

# Lab #19 - Minimum Wages and Unemployment

*Econ 224*

*November 13th, 2018*

## Introduction

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 & Krueger. 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
<code>state</code>	Dummy variable =1 for NJ, =0 for PA
<code>wage_st</code> / <code>wage_st2</code>	Starting wage in dollars/hour at the restaurant
<code>fte</code> / <code>fte2</code>	Full-time equiv. employment = $\#(\text{Full time employees}) + \#(\text{Part-time Employees})/2$ . Excludes managers.
<code>chain</code>	Categorical variable taking values in $\{1, 2, 3, 4\}$ to indicate the four chains in the dataset: Burger King, KFC, Roy Rogers, and Wendy’s
<code>co_owned</code>	Dummy variable =1 if restaurant is company-owned, =0 if franchised
<code>sample</code>	Dummy variable =1 if wage and employment data are available for both survey waves at this restaurant

## Exercises

### 1. Preliminaries:

- Download the data and load it in R using an appropriate package.
- 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 lab.
- Rename the column `state` to `treat`.
- Create a *new* column called `state` that equals PA if `treat` is 0 and NJ if `treat` is 1.
- Create a column called `low_wage` that takes the value 1 if `wage_st` is less than 5.

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

- Calculate the average wage in each survey wave separately for each state.
- Calculate the within-state time-differences based on (a).
- Calculate the between-state difference-in-differences based on (c).
- 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.
4. Reshape `minwage` for Diff-in-Diff Regression Estimation:
  - (a) Create a tibble called `wave1` containing only the columns `state`, `treat`, `wage_st`, `fte`, `chain`, `co_owned`, and `low_wage`. Add a column called `post` to `wave1` that equals 0 for every observation.
  - (b) Create a tibble called `wave2` containing only the columns `state`, `treat`, `wage_st2`, `fte2`, `chain`, `co_owned`, and `low_wage`. Rename `wage_st2` to `wage_st` and `fte2` to `fte`. Then add a column called `post` to `wave2` that equals 1 for every observation.
  - (c) Create a tibble called `both_waves` by *stacking* `wave1` on top of `wave2`. You can do this using the `bind_rows` command from `dplyr`. (Read the help file for more details.)
5. Diff-in-Diff Regression Estimates:
  - (a) Consider the following regression model using the variables `treat` and `post` constructed above:
 
$$Y_{i,s,t} = \beta_0 + \beta_1(\text{treat}_{i,s}) + \beta_2(\text{post}_t) + \beta_3(\text{treat}_{i,s} \times \text{post}_t) + \epsilon_{i,s,t}$$

where  $i$  indexes *restaurants*,  $s$  indexes *states*, and  $t$  indexes *time periods*, i.e. the two survey waves. Explain the meaning of each of the four regression coefficients. Which one gives the Regression differences-in-differences effect?
  - (b) Estimate the regression from part (a) based on `both_waves` using `wage_st` as the outcome variable. Summarize your results, including appropriate statistical inference. How do they compare to those that you calculated in question 2 above?
  - (c) Estimate the regression from part (a) based on `both_waves` using `fte` as the outcome variable. Summarize your results, including appropriate statistical inference. How do they compare to those that you calculated in question 3 above?
  - (d) An advantage of the regression-based formulation of differences-in-differences is that it allows us to control for other variables that might affect wages and employment. Repeat parts (b) and (c) adding `co_owned` and dummy variables for each of the four restaurant chains to your regression. Hint: rather than creating separate dummy variables from each of the values that `chain` can take, use `as.factor()` to convert `chain` to a factor. Then if you include `chain` in a regression, R will automatically create the dummy variables for you.
  - (e) How do your results from part (d) compare with those of parts (b) and (c)?
6. Probing the Diff-in-Diff Assumption:
  - (a) What assumption is required for the diff-in-diff approach to provide a valid causal estimate of the effects of New Jersey raising its minimum wage?
  - (b) An alternative to the comparison of NJ and PA restaurants is a *within* NJ comparison. The key insight here is that only restaurants with starting wages below \$5 per hour in the first wave will be affected by the change in minimum wages. Use the variable `low_wage` to run this alternative to the regression from 5(a) using only observations from NJ. Discuss your findings.
  - (c) What assumption is needed for the DD estimate from (b) to be reliable? How plausible is this assumption compared to the assumption from (a)?
  - (d) Repeat part (b) but restrict attention to restaurants in PA where there was no change in minimum wages. Discuss your findings. What do these results suggest about the plausibility of the diff-in-diff assumption in part (b)?

## Solutions

```
# 1 - Preliminaries
library(tidyverse)
library(haven)
minwage <- read_dta('~/.econ224/labs/minwage.dta')
minwage <- minwage %>% filter(sample == 1) %>%
  rename(treat = state) %>%
  mutate(state = case_when(treat == 0 ~ 'PA',
                           treat == 1 ~ 'NJ'),
         low_wage = 1 * (wage_st < 5))
```

```
# 2 - Baseline Diff-in-Diff: starting wages
DinD_wage <- minwage %>% 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_wage
```

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

```
with(DinD_wage, diff[1] - diff[2])
```

```
[1] 0.5040066
```

```
# 3 - Baseline Diff-in-Diff: full-time equivalent employment
DinD_emp <- minwage %>% group_by(state) %>%
  summarize(mean_fte = mean(fte),
            mean_fte2 = mean(fte2)) %>%
  mutate(diff = mean_fte2 - mean_fte)
DinD_emp
```

```
# A tibble: 2 x 4
  state mean_fte mean_fte2 diff
  <chr>      <dbl>      <dbl>  <dbl>
1 NJ          17.3          17.6  0.287
2 PA          20.1          18.1 -2.02
```

```
with(DinD_emp, diff[1] - diff[2])
```

```
[1] 2.301994
```

```
# 4 - Reshape dataset for Diff-in-Diff regression estimation
wave1 <- minwage %>%
  select(state, treat, wage_st, fte, chain, co_owned, low_wage) %>%
  mutate(post = 0)
wave2 <- minwage %>%
  select(state, treat, wage_st2, fte2, chain, co_owned, low_wage) %>%
```

```
mutate(post = 1) %>%
  rename(wage_st = wage_st2, fte = fte2)
both_waves <- bind_rows(wave1, wave2)
```

## Solution to 5(a)

Blah blah blah...

```
# 5 - Diff-in-Diff regression results
library(stargazer)
reg_wage1 <- lm(wage_st ~ treat + post + treat:post, both_waves)
reg_emp1 <- lm(fte ~ treat + post + treat:post, both_waves)

stargazer(reg_wage1, reg_emp1, type = 'latex', header = FALSE, digits = 2,
  dep.var.labels = c('Starting Wage', 'Full-time Equiv. Employment'),
  omit.stat = c('f', 'ser', 'adj.rsq', 'rsq'))
```

Table 2:

	<i>Dependent variable:</i>	
	Starting Wage	Full-time Equiv. Employment
	(1)	(2)
treat	-0.04 (0.04)	-2.84** (1.22)
post	-0.03 (0.05)	-2.02 (1.56)
treat:post	0.50*** (0.05)	2.30 (1.73)
Constant	4.65*** (0.03)	20.11*** (1.10)
Observations	702	702

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

```
#----- (d) control for co_owned and chain
both_waves <- both_waves %>% mutate(chain = as.factor(chain))
reg_wage2 <- lm(wage_st ~ treat + post + treat:post + co_owned + chain, both_waves)

reg_emp2 <- lm(fte ~ treat + post + treat:post + co_owned + chain, both_waves)

stargazer(reg_wage2, reg_emp2, type = 'latex', header = FALSE, digits = 2,
  dep.var.labels = c('Starting Wage', 'Full-time Equiv. Employment'),
  omit.stat = c('f', 'ser', 'adj.rsq', 'rsq'))
```

Table 3:

	<i>Dependent variable:</i>	
	Starting Wage	Full-time Equiv. Employment
	(1)	(2)
treat	−0.04 (0.04)	−2.14* (1.10)
post	−0.03 (0.05)	−2.02 (1.40)
co_owned	0.07*** (0.02)	−1.01 (0.72)
chain2	0.02 (0.03)	−10.16*** (0.84)
chain3	0.05 (0.03)	−1.35 (0.86)
chain4	0.12*** (0.03)	−1.37 (0.97)
treat:post	0.50*** (0.05)	2.30 (1.55)
Constant	4.59*** (0.04)	22.56*** (1.05)
Observations	702	702
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01		

```

# 6 - Probing the diff-in-diff assumption
nj_only <- both_waves %>% filter(state == 'NJ')
pa_only <- both_waves %>% filter(state == 'PA')

# Within NJ comparison
nj_wage <- lm(wage_st ~ low_wage + post + low_wage:post, nj_only)
nj_fte <- lm(fte ~ low_wage + post + low_wage:post, nj_only)
stargazer(nj_wage, nj_fte, type = 'latex', header = FALSE, digits = 2,
  dep.var.labels = c('Starting Wage', 'Full-time Equiv. Employment'), omit.stat = c('f', 'ser',
  title = 'Within NJ comparison')

```

Table 4: Within NJ comparison

	<i>Dependent variable:</i>	
	Starting Wage	Full-time Equiv. Employment
	(1)	(2)
low_wage	-0.65*** (0.02)	-2.23* (1.21)
post	-0.004 (0.03)	-2.25 (1.50)
low_wage:post	0.62*** (0.03)	3.30* (1.71)
Constant	5.11*** (0.02)	18.99*** (1.06)
Observations	570	570

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

```

# Within PA comparison
pa_wage <- lm(wage_st ~ low_wage + post + low_wage:post, pa_only)
pa_fte <- lm(fte ~ low_wage + post + low_wage:post, pa_only)
stargazer(pa_wage, pa_fte, type = 'latex', header = FALSE, digits = 2,
  dep.var.labels = c('Starting Wage', 'Full-time Equiv. Employment'), omit.stat = c('f', 'ser',
  title = 'Within PA comparison')

```

Table 5: Within PA comparison

	<i>Dependent variable:</i>	
	Starting Wage	Full-time Equiv. Employment
	(1)	(2)
low_wage	−0.63*** (0.07)	−0.89 (2.67)
post	−0.27*** (0.08)	−3.85 (3.04)
low_wage:post	0.35*** (0.10)	2.81 (3.77)
Constant	5.07*** (0.06)	20.70*** (2.15)
Observations	132	132

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01