Lab #19 - Differences-in-Differences

Econ 224

November 13th, 2018

Introduction

The exercises in this lab are adapted from a problem set by Steve Pischke. In the first part, you'll look at a conceptual example to improve your understanding of how and when diff-in-diff works and its relation to other methods of causal inference. In the second part, you'll work through an extended empirical example about the effect of raising the minimum wage. My expectation is that you will need to work on this lab on both Tuesday and Thursday. Please submit it as your final problem set, just as we did with the RD labs last week.

Exercise I: Conceptual

Suppose we observe a dataset with information on a sample of $i = 1, ..., N$ farms over $t = 1, ..., T$ years:

Name	Description
yield	yield of wheat (bushels/acre) on farm i in year t
capital	value of machinery (in constant dollars) used on farm i in year t
labor	number of workers employed on farm i in year t
rain	rainfall (inches) on farm i in year t
fertilizer	amount of fertilizer (1000 of pounds) used on farm i in year t
size	size in acres of farm i

Solutions

Exercise II: Applied

The following questions are based on a dataset called `minwage.dta` that you can download from the Mastering 'Metrics website: click on "Instructor's Corner," then scroll down to the bottom of the page. This dataset contains information collected from fast food restaurants in New Jersey and eastern Pennsylvania during two interview waves, the first in March of 1992 and the second in November-December of the same year. Between these two interview waves – on April 1st to be precise – the New Jersey minimum wage increased by just under 19%, from $4.25 to $5.05 per hour. The minimum wage in Pennsylvania was unchanged during this period: $4.25 per hour. The `minwage.dta` dataset is drawn from a famous but controversial study of the effects of minimum wages by Angrist & Kreuger. The study is so famous that there is even an oblique reference to it on the label of a certain brand of shampoo! (Sadly they do not provide the full citation.) Here is a description of the variables that you will need to carry out this exercise. When you see a pair of variables in the table below, e.g. `fte`/`fte2`, both measure the same thing but the one with the `2` is based on the *second* survey wave, while the one without the `2` is based on the *first* survey wave.

Name	Description
`wage_st`/`wage_st2`	Starting wage in dollars/hour at the restaurant

Name	Description
`fte`/`fte2`	Full-time equiv. employment = #(Full time employees) + #(Part-time Employees)/2. Excludes managers.
`chain`	Categorical variable taking values in $\{1, 2, 3, 4\}$ to indicate the four different chains in the dataset: Burger King, KFC, Roy Rogers, and Wendy's
`co_owned`	Dummy variable (=1 if restaurant is company-owned; =0 if it is franchised)
`sample`	Dummy variable (=1 if wage and employment data are available for both survey waves at this restaurant)

1. Preliminaries:

   (a) Download the data and convert it to a format that R can understand, using an appropriate package.
   (b) Restrict the sample to only those restaurants with `sample` equal to 1 to ensure that we are making an apples-to-apples comparison throughout the remainder of this exercise.

2. Baseline Diff-in-Diff Estimate: starting wages

   (a) Calculate the average wage in each survey wave separately for each state.
   (b) Calculate the within-state time-differences based on (a).
   (c) Calculate the between-state difference-in-differences based on (c).
   (d) Interpret your findings from (c). What do they tell us about the causal effect of increasing the minimum wage? What assumptions are required for this interpretation to be valid?

3. Baseline Diff-in-Diff Estimate: full time equivalent employment

   (a) Repeat question 2 but using full-time equivalent employment as the outcome variable rather than starting wages.

Solutions

```r
library(tidyverse)
library(haven)
ak <- read_dta('~/econ224/labs/minwage.dta')
ak <- ak %>% filter(sample == 1)

DinD <- ak %>% group_by(state) %>%
  summarize(mean_wage_st = mean(wage_st),
            mean_wage_st2 = mean(wage_st2)) %>%
  mutate(diff = mean_wage_st2 - mean_wage_st)
DinD
```

```
# A tibble: 2 x 4
  state mean_wage_st mean_wage_st2    diff
  <dbl>        <dbl>         <dbl>   <dbl>
1     0         4.65          4.62 -0.0348
2     1         4.61          5.08  0.469
```

```r
names(ak)
```

```
 [1] "sheet"    "chain"    "co_owned" "state"    "empft"    "emppt"
 [7] "wage_st"  "empft2"   "emppt2"   "wage_st2" "fte"      "fte2"
[13] "dfte"     "gap"      "dw"       "sample"
```

```r
table(ak$chain)
```

```
  1   2   3   4
145  74  87  45
```