Assignment 2

HS 649 Public Policy and its Impact Evaluation

Semester-II 2022-2023

Due Date/time: 11.59PM on 12th March, 2023

Submission Instructions:

- *You are required to submit the following documents as part of your assignment:*
  1. *A do-file (it has to be commented out properly, otherwise points will be deducted).*
  2. *A log file that shows the output from the do-file*
  3. *A pdf or word doc that has all the results along with descriptions where asked.*
  4. *Final .dta that will be produced through this exercise.*
- *The responses to most of the questions are Stata commands. Wherever you are asked to provide a description, make sure that you do that in your own words. Copying will not be tolerated.*
- *Make sure that your do-files are properly commented out.*
- *Responses submitted after due date/time will not be accepted.* **If late by 1-5 minutes, then 50% points will be deducted (without negotiation). If late by more than 5-10 minutes, it will be considered as fail.** *So, plan your submission in advance of the deadline.*

Objective of this assignment:

In this assignment, we are going to learn running regressions in Stata, exporting regression outputs and interpreting them, among other things. We will analyze if electrification reduces gender differences. For this we will rely on IHDS-2. You will be referring this paper for your exercises (also attached).

- Note that the paper has used 2005 and 2012 waves of IHDS, but we will be using only 2012 because we haven't learnt dealing with panel data yet.

Please keep the following in mind while producing and interpreting your results:

- You will indeed want to refer (or read) certain sections of the paper, if not all of it. My suggestion would be the following sections: the Introduction (particularly where they discuss data and results), Data and Empirical work, and the Results section.
- Since we will be using only IHDS-2 in this assignment, you should also be concerned only with the results that rely on IHDS-2.

Exercise/Questions: [Total Marks: 50 points]

IMPORTANT NOTE: For each of the results below, in addition to showing your codes in do-file for producing them, you are also required to use codes to export the results. DO NOT simply copy-paste the results from Stata window to Word.

1. Prepare a data set that would produce the results from the paper. The output here will be your codes that prepares such a data-set. (Hint: see the third bullet point in the 'suggestions' section below) [10 points]

2. Reproduce the results of Table 1, only for the column 2012. [5 points]

3. Reproduce the results of Table 2, but only for 2012. [5 points]

4. Reproduce the results of Table 3, but only for 2012. [5 points]

5. Finally, reproduce the first three columns of Table 5, but using the following specification only with IHDS-2 data. [10 points]

$$Y_i = \beta EH_i + X'_i\delta + \alpha D_j + \lambda P_j + \epsilon_i$$

Clearly mention what control variables you are able to use, and what not.

6. Based on part Q.5, [15 points]

    a. Interpret the results. What is the effect of electricity on the outcomes? Explain the meaning both of the point estimate and of the statistical significance. (No more than 150 words)

    b. Briefly justify your choice of standard errors. Would you want to cluster your standard errors? What does clustering do? (No more than 100 words)

    c. Briefly discuss the causality claim based on your results only. Talk about the potential bias in your coefficients. If you were restricted to only one wave of IHDS data, what other controls you would want to include? (No more than 100 words)

Some suggestions to execute this exercise:
- First try and understand the variable mentioned in the paper, including their modifications.
- Next, go to the IHDS questionnaire to find the variables, and then identify the correct data files that has those variables.
- You will notice that electricity related variables are in the household data file, and the employment related variables are in individual data file.
    - Clean both the data files (i.e., keep only those variables that you will be using) and then merge them using unique HH IDs.

- Basically, create your main analysis data with only those modified variables that you will eventually need.
- Finally, your results have to be as close to those in the paper as possible. It is fine if they are not exactly the same. Grading will also be relative and will depend how close the results are for anyone in the class.
    - Similar to the above point, not all variables are easy to comprehend at times, so as long as you get most of the variables, you are good. But again, grading will be relative to the closest results in the class.