Problem Set 4

Your id here

2025-03-10

1 Difference-in-difference

(From Hazlett)

In this problem, you will analyze data from the Card and Krueger paper discussed in lecture.

> Card, D. and A. B. Krueger (1994), "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania," *American Economic Review*, vol. 84, 772-793.

Card and Krueger are interested in estimating the impact of minimum wage on teenage employment. Conventional economic wisdom states that raises in minimum wages hurt employment, especially teenage employment, which often takes wages that will be affected by minimum wage law. However, empirical analysis has failed to find evidence of employment responses to raises in minimum wages. In 1992, New Jersey's minimum wage increased from $4.25 to $5.05 while the minimum wage in Pennsylvania remained at $4.25. The authors used data on employment at fast-food establishments in New Jersey and Pennsylvania before and after the increase in the minimum wage to measure the impact of the increase in minimum wage on teenage employment. Download and import the dataset, `card_krueger.RData`, from the course website.

The dataset contains the following variables, with one row corresponding to a single fast-food restaurant:

- `state`: New Jersey or Pennsylvania
- `chain`: the fast-food chain to which the restaurant belongs
- `wage_pre`: starting wage in February 1992, in dollars per hour
- `wage_post`: starting wage in November 1992, in dollars per hour
- `emp_pre`: employment in February 1992, in number of full-time equivalent (FTE) employees

- `emp_post`: employment in November 1992, in number of full-time equivalent (FTE) employees
- `closed`: whether the store was closed in November 1992

1. Assume that the fast-food restaurants surveyed by Card and Krueger represent a random sample from a larger population of all fast-food restaurants in New Jersey and eastern Pennsylvania. Consider the estimands in table **??**, which correspond to the mean level of full-time equivalent (FTE) employment for population subgroups (restaurants within a given state-time). For example, $\alpha = \mathbb{E}[Y_i|T = \text{February}, \text{state} = \text{NJ}]$.

	February	November
New Jersey	α	β
Pennsylvania	γ	δ

Table 1: Average population FTE employment, by state and time

(i) Define the difference-in-differences (DID) estimand in terms of these values, and concisely explain what this represents.

(ii) Consider the eight potential quantities $\{\alpha(0), \alpha(1), \beta(0)...\}$. Let these represent the mean *potential* level of FTE employment levels that *would have realized* if the minimum wage had been raised in each state at each time. For example, $\alpha(0) = \mathbb{E}[Y_i(D = 0)|T = \text{February}, \text{state} = \text{NJ}]$. Define the causal quantity of interest, the ATT, in terms of these potential outcomes. Describe which of these are observed.

(iii) Discuss the identification assumption required for DID to yield a causal effect: (a) state it in general terms and then as applied to this problem; (b) explain in simple and intuitive terms why this is a useful assumption; (c) finally, tell us whether you think it is a good assumption here, and if not, what violations you imagine are problematic.

(iv) Making the required identification assumption from above, show that the DID estimand (part (i)) can identify the ATT (part (ii)). Show all your work and note where the identification assumption comes in.

2. Estimate all quantities in Table 1 and write them in a similar table.
3. Calculate your DID estimate (with a standard error and p-value) and interpret the result.

2 Sensitivity analysis

For this section, we are going to reanalyze the data from LaLonde's study that we used in problem set 2. Technical details for sensitivity analysis can be found in Rosenbaum's books *Design of Observational Studies*, and *Observational studies*. This guide can also be helpful.

You may want to check the following `R` packages: `rbounds`, `sensitivitymv`, `sensitivitymw`, and `sensitivityfull`. Consult the documentation and decide which method would be appropriate for the following questions.

2.1 Rosenbaum bounds

In problem set 2, you were asked to estimate the treatment effect of a training program on future income using observational data and a matching method. For this exercise, you're asked to conduct an appropriate sensitivity analysis.

(a) Reproduce the results from problem set 2, section 2, using pair matching on the propensity score. Show the estimated treatment effect.

(b) Reproduce the results from problem set 2, section 2, using full matching on the propensity score. Show the estimated treatment effect.

(c) Compare the results obtained in sections (a) and (b). How similar are they? Do you have any reason to prefer one model over the other?

(d) Explain what Rosenbaum bounds are, which sensitivity parameters are used, and what is the interpretation of the Γ magnitude.

(e) Conduct a sensitivity analysis for the results in (a) and (b) using the appropriate version of Rosenbaum bounds. Explain your results.

2.2 Cinelli and Hazlett

Before attempting these questions, make sure to read Cinelli and Hazlett (2020), and the introduction to `sensemakr`

Now, we will try to replicate the results in problem set 2, section 2, using regression. We will assume the effect is homogeneous so it can be captured in a single parameter.

(a) Run a multiple regression analogue of the matched analysis in the previous questions. Include the raw covariates as controls, and report the point estimate and confidence intervals for the regression coefficient of the treatment.

(b) Based on Cinelli and Hazlett (2020), explain the sensitivity analysis they propose. Pay special attention at explaining the *Robustness value* (R-value) and the $R^2_{Y\sim D|X}$ parameter.

(c) Now implement and report results from the sensitivity analysis following Cinelli and Hazlett (2020). How would you read the results? Would this results increase or decrease your confidence in the original results?

(d) Select and provide a justification to choose one of the covariates as your benchmark. Report the results graphically, and explain if these results would increase or decrease your confidence in the original results.