states which do and do not have voter-protection laws at any point pre- or post-treatment. PostLaw refers to pre- and post-treatment when voter-protection laws are implemented in 2016. The interaction term then represents the DID estimate for the effect of the voter-protection laws on voter turnout (relative to similar states who didn’t implement voter-protection in the same period)

We can also use TWFE more directly without the interaction term where captures the DID estimate:

* + 1. An indicator for whether the state is treated, and an indicator for whether the year is 2016. (nope, need interaction term for this one)
    2. A set of fixed effects for state, and a set of fixed effects for year. (nope - on the right track, but just the fixed effects don’t estimate the DID)
    3. An indicator for whether the state is treated, a set of fixed effects for year, and an indicator for whether the state is currently treated. (nope)
    4. A set of fixed effects for state, and for year, and an interaction between “is 2016” and “is a treated state”. (this will work)
    5. This design should not be estimated using a regression.
  1. Unless you chose the final option in the previous question, specify which coefficient in that regression would give you the DID estimate.
     1. Answered above. In short, the interaction term in the first specification and in the second.

1. You are looking at a difference-in-difference design to estimate the effect of providing laptops to school children on their test scores. Look at the below regression output, in which “Treated” is an indicator that the school received laptops in 2008 as part of a new program (the untreated group did not receive any laptops until years after the sample window for this study ended), and “After” is an indicator for being after the year 2008.  
     
   Using the table, fill in the blanks in the sentence “Assuming that \_\_\_\_\_, the effect of laptops on test scores was \_\_\_\_\_, and this effect (was/was not) statistically significant at the 95% level.”
   1. Assuming that the treated and control groups are plausible counterfactuals (via there being no relevant time-varying differences between them, the effect of laptops on test scores was about a 5 point increase in test scores, and this effect was statistically significant at the 95% confidence level.

|  |  |
| --- | --- |
|  | **Test Scores** |
| (Intercept) | 80.342\*\*\* |
|  | (0.501) |
| After | 3.369\*\*\* |
|  | (0.696) |
| Treated | 4.116\*\*\* |
|  | (0.718) |
| After× Treated | 5.034\*\*\* |
|  | (0.993) |
| Num.Obs. | 1523 |
|  | 0.188 |
| Standard errors in parentheses.  + p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001 | |

1. A standard “prior trends” test might estimate a regression using the model (only using data from before-treatment), where is a time variable, is an indicator for being in the treated group, and is an outcome variable, and look for a large/significant estimate of . Explain why this test is performed, and specifically what it shows.
   1. To make this more visual, here is a plot showing two sets of prior trends – one good and one not so good.

Chart, line chart

Description automatically generated

* 1. What are tests of prior trends and why do we do them?
     1. Testing prior trends is all about getting some traction on assessing the parallel trends assumption. If the prior trends are reasonably parallel, then we have some basis to think that the post-treatment trends would be parallel if it weren’t for the treatment.
     2. The test for whether or not prior trends are parallel is actually a DID itself where we pick some pre-treatment date as the treatment and the actual treatment as the outcome. If we observe a significant interaction coefficient, then this is telling us that the trends between the treatment and control group are unlikely to be the same (i.e., not parallel trends). If the coefficient is not significant (across a range of available pre-treatment dates as treatments), then this is circumstantial evidence that our prior trends are parallel.
     3. Let’s look at this visually! The top plot on the image above shows prior trends which diverge from the first wave. These are indicative of prior trends which would produce a significant interaction coefficient if we were to conduct a prior trends test. The bottom plot, however, looks better because the prior trends are more close to parallel (giving a suggestion of parallel trends and therefore plausible counterfactuality).
        1. In the bottom plot, there is still some movement after the treatment which may present some concerns about time-varying differences, but these can be investigated separately.

1. Consider the below graph with estimates from a dynamic difference-in-differences model for a treatment that occurs between periods 4 and 5, with 95% confidence intervals shown.

Chart, line chart

Description automatically generated

* 1. What about this graph might make us concerned about our identification assumptions?
     1. The plot suggests that there is a significant difference between treatment and control groups in wave 1 (relative to wave 4). This is concerning because this indicates that the prior trends are different and gives a warning that the parallel trends assumption may not be met.
  2. Ignoring any concerns we have, what would we say is the effect of treatment on Y in this case? (note the height of the line in period 5 is about 3, in period 6 is about 1, and in period 7 is about .5).
     1. The effect of treatment on Y appears quite clearly to have a diminishing effect from wave 5 to 6 and 7. The effect of the treatment on Y is about 3 in the wave immediately after the treatment. This effect decreased by a factor of about 1/3 in the following wave and is not significant in the next wave.

1. Chapter 18.2.5 points out a problem with two-way fixed effects in cases where treatment is not all assigned at the same time, but rather different groups get treated at different times (a “rollout” design). In these designs, two-way fixed effects treats “already-treated” units, who were treated in earlier periods, as “control” units, as though they hadn’t gotten treated at all. However, there’s nothing theoretically wrong about using an already-treated unit as a control; the DID assumptions don’t require that the control group be untreated, just that the gap between treated and control doesn’t change when the treated group’s treatment goes into effect. Why are we so concerned, then, about using an already-treated group as a control? You can answer generally, or use as an example a DID with only two groups – an already-treated group and a newly-treated group. (hint: to do the example, try assuming the treatment only has an effect for the single period after treatment, and the already-treated group is treated exactly one period before the treated group)
   1. We have indirectly discussed this issue in class in that two-period DID doesn’t directly generalize to cases where they have always been treated or never treated. Why do I mention this? Because two-period DID estimates concern those with variable treatment statuses, there is a possibility that there are systematic differences between the subset of cases used to estimate DID and those who always or never get treatment. As it related to rollout designs, we are including cases which get treatment at, say t = x-1, as a control case for the estimation of the effect of treatment at t = x. If those different cases have systematically different effects over time, then this comparison presents an identification concern because we assume that the effects are the same for them for calculation of the effect by wave.