technische universität dortmund

Fakultät Wirtschaftswissenschaften

Problem Set 2

Labour Economics, Winter Semester 2025/26

Submit by Sunday, 30 November, 22:45h ***on Moodle!***

Learning objectives

- Practical implementation of difference-in-differences analysis. Equivalence with implementations in different regression specifications.
- Presentation of identification assumptions and discussion regarding their plausibility.
- Interpretation of empirical results.

Tasks

Download the dataset called `minwage.dta`. It contains data collected by David Card and Alan Krueger on fast food restaurants in New Jersey (NJ) and eastern Pennsylvania (PA) during two interview waves in March and November/December of 1992. On April 1, 1992 New Jersey raised its minimum wage from $4.25 to $5.05. The minimum in Pennsylvania remained at the federal level of $4.25. Use this data to analyze the impact of the minimum wage increase in New Jersey on employment in the fast food industry.

Throughout, variable names with a trailing "2" refer to the second (Nov./Dec.) wave of the data, the same names without any number refer to the corresponding variable from the March wave. `fte` and `fte2` are full time equivalent employment, it is the sum of the number of full time employees and one half the number of part time employees, excluding managers; `dfte` refers to the change in full time equivalent employment between the second and first interview (`fte2 - fte`); `dw` refers to the change in the starting wage between the second and first interview, and `sample` is a dummy variable which is 1 if both wage and employment data are available in both the first and second interview wave, and 0 otherwise. I want you to do the following analysis for the part of the data with sample equal to 1. If you don't specify this, R will make calculations with the full set of available observations for each variable, so you may not be comparing the same set of restaurants between March and November, or you may compare wages and employment for different restaurants.

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

(i) Calculate the difference in the average wages between the second and first interviews.

Solution

The difference between the average starting wage of the second compared to first interview wave is 0.4691 in New Jersey and -0.0348 in Pennsylvania.

State	First Interview	Second Interview	Difference
New Jersey	4.6130	5.0821	0.4691
Pennsylvania	4.6536	4.6188	-0.0348
Difference	-0.0406	0.4633	0.5039

(ii) Now calculate the difference between NJ and PA of the time differences just obtained.

Solution

The difference of the time differences between NJ and PA is 0.5039.

(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?

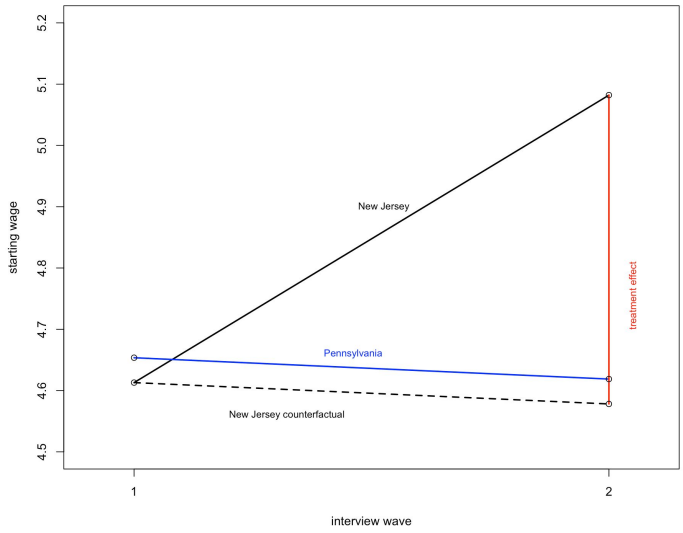
Solution

This difference can be interpreted as the effect of the treatment on the starting wage if one assumes that there would be a parallel trend in NJ and PA without a treatment.

(iv) Interpret your finding.

Solution

If one assumes parallel trends, this means that the starting wage increased by about 0.5 due to the minimum wage raise which is approximately 11% of the average starting wage at the first interview wave. Since the minimum wage raise increases the minimum wage by approximately 19%, this shows that there is no equivalence between a minimum wage raise and a starting wage raise. This is likely because not all starting wages are bound at the minimum wage and some restaurants pay higher starting wages, which is visible also in the averages of the table.

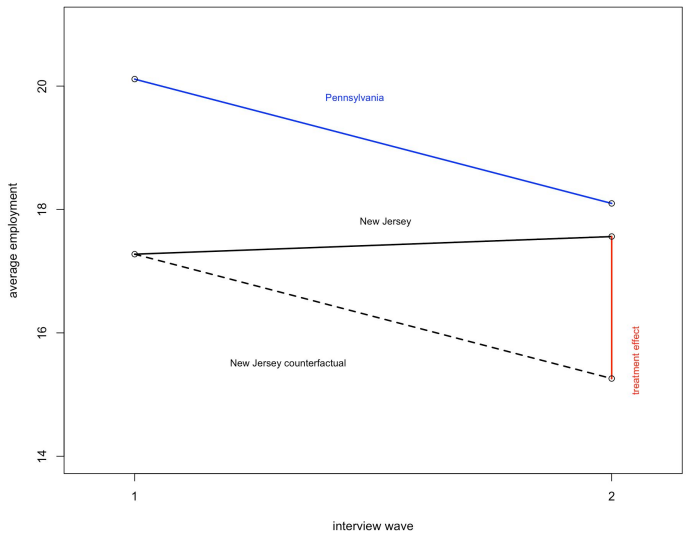


(b) Repeat the same exercise as in (a) for full time equivalent employment. What is the impact of the minimum wage increase on relative employment in NJ restaurants?

Solution

The relative employment in NJ increased by 2.3020 due to the minimum wage increase. This is approximately 13% higher than the average employment in March. The number of jobs increased although the wage was higher and therefore it seems that worries about axing jobs because of a higher minimum wage might be unfounded.

State	First Interview	Second Interview	Difference
New Jersey	17.2754	17.5623	0.2869
Pennsylvania	20.1136	18.0985	-2.0151
Difference	-2.8382	-0.5362	2.3020



(c) Difference-in-difference estimates can also be calculated from the regression

$$Y_{ist} = \beta TREAT_{is} + \gamma POST_t + \delta_{rDD}(TREAT_{is} * POST_t) + e_{ist},$$

where Y_{ist} is employment in restaurant i in state s and period t, $TREAT_{is}$ is an indicator for the treatment area (NJ or low wage restaurants in NJ), $POST_t$ is an indicator for the treatment period (Nov/Dec) and $TREAT_{is} * POST_t$, is the interaction of these two dummies. Note that this regression uses the data for individual restaurants i and we leave the averaging to the regression.

(i) Write the equation separately for March and Nov/Dec and show that the DD model for two periods ($t = 1, 2$) can be estimated as

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

Solution

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

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

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

where in the last line γ is the regression constant reflecting general (i.e., PA) changes of wages or employment. Coefficient of interest is δ_{rDD}, which reflects the differential changes in NY. Last is the new regression residual $e_{is2} - e_{is1}$.

(ii) What are the regression DD estimates on wages and employment using this regression? How do they compare to the results you found in (a) and (b)?

Solution

The estimates are exactly the result of a) and b).

wages

```
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)  0.46916    0.02063   22.74   <2e-16 ***
statePA     -0.50401    0.04757  -10.60   <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
```

employment

```
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
```

(iii) The regression allows you to control for other factors. Repeat the regressions, entering a dummy variable for whether the restaurant is company owned (`co_owned`, as compared to franchised) and three dummy variables for three of the four chains in the dataset (Burger King, KFC, Roy Rogers, and Wendy's; you will have to construct the dummies from the variable `chain` or use `factor(chain)`.

Solution

The coefficient on the treatment in the wage-regression changes just very little and is now 0.5037.The coefficient on the treatment in the employment-regression stays as well very similar and is now 2.2973. We would not have expected the results to change because we assumed that the different chains and the fraction of company owned restaurants are quite balanced across the two states and are therefore not correlated with the treatment variable.

wages

```
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
```

employment

```
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
```

(iv) Do your results change when you enter restaurant specific covariates? Would you have expected the results to change? Explain why or why not.

Solution

The coefficient on the treatment in the wage-regression changes just very little and is now 0.5037.The coefficient on the treatment in the employment-regression stays as well very similar and is now 2.2973. We would not have expected the results to change because we assumed that the different chains and the fraction of company owned restaurants are quite balanced across the two states and are therefore not correlated with the treatment variable.

(d) An alternative to comparing NJ and PA restaurants is comparing restaurants within NJ which have high and low wages before the minimum wage increase. Restrict your sample to restaurants in NJ.

(i) Would you expect the DD assumptions to be satisfied more easily for the within NJ comparison than for the NJ - PA comparison?

Solution

We could think that the DD assumptions would be satisfied more easily for the within NJ comparison than for the NJ - PA comparison because there might be some different time trends in NJ and PA. For instance there might be laws affecting restaurants becoming effective in NJ but not in PA. If one only looks at restaurants in one of the states, those problems do not remain.

On the other hand, at least in terms of wages, one could expect that there may be regression to the mean. That is, some of the low wages in low-wage restaurants are temporary, and some of the high wages in high-wage restaurants are too, such that over time (currently) low- and high-wage restaurants' wages move toward one another.

(ii) Construct a variable for those restaurants paying starting wages of less than $5.00 before the minimum wage increase. Use the regression to obtain a DD estimate of the employment and wage effects of the minimum wage increase. What is the relative impact of the minimum wage on starting wages and employment within NJ?

Solution

The relative impact of the minimum wage on starting wages within NJ is 0.6159 and 3.3014 on employment.

wages

```
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
```

employment

```
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
```

(iii) How do your witin NJ estimates compare to those obtained in part (c) for the NJ - PA comparison?

Solution

The results are 22% (for wages) and 43% (for employment), i.e., bigger than the ones in c).

(e) You can create a variable for those restaurants paying starting wages of less than $5.00 in PA in the initial period. There is no minimum wage forcing those restaurants to pay more in the second period but there may be general wage growth.

(i) Now run a regression of changes in employment and wages just for PA using this new variable for low paying restaurants in PA. How do your results differ from those just for NJ?

Solution

The wage increases by 0.3536 and the employment increases by 2.813 for restaurants which had a low wage in March. So even in a state that had no minimum wage increase, wages and employment of low-wage restaurants increased compared to high-wage

restaurant. This may indicate some regression-to-the-mean effects hypothesized in d(i).

wages

```
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
```

employment

```
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
```

(ii) Carry out a statistical test of the hypothesis that the coefficient on the low wage dummy is the same in NJ and in PA.

Solution

If one runs a regression of the difference in wages on the state dummy, the dummy for restaurants that paid less than 5$ in March and an interaction term of those two variables, one can see that the t-statistic of the interaction term is 3.608. Therefore, one can reject that the coefficients are the same in the two states. If one runs the same regression but this time with the difference in employment as the dependent variable, the picture looks different as the t-statistic of the interaction term is 0.196. Hence, one cannot reject the hypothesis of equality in low-wage restaurants' relative wage increases between the two states.

wages

```
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
```

employment

```
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) Why is this a check on how well the methodology is doing in uncovering effects of the minimum wage increase? What do you conclude?

Solution

This is a check on how well the methodology is doing in uncovering effects of the minimum wage increase because there should be a different effect in PA and in NJ as only the latter implemented a higher minimum wage. We could think of PA as providing a "Placebo Test", since any "effects on" PA cannot be due to the treatment.

Indeed, there were some increases of employment and wages in PA's low-wage restaurants, too. Still, there is a statistically significant stronger increase of low-wage restaurants' wages in NJ than in PA, which is consistent with the direct effect of the minimum wage.

Results for employment are smaller than in the DD study of averages across the state from question b) and not statistically significant. This indicates that results in b) may be a bit too high and, indeed, later studies (also by Card and Krueger themselves) found essentially a zero impact of the minimum wage on employment.

Notes: You can work in teams of 1–3 students. Please upload your code as well as a pdf-file with discussions on what you found in the data in response to the tasks above. It should be clear which lines of code and answers in the `.pdf` refer to which question. If you work in a team, each member has to upload the group's solution and note whom they worked with.