Econometrics 2 – M1 PPD, Paris School of Economics – 2024-2025

Problem set no 1

Exercise: Rubin model and Roy model (and conditional expectation manipulation)

We consider a labor market training program that is offered to workers. This training may increase a given individual's wage from w_0 to w_1. Attending the program has a cost c. We assume that each agent knows his or her wage outcome, with and without training, with certainty. We also assume that both counterfactual wages in the population are heterogeneous, but the cost c is common to everyone. We write:

$$w_1 = w_0 + \delta$$

where δ and w_0 are heterogeneous in the population. We impose $\delta > 0$ and, at the beginning of this exercise, we assume that δ and w_0 are mean-independent ($E(w_0|\delta) = E(w_0)$). We will wave this assumption later.

1. What is the treatment impact of a given individual i? What is the average treatment impact in the population?

2. Write the decision model of attending the training or not (called a Roy model).

3. Based on that decision rule, people sort themselves into the training program. We observe average wages of treated and untreated. What does each of those averages measure.

4. If you compare average wages of the treated and untreated, what parameter do you estimate? Interpret that parameter.

   Now assume that a randomized control trial is run. Three groups are randomly formed in the population. The first group ($Z = 0$) faces the normal rules (cost c). The second group ($Z = 1$) cannot access the training at all. The third group ($Z = 2$) is offered a subsidy $s < c$ **if** they choose to attend the training.

5. Compute the value of the average wage in each of these populations.

6. Note that you can estimate the proportion of trainees in each random group. Given this, show that comparing group $Z = 1$ with group $Z = 0$ identifies the same parameter as in the absence of an experiment.

7. Explain why the mean-independence between δ and w_0 ensures that the naive wage comparison can estimate a treatment parameter without the experiment. Why can't we obtain ATE, though?

8. Show that is it possible to identify the impact of the training on the population that is induced to participate by the subsidy s (and would not participate otherwise), using group $Z = 2$ and group $Z = 0$. Define that parameter formally and explain how can you compute it?

9. What can you estimate if $s = c$?

   Now assume that δ and w_0 are correlated and assume that $w_0 = a+\rho\delta+\varepsilon$ where a and ρ are fixed parameters and ε is a residual uncorrelated with δ with mean zero.

10. Show that the naive comparison of the wages of treated and untreated absent an experiment would no longer identify a treatment parameter. Discuss the sign of the bias depending on ρ.

11. Show that comparing group $Z = 1$ with group $Z = 0$ identifies the same treatment parameter as before.