Problem Set 2 - A.I. and Political Knowledge

Professor: Chiara Monfardini
Teaching Assistant: Giacomo Opocher

- This Problem Set (PS2) was sent on November 29, 2024.
- Due date: **December 6, 11.59 pm**.
- Submit your PS on the web page of the course, in the "Problem Set 2" 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 *log file*; (iii) the Stata *do file*. The pdf should be no longer than 5 pages (w/out tables).
- 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.

A well-informed electorate is essential for the proper functioning of a democracy. Political knowledge serves as the cornerstone for voters to form opinions and express their preferences effectively. However, despite the widespread availability of information in today's world, voters are, on average, poorly informed. One way to explain this paradox of abundant information (input) and limited political knowledge (output) is to recognize that while information is easily accessible, absorbing and processing it can be both time-consuming and mentally demanding.

You are a team of talented young applied microeconomists who have developed an innovative mobile application powered by advanced machine learning algorithms. The app is designed to answer any politics-related question by gathering information from a vast array of news sources and official documents, ensuring neutrality and avoiding any bias toward specific political parties. It is very simple to use: you can think about it as a *Chat GPT*

that helps you understand politically relevant information. It is also very successful: at the first launch, a lot of people use it and are satisfied with it.

Recognizing its potential, you approach a team of researchers at the Italian Ministry of the Interior to propose the large-scale public distribution of the app. The Minister expresses enthusiasm for the project but requests that you provide empirical evidence demonstrating that using the app significantly improves voters' political knowledge before moving forward.

To fulfill this request, you run a field experiment with the support of the Ministry. Considering a random sample of 100,000 voters, you **randomly** select half of the sample to be **encouraged** directly by the Ministry to use the app ($\texttt{enc}_i = 1$). Notice that, since it is already on the market, also individuals with $\texttt{enc}_i = 0$ can use it, although **they are not encouraged** to.

After 2 months, you receive from the Ministry the data collected so far: `pol_kn.dta`. About the whole set of participants (both $\texttt{enc}_i = 0$ and $\texttt{enc}_i = 1$), you know the age, gender, education, and political party membership. Importantly you have a survey-based measure of each participant's political knowledge. Finally, for the whole set of respondents, you know the time of use of the application in the last 2 months.

Part 1: Data Exploration

1. What are the main characteristics of the sample? Provide a table that displays the mean, standard deviation, minimum, and maximum value of all the variables available. Comment on it.

2. Are the two experimental groups balanced? Meaning: is the average voter in the group $\texttt{enc}_i = 0$ similar to the average voter in the group $\texttt{enc}_i = 1$ over the characteristics you can observe? Make a single table that shows the average of each variable for each group. You can use this table as a reference. This command might be useful to replicate it easily, but feel free to use another one.

3. Is receiving the encouragement exogenous to political knowledge before the experiment? What about the time each respondent spends inside the app?

Part 2: Impact Evaluation

1. Assume you **did not** run the experiment and therefore you can observe only the usage time of the app and the political knowledge. Estimate with OLS the following regression:

$$\texttt{pol\_kn}_i = \alpha_0 + \alpha_1 \texttt{app}_i + \delta X_i + \mu_i \tag{1}$$

   Where X_i is a vector of controls of your choice and μ_i is the error term. Represent the relation of interest in a scatterplot: what feature of the graph is $\hat{\alpha}_1$? Can this model deliver an unbiased estimator of α_1: the causal effect of one more hour of usage of the application on political knowledge? Why?

2. Use an Instrumental Variable strategy to estimate the effect of one additional hour of use of the application (`app`) on political knowledge (`pol_kn`). Which variable is a good candidate as an instrument? Write down the first and the second stages.

3. Estimate the model using the command `ivreg`. What is the effect of one hour more of use on the respondents' political knowledge? Is it statistically significant?

4. Is the instrument you chose a valid instrument? Provide evidence for its relevance and argue about its exogeneity.

5. Estimate the two stages separately using the command `reg`. Is the second stage coefficient different from the one in the previous point? And the standard error? Which estimation technique is better? Why?

Codebook

- `id`: unique identifier.
- `age`: age in years.
- `educ`: years of education.
- `fem`: female indicator.
- `party`: political party membership (=1 if *left* party, =0 if *right* party).
- `pol_kn0`: level of political knowledge (0-300) before the experiment.
- `pol_kn`: level of political knowledge (0-300) after the experiment.
- `app`: total time in hours on the app since the starting date of the experiment.
- `enc`: dummy equal to 1 if the respondent received the encouragement.
- `empl`: =1 if employed.