Empirical Methods for Policy Evaluation

TSE MRes Program 2024-2025

Takehome for Part 2

Directions. You can work in pairs. Send your answers in a pdf file along with the code you used with whatever statistical software you want to matteo.bobba@tse-fr.eu by **January 15, 2025**. Do not submit your work after the deadline.

For this exercise you will use the 1998 evaluation sample of the *Progresa* CCT program in Mexico, which is made available in this folder.

1. The assignment of the program (variable `progresa`) was randomized at the village-level. What are the identification assumptions? Provide supportive evidence for those assumptions.

2. The randomization was not stratified. What would be the benefit of stratification in this context? Propose some variables in the dataset that you would use as strata. Explain in some detail how the randomization would be performed in such a case.

3. Using the variables that you chose as strata in point 2, explain the difference between re-randomization and stratified randomization, as well as the pros and cons of each approach.

4. Estimate the effect of the *Progresa* program on school enrollment by performing a series of OLS regressions by indigenous status and gender. Report in a table the estimated coefficients along with the asymptotic p-values. Do your results vary if you include covariates in the regression models? Discuss.

5. Test both the Sharp Null Hypothesis (randomization-c) and the Average Null Hypothesis (randomization-t) of no effect of the program by indigenous status and gender, and compute the Fisher's exact p-values. Compare and interpret your results between these two approaches with those of point 4.

6. Assume households solve the following problem:

$$\begin{aligned} \max U(C, s; \epsilon) &= C + \alpha s + \beta C s + \epsilon s, \epsilon \sim N(0, \sigma_\epsilon^2) \\ C &= y + w(1 - s) + \tau s, \end{aligned} \tag{1}$$

   where s is children's school enrollment (variable `enroll`), y is household earnings (variable `household wage`), and w is the child wage in the local labor market, which you can approximate as the minimum value of the village-level distribution of adults' earnings.

Use a Probit model to recover the model parameters. Perform the estimation using only the control sample.

7. Predict the counterfactual enrollment under the subsidy scheme ($\tau \approx \frac{1}{4}\bar{y}$). Compute the difference in mean enrollment by indigenous status and gender. Compare the results with those found on point 3. What do you conclude?

8. Consider this alternative model, which we saw in class:

$$U^s - U^w = \alpha + (\beta^s - \beta^w)Y + \theta^s\tau - \theta^w w \tag{2}$$

What are the fundamental differences of this model when compared to the model depicted by equation (1)?

9. Use a Probit model to recover the underlying parameters. Perform the estimation with the full sample (both treated and control villages). Compute the difference in mean enrollment by indigenous status and gender. Compare your results with those of points 3 and 6. Discuss.

10. Plot the relationship between educational attainment (variable `grade enr 1997`) and average enrollment rates pre-program (variable `enroll 1997`). Based on this evidence, what would be a more effective subsidy scheme? Simulate both estimated models (1) and (2) and quantify the average gains in enrollment by grade from your counterfactual subsidy.