In-Class Lab 16

ECON 4223 (Prof. Tyler Ransom, U of Oklahoma)

April 14, 2022

The purpose of this in-class lab is to use R to practice with difference-in-differences. To get credit, upload your .R script to the appropriate place on Canvas.

For starters

Open up a new R script (named `ICL16_XYZ.R`, where `XYZ` are your initials) and add the usual "preamble" to the top:

```
library(tidyverse)
library(wooldridge)
library(broom)
library(magrittr)
library(modelsummary)
library(estimatr)
```

Load the data

Our data set will be a pooled cross section of workers in Kentucky and Michigan, as analyzed by Meyer, Viscusi, and Durbin (1995). Load the data from the `wooldridge` package and restrict to those living in Kentucky:

```
df <- as_tibble(injury)
df %<>% filter(ky==1)
```

Policy setting

In 1980, Kentucky raised its cap on weekly earnings that were covered by worker's compensation.[1] The outcome variable is `ldurat`, which is the log duration (in weeks) of worker's compensation benefits. The policy was such that the cap increase did not affect low-earnings workers, but did affect high-earnings workers. Thus, low-earnings workers serve as the control group, while high-earnings workers serve as the treatment group.

We are interested to know if the policy caused workers to spend more time out of work. If benefits are not generous enough, then workers may sue the company for on-the-job injuries. On the other hand, benefits that are too generous may induce workers to be more reckless on the job, or to claim that off-the-job injuries were incurred while at work.

Summary statistics

In difference-in-differences settings, it is often helpful to see if there even is a policy effect. Let's compute the difference-in-differences (with no regression) based on the pre- and post-means for the treatment and control groups

[1] Worker's compensation is "a form of insurance providing wage replacement and medical benefits to employees injured in the course of employment in exchange for mandatory relinquishment of the employee's right to sue their employer for ... negligence." (Wikipedia)

```
df %>% group_by(afchnge,highearn) %>% summarize(mean.ldurat = mean(ldurat))
```

1. What is the difference in the differences? What is its interpretation? (Hint: recall that the dependent variable is logged)

Difference-in-differences regression

The basic regression model to analyze the policy's impact is

$$\log(durat) = \beta_0 + \delta_0 afchnge + \beta_1 highearn + \delta_1 afchnge \cdot highearn + u$$

Estimate this model, calling it `est.did`. I suppress the code here, since this should be old hat by now. (Hint: recall that, for two dummy variables `x1` and `x2`, putting `x1*x2` in the `lm()` formula will include in the formula each dummy and the interaction of the two.)

2. Verify that your estimated $\hat{\delta}_1$ is the same as your answer in (1). Is the effect significant?

Including more x's

We might want to control for other aspects of our workers. For example, perhaps claims made by construction or manufacturing workers tend to have longer duration than claims made workers in other industries. Or maybe those claiming back injuries tend to have longer claims than those claiming head injuries. One may also want to control for worker demographics such as gender, marital status, and age.

Estimate an expanded version of the basic regression model, where the following additional variables are included:

- `male`
- `married`
- Quadratic in `age`
- `hosp` (1 = hospitalized)
- `indust` (1 = manuf, 2 = construc, 3 = other)
- `injtype` (1-8; categories for different types of injury)
- `lprewage` (log of wage prior to filing a claim)

Be sure to format `indust` and `injtyp` as factors before proceeding. Call your model `est.did.x`.

Calling `modelsummary` can help view the results:

```
modelsummary(list(est.did,est.did.x),output="latex")
```

3. How different is your estimate of $\hat{\delta}_1$ from that in (2)?
4. Do you see any interesting patterns among the x's in predicting log duration of benefits?

References

Meyer, Bruce D., W. Kip Viscusi, and David L. Durbin. 1995. "Workers' Compensation and Injury Duration: Evidence from a Natural Experiment." *American Economic Review* 85 (3): 322–40.

	Model 1	Model 2
(Intercept)	1.126	−1.559
	(0.030)	(0.457)
afchnge	0.008	0.017
	(0.044)	(0.044)
highearn	0.256	−0.160
	(0.047)	(0.094)
afchnge × highearn	0.191	0.215
	(0.069)	(0.070)
male		−0.071
		(0.046)
married		0.041
		(0.040)
age		0.025
		(0.008)
I(age^2)		0.000
		(0.000)
injtype2		0.775
		(0.170)
injtype3		0.346
		(0.101)
injtype4		0.632
		(0.108)
injtype5		0.493
		(0.103)
injtype6		0.392
		(0.102)
injtype7		0.773
		(0.239)
injtype8		0.510
		(0.142)
lprewage		0.314
		(0.087)
indust2		0.262
		(0.056)
indust3		0.182
		(0.040)
Num.Obs.	5626	5347
R2	0.021	0.049
R2 Adj.	0.020	0.046
Std.Errors	HC2	HC2