Problem Set 2 Labour Economics

Alvaro Rodriguez, Benedikt Schmidt, Nhat Huynh

2025-11-12

Imports, working directory, and dataset.

```
library(MASS)
library(tidyverse)
library(GGally)
library(labelled)
library(dplyr)
library(rstatix)
library(ivreg)
library(simstudy)
library(ivmodel)
library(ggplot2)
library(car)
library(lmtest)
library(tseries)
library(ggfortify)
library(plotly)
library(haven)
library(broom)

setwd("C:/Users/alvar/Documents/TU_DORTMUND_4_semester/Labour_economics/")
rm(list = ls())

# load data
minwage <- read_dta("Data and Code-20251030/minwage.dta")
```

Task a)

Calculate the average starting wage (wage_st) separately for restaurants in NJ and in PA, both for each interview wave.

```
# swap the "state" variable with the state names (not a must, but makes it clearer)
minwage$state <- recode(minwage$state, "0='PA'; 1='NJ'")

# select data that is available in both survey waves
df <- minwage %>%
  filter(sample == 1)

######## a) ########
wages <- df %>%
```

```r
  group_by(state) %>%
  summarise(wage_st2 = mean(wage_st2),
            wage_st = mean(wage_st),
            dw = mean(dw))

# i) Calculate the difference in the average wages between the second and first interviews.
difference <- wages$wage_st2 - wages$wage_st

# ii) Now calculate the difference between NJ and PA of the time differences just obtained.
diff_row <- wages %>%
  filter(state %in% c("NJ", "PA")) %>%
  summarise(wage_st2 = -diff(wage_st2),
            wage_st = -diff(wage_st),
            dw = -diff(dw),
            state = "diff")

wages <- bind_rows(wages, diff_row)
print(wages)
```

```
## # A tibble: 3 x 4
##   state wage_st2 wage_st      dw
##   <chr>    <dbl>   <dbl>   <dbl>
## 1 NJ       5.08   4.61    0.469
## 2 PA       4.62   4.65   -0.0348
## 3 diff     0.463 -0.0407  0.504
```

iii) What is the interpretation of such a difference-in-differences estimate of the wage effect? Under what conditions does this provide a valid estimate of the minimum wage increase on wages in the fast food industry?

Answer: The DiD estimate of 0.504 represents the causal effect of the new minimum wage law on starting wages in New Jersey. it implies that the legislative increase in the minimum wage caused average starting wages in NJ fast food restaurants to rise by approximately 0.50$ more than they would have in the absence of the policy change.This assumes that, in the absence of the minimum wage increase in NJ, the trend in starting wages in New Jersey would have been the same as the trend observed in Pennsylvania.This assumes that, in the absence of the minimum wage increase in NJ, the trend in starting wages in New Jersey would have been the same as the trend observed in Pennsylvania.

iv) Interpret your finding.

In NJ average wages increased significantly by \$0.469 (from \$4.61 to \$5.08). This is expected, as the minimum wage was raised to \$5.05. The post-treatment average (\$5.08) is slightly above the new binding minimum.

In Pennsylvania the average wages remained largely stagnant, actually decreasing slightly by \$0.035 (from \$4.65 to \$4.62). This suggests that without the law, wages were not naturally rising.

The finding indicates that the policy was effective. The law successfully raised wages in the targeted sector. The increase of \$0.50 is substantial, though it is less than the full statutory increase of \$0.80 (from \$4.25 to \$5.05). This is likely because some restaurants were already paying above the old minimum wage of \$4.25 before the hike occurred, so their wages did not need to increase by the full \$0.80 margin to comply.

Task b)

Repeat the same exercise as in (a) for full time equivalent employment. What is the impact of the minimum wage increase on relative employment in NJ restaurants?

```
## # A tibble: 3 x 4
##   state   fte2   fte   dfte
##   <chr>  <dbl> <dbl>  <dbl>
## 1 NJ    17.6   17.3   0.287
## 2 PA    18.1   20.1  -2.02
## 3 diff  -0.536 -2.84  2.30
```

This indicates that, relative to Pennsylvania, full-time equivalent employment in New Jersey fast-food restaurants increased by approximately 2.3 employees per restaurant. This suggests that the minimum wage increase did not reduce employment in New Jersey. In fact, relative to the control group, employment appeared to expand.

Task c)

#i)

Write for period 1 and 2 (t = 0, POST = 0, t = 1, POST = 1):

$$Y_{is1} = \beta TREAT_{is} + e_{is1}$$

$$Y_{is2} = \beta TREAT_{is} + \gamma + \delta_{rDD} TREAT_{is} + e_{is2}$$

Then subtract period 1 from 2:

$$Y_{is2} - Y_{is1} = (\beta TREAT_{is} + \gamma + \delta_{rDD} TREAT_{is} + e_{is2}) - (\beta TREAT_{is} + e_{is1})$$

which simplifies to:

$$Y_{is2} - Y_{is1} = \gamma + \delta_{rDD} TREAT_{is} + (e_{is2} - e_{is1})$$

```
##
## Call:
## lm(formula = dw ~ state, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.03916 -0.21515  0.03485  0.33084  2.03485
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.03485    0.04287  -0.813    0.417
## state        0.50401    0.04757  10.595   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3483 on 349 degrees of freedom
```

```
## Multiple R-squared:  0.2434, Adjusted R-squared:  0.2412
## F-statistic: 112.2 on 1 and 349 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = dfte ~ state, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -41.485  -3.287   0.213   4.463  25.765
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -2.015      1.052  -1.916   0.0562 .
## state          2.302      1.167   1.972   0.0494 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.546 on 349 degrees of freedom
## Multiple R-squared:  0.01102,    Adjusted R-squared:  0.008184
## F-statistic: 3.888 on 1 and 349 DF,  p-value: 0.04942
```

```
##
## Call:
## lm(formula = dw ~ state + co_owned + as.factor(chain), data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.05863 -0.21156  0.00137  0.25137  1.95503
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.04497    0.04744   0.948  0.34379
## state              0.50366    0.04693  10.731  < 2e-16 ***
## co_owned          -0.03676    0.04308  -0.853  0.39413
## as.factor(chain)2 -0.04665    0.05084  -0.918  0.35945
## as.factor(chain)3 -0.15112    0.05180  -2.917  0.00376 **
## as.factor(chain)4 -0.15024    0.05846  -2.570  0.01060 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3419 on 345 degrees of freedom
## Multiple R-squared:  0.279,  Adjusted R-squared:  0.2685
## F-statistic:  26.7 on 5 and 345 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = dfte ~ state + co_owned + as.factor(chain), data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -41.893  -3.628   0.469   4.372  25.357
##
```

```
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -1.6073     1.1867  -1.354   0.1765
## state               2.2973     1.1741   1.957   0.0512 .
## co_owned            0.3394     1.0777   0.315   0.7530
## as.factor(chain)2   0.2990     1.2719   0.235   0.8143
## as.factor(chain)3  -1.9637     1.2960  -1.515   0.1306
## as.factor(chain)4  -0.7816     1.4626  -0.534   0.5934
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.554 on 345 degrees of freedom
## Multiple R-squared:  0.0207, Adjusted R-squared:  0.006506
## F-statistic: 1.458 on 5 and 345 DF,  p-value: 0.2029
```

#ii) The regression approach yields the exact same coefficients as the manual "means-comparison" approach. This confirms that regressing the change in the outcome on the treatment dummy is mathematically equivalent to the standard Difference-in-Differences formula.

#iv) We would generally not expect the results to change significantly. For the inclusion of control variables (like chain or co_owned) to change the treatment coefficient, those variables would need to be correlated with the treatment variable (state). Since the treatment is roughly exogenous to these restaurant characteristics, there is no omitted variable bias to correct. Adding controls in this context mainly serves to reduce the standard errors (increase precision) rather than correct the coefficient estimate. ## Task d)

#i) Generally, we would expect that the assumptions are satisfied more easily. The within analysis of NJ restaurants is (more) robust by design to different regional specific economic conditions. Thus, the parallel trend assumption may be more plausible, holding both groups geographically (and thereby institutionally) fixed.

```
##
## Call:
## lm(formula = dw ~ low_wage, data = nj)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.56591 -0.16178  0.05409  0.18822  0.55822
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.004091   0.026719  -0.153    0.878
## low_wage     0.615872   0.030480  20.206   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2171 on 283 degrees of freedom
## Multiple R-squared:  0.5906, Adjusted R-squared:  0.5892
## F-statistic: 408.3 on 1 and 283 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = dfte ~ low_wage, data = nj)
##
## Residuals:
```

```
##      Min      1Q  Median      3Q     Max
## -35.051  -3.551  -0.051   3.949  24.949
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -2.2500     0.9472  -2.375  0.01820 *
## low_wage      3.3014     1.0806   3.055  0.00246 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.695 on 283 degrees of freedom
## Multiple R-squared:  0.03193,    Adjusted R-squared:  0.02851
## F-statistic: 9.334 on 1 and 283 DF,  p-value: 0.002464
```

#ii)

Wage effect: The coefficient on low_wage is 0.616 ($p<0.001$). This implies that starting wages in the restaurants directly affected by the minimum wage reform (those paying below $5.00 beforehand) shifted positively by about $0.62 more than in restaurants already paying above the threshold. This confirms that the policy was binding for the low-wage group, and the magnitude is close to the statutory increase of wages, again matching our expectations.

Employment effect: The coefficient on low_wage is 3.30 ($p \approx 0.002$), implying that employment in affected low-wage restaurants increased by approx. 3.3 full-time equivalent workers relative to the higher-wage restaurants.

The intercept of -2.25 indicates that high-wage restaurants experienced a declie of about 2.25 (fte) jobs, while low-wage restaurants saw a net gain of approx. 1.05 jobs (-2.25+3.30).

#iii) The estimates are consistent with the analysis in part c) and show even stronger effects on both wages and employment. For wages, the NJ-PA estimate from part c) is approx. 0.50, while our within estimate is bigger with 0.62. This is likely due to the fact that the control group (high-wage NJ stores) had very little wage growth compared to PA stores. A similar pattern appears for employment. The NJ-PA DiD estimate is approx. 2.3 (fte) jobs, while the within-NJ estimate is 3.3 jobs. The fact that the within-NJ estimate is larger suggests that when using a more comparable control group inside the treated state, the relative employment gains appear stronger.

##Task e)

#i) Compared to the within-NJ results from part (d), where the corresponding coefficients were about 0.62 for wages and 3.3 for employment, the PA effects are smaller in magnitude for wages and somewhat smaller for employment. Moreover, according to the regression output, the employment effect in PA is not statistically significant at the 5% level, so we cannot reject the null hypothesis of no differential employment change between low- and high-wage restaurants in PA.

#ii) To test whether the effect of being a low-wage restaurant is the same in New Jersey and Pennsylvania, we run the interacted regression:

$$Y = \beta_0 + \beta_1 \text{low\_wage} + \beta_3(\text{NJ} \times \text{low\_wage}) + e$$

Our parameter of interest is:

β_3 the interaction term = difference between the NJ low-wage premium and the PA low-wage premium. If $\beta_3 = 0$ the low wage effect is the same in both states. but if $\beta_3 \neq 0$ then the low wage effect is different in each state.

```r
# only consider restaurants in PA
pa <- df %>%
  filter(state == 0)

pa$low_wage <- ifelse(pa$wage_st < 5, 1, 0)

# i)
## for wages
reg_wages_e <- lm(dw ~ low_wage, data = pa)
summary(reg_wages_e)
```

```
##
## Call:
## lm(formula = dw ~ low_wage, data = pa)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.58837 -0.08837 -0.08837  0.13913  1.91163
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.26522    0.07318  -3.624 0.000575 ***
## low_wage     0.35359    0.09066   3.900 0.000233 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3509 on 64 degrees of freedom
## Multiple R-squared:  0.192,  Adjusted R-squared:  0.1794
## F-statistic: 15.21 on 1 and 64 DF,  p-value: 0.0002331
```

```r
## for employment
reg_emp_e <- lm(dfte ~ low_wage, data = pa)
summary(reg_emp_e)
```

```
##
## Call:
## lm(formula = dfte ~ low_wage, data = pa)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -39.652  -5.465   1.441   5.988  24.785
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -3.848      2.340  -1.644    0.105
## low_wage       2.813      2.899   0.970    0.336
##
## Residual standard error: 11.22 on 64 degrees of freedom
## Multiple R-squared:  0.01449,    Adjusted R-squared:  -0.0009036
## F-statistic: 0.9413 on 1 and 64 DF,  p-value: 0.3356
```

```r
# ii)
df$NJ <- df$state
df$low_wage <- ifelse(df$wage_st < 5, 1, 0)

reg_f <- lm(dw ~ NJ + low_wage + NJ * low_wage, data = df)
summary(reg_f)
```

```
##
## Call:
## lm(formula = dw ~ NJ + low_wage + NJ * low_wage, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.58837 -0.16178  0.01522  0.18822  1.91163
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.26522    0.05156  -5.144 4.51e-07 ***
## NJ           0.26113    0.05987   4.361 1.71e-05 ***
## low_wage     0.35359    0.06388   5.536 6.13e-08 ***
## NJ:low_wage  0.26228    0.07270   3.608 0.000354 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2473 on 347 degrees of freedom
## Multiple R-squared:  0.6207, Adjusted R-squared:  0.6174
## F-statistic: 189.3 on 3 and 347 DF,  p-value: < 2.2e-16
```

```r
reg_g <- lm(dfte ~ NJ + low_wage + NJ * low_wage, data = df)
summary(reg_g)
```

```
##
## Call:
## lm(formula = dfte ~ NJ + low_wage + NJ * low_wage, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -39.652  -3.551   0.035   4.750  24.949
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -3.8478     1.7635  -2.182   0.0298 *
## NJ            1.5978     2.0478   0.780   0.4358
## low_wage      2.8129     2.1848   1.288   0.1988
## NJ:low_wage   0.4884     2.4867   0.196   0.8444
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.457 on 347 degrees of freedom
## Multiple R-squared:  0.03706,    Adjusted R-squared:  0.02874
## F-statistic: 4.452 on 3 and 347 DF,  p-value: 0.004376
```

#iii)

This comparison works like a kind a placebo test to see whether our method is actually picking up the real effect of the minimum-wage increase. Since PA didn't raise its minimum wage, we shouldn't see a big jump in wages for low-wage restaurants. And that's basically what we find: the low-wage effect in PA is small, while in NJ it's much larger (twice as large) and driven by the policy. This gives us confidence that the DiD approach is working properly and isn't just capturing general wage changes. For employment, the differences between NJ and PA aren't statistically strong, which fits with the earlier results showing little evidence of job losses. Overall, the PA results support the idea that the methodology is doing what it's supposed to do.