Problem Set 1 - The Effect of Public Employment Centers

Professor: Chiara Monfardini
Teaching Assistant: Giacomo Opocher

- This Problem Set (PS) was sent on November 22, 2023.
- Due date: **November 29, 11.59 pm**.
- Submit your PS on the web page of the course, in the "Problem Set 1" dedicated section, no later than 11.59 p.m. PSs sent after the deadline will not be graded (no exceptions).
- You must attach a single zip file containing: (i) a pdf answer sheet; (ii) the Stata *do file*. The pdf should be no longer than 5 pages (w/out tables and figures).
- Each table and graph in the pdf file should be fully reproducible. By simply running the *do file* I should be able to reproduce the exact table or graph you are showing in the pdf, including title, variable names, and numbers. In the *do file* you have to signal clearly which chunk of code reproduces which table or which chart. I will assign up to 2 bonus points if your code is readable, replicable, and efficient.
- Name the zip file as `surname1_surname2_surname3.zip`; remember to write the name, surname, and student number of each student in the answer sheet.
- Each Problem Set is graded from 1 to 30. You will know the grade only **after** you took the exam. The average grade over the 3 PSs will make up the 40% of the final grade.
- Please follow carefully the instructions detailed above. Any misconduct will negatively impact the grading of the PS.

The national government of *Metricsland* contacts you to evaluate the impact of a policy the Ministry of Labor put in place in 2023 to increase the employment rate.

This policy involves building new public employment centers around the country. These centers are places where citizens can receive guidance during their job search process.

You observe a random sample of 2400 neighborhoods in the three years before (2020, 2021, 2022) and after (2023, 2024, 2025) the introduction of the policy. Your final goal is to report to the Ministry whether the policy was effective and provide some indications on how to improve it.

Part 1: Data Exploration

1. Load `MLand_jobpol.dta` into Stata and define it as a panel dataset[1] using the `xtset` command. What are the main variables?

2. What are the characteristics of the sample in 2022?
   Do some of these characteristics change over time? Choose one variable and plot its sample distribution in 2022 and 2024.

3. Focus on the main variable of interest: `unempl`. Use the command `xtsum` to compute and interpret the overall, within, and between variation.
   Does the average `unempl` increase or decrease over time?

4. Focus on the center presence indicator: `center`. Considering only data from 2022, check whether neighborhoods where an employment center was implemented or not differ, on average, along any of the observable variables.

Part 2: Impact Evaluation

1. Consider the data from 2023, 24, and 25. Using (simple[2]) OLS regression, estimate the effect of the presence of an employment center on the unemployment rate, for each post-policy year separately, and represent it graphically. In the pdf, write down the model you estimated. Given the information you have, can you claim that this estimator is unbiased in the causal effect of interest? Why?

2. Can you use a Diff-in-Diff strategy to identify the causal effect of interest? If so, under which critical assumption? Is it plausible that this assumption holds in our setting? Does this approach cancel out the source of bias of the previous estimator?

3. Write down the DiD estimator by means of a double difference between averages. Estimate it manually.

4. Write down the DiD estimator employing a first difference regression model. Estimate it with the `reg` command.

5. Write down the DiD estimator as a fixed effect regression model. Estimate it with the `xtreg` command.

6. *Optional:* Represent graphically the effects you estimated in the previous points. Consider the magnitude of the effect on the Y axes and year on the X axes.

7. Comment on your results: was the intervention effective? Would you suggest the Minister to scale it up?

[1] Be careful: data are in *wide* format: you first need to reshape to *long* format.
[2] Meaning, without control variables.

8. *Bonus Question*: Would you still need panel data to evaluate the impact of the information campaign if the neighborhoods where a center was located were randomly selected? Why?

Codebook

- `id_zip`: unique identifier of the neighborhood.
- `city`: city code.
- `region`: region code.
- `unempl`: share of unemployed.
- `avg_educ`: average years of education.
- `econ_conn`: share of individuals in the bottom decile of income connected with the top decile.
- `share_imm`: share of residents born outside the country.
- `size`: number of inhabitants in hundreds.
- `crimes_pc`: number of crimes per capita.
- `center`: treatment indicator (=1 if an employment center was located there in 2023, =0 otherwise).