

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.
 - (ii) Now calculate the difference between NJ and PA of the time differences just obtained.
 - (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?
 - (iv) Interpret your finding.
- (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?
- (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

$$Y_{is2} - Y_{is1} = \gamma + \delta_{rDD} TREAT_{is} + e_{is2} - e_{is1}$$
 - (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)?
 - (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)).
 - (iv) Do your results change when you enter restaurant specific covariates? Would you have expected the results to change? Explain why or why not.
- (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?
 - (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?

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

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.