Day 4 - Part 1: Public Policy Design and Evaluation

Instructors: Cristián Jara[1], Javiera Petersen[2] and Raimundo Smith[3]

1 Impact Evaluation

The goal of section is to present applications of experimental and quasi-experimental methods that are frequently used in public policy evaluation. Particularly, we present applications on: Randomized Control Trials, Matching Estimators, Difference-in-Difference and panel regression analysis.

Random Control Trials (RCT)

Exercise 1

To motivate this section, we are going to work with you are asked to read: Nyqvist, M. B. and S. Jayachandran (2017): "Mothers Care More,But Fathers Decide: Educating Parents about Child Health in Uganda", *American Economic Review: Papers and Proceedings.*. This paper investigates intra-household decision making in Uganda and, specifically, the impact of nutrition classes provided to parents of young children with knowledge to improve their children's health.
The exercises you must solve are:

1. Replicate balance table (Table 1 of the paper).
2. Replicate table 2 of the paper. Particularly, columns 2 and 3, relate to impacts on health knowledge and health behavior.[4]

Exercise 2

In the second exercise, you are asked to read: Blattman, Christopher, and Stefan Dercon. 2018. "The Impacts of Industrial and Entrepreneurial Work on Income and Health: Experimental Evidence from Ethiopia." American Economic Journal: Applied Economics, 10 (3): 1-38. This paper evaluates the impact of individual investments on skills and training in an experimental setting.
Now you are asked to replicate to basic results from this paper:

1. Replicate balance table
2. Run ITT regression for the following income variables: *incomefamilyindex_e, totalprofit7_av_p99_e, totalprofit_h_av_p99_e, totalprofit_sqdev_e, consdur_p99_std_e, cons_month_p99_e, proddur_p99_std_e.* Use robust standard errors and clusters using the variable *appid*.

[1]National Institute of Statistics Chile: cristianjara21@gmail.com
[2]University College London (UCL): javierapertersenm@gmail.com
[3]World Bank: raimundo.smith.m@gmail.com
[4]Pay attention to the notes on the table. Specifically, for the covariates used in the paper.

Matching and Propensity Score (PSM)

Exercise 1

For this exercise, we are going to use *4_ matching_ 1.dta* data, which is a subset of the data used in Dehejia, R. H., and Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. Journal of the American statistical Association, 94(448), 1053-1062. The idea is introduce students to the basic steps of matching estimation, and present the set of tools that Stata provide for this purpose. In this context, the, students are asked to do the following calculations:

1. Compute descriptive statistics using the entire data and by treatment.
2. Compute the propensity score using the `pscore` command.
3. Analyze the distribution of the common support using `kdensity` or using `psgraph` command.
4. Compute Nearest Neighbor and Radius matching using the `attnd` command.
5. Repeat steps 1 to 4 but using the command `psmatch2`.
6. Test the balance of the common support after the matching using the command `pstest`.

Exercise 2

In this exercise, students are going to work with the *Social Protection Survey* data of 2006 from Chile. The idea is to model the relation between income and marital status using a matching estimator. For this purpose, you will have to use income provided by the "labor history" data from the survey, and then merge it with socio-economic information from respondents. The steps are the following:

1. Use *day_ 4_ matching_ historialaboral_ 06.dta* data and keep last job history for each participant. Then merge this data with socioeconomic data using *day_ 4_ matching_ entrevistado_ 06.dta* data.
2. Generate variables for analysis: *Gender, income and married.*
3. Compute the propensity score using a probit model.
4. Check the common support using graphical analysis.
5. Compute ATT using "1-1" matching and nearest neighbour using `psmatch2` command.

Difference in Difference (DID) and Panel regression

Exercise 1

To illustrate the basics of panel regression analysis, we are going to use *day_ 4_ panel_ data_ guns.dta* data, which is a subset of the data used in Ayres, I., and Donohue III, J. J. (2003). Shooting down the "More Guns, Less Crime" Hypothesis. Stanford Law Review, 1193-1312. Particularly, this data is a balanced panel on 50 US states, plus the District of Columbia (for a total of 51 "states"), by year for 1977 – 1999. Each observation is a given state in a given year. We will analyse the relation between states that have a shall-carry law in year x and violent crimes. For this purpose, you will have to:

1. Generate the logarithm of the violent rate.
2. Check if there is variation over the variable of interest (*shall*).
3. Analyse two periods: '95 and '96. Then run a simple regression on the whole sample: *lnvio vs. shall*
4. Set panel data (by state and year) and run fixed effects regressions using `xtreg` command and using factor syntax (`xi:`). Use the same model of 3.

Exercise 2:

In this second exercise, we will analyse the *Medicare Modernization Act* of 2003 following the work of Engelhardt, Gary V and Jonathan Gruber (2011). "Medicare Part D and the Financial protection of the elderly". In: American Economic Journal: Economic Policy 3.4, pp. 77-102.
In this exercise, your are asked to:

1. Replicate the graphical evidence on the age of profile of prescription-drug coverage before and after the enactment of Part D of Medicare (Figure 1 of the paper)[5].
2. Replicate table 3 of the paper: Difference-in-Difference evidence.

   *Use the variable `clustid` for clustering in each of the regressions.

2 Social Programs

A Public Infrastructure Program

One the most common programs implemented by the public sector are infrastructure projects. These type of programs are supposed to have a huge impact on different aspects of society. However, to quantify this impact is not straightforward, and represents a big challenge for academics and policy makers. In this exercise, we are going to work with the data provided by: McIntosh, Craig, Tito Alegría, Gerardo Ordóñez, and René Zenteno. 2018. "The Neighborhood Impacts of Local Infrastructure Investment: Evidence from Urban Mexico." American Economic Journal: Applied Economics, 10 (3): 263-86. This paper evaluates the impact of a large scale infrastructure investment experiment in Mexico on public infrastructure access and private investment in housing.
In this exercise we are going to replicate two of the tables from this paper. On one hand, we are going to replicate the table with the balance of the main variables, which is key information not only for RCTs or experiments, but for every evaluation. Then we are going to run fixed effect regressions on public infrastructure and on private investment. The steps are the following:

1. Run simple regressions from each of these variables *Disp_ Infr_ Bas, Disp_ Agua, Disp_ Drenaje, Disp_ Luz, Alumbrado_ Siempre_ Enc, Disp_ Guarniciones, Disp_ Banquetas, Disp_ Pavimento* against the treatment. Use proportion weights and cluster errors by polygon using *cve_ mun* variable. The export the results using the command `outreg2`.
2. Using the same variables from 1., run panel fixed effect regressions on the treatment and use these variables as controls: *satis_ colon_ fis, satis_ colon_ soc.* Also, use proportion weights and cluster errors by polygon using *cve_ mun* variable. The export the results using the command `outreg2`.
3. Using these variables: *brick_ walls, concrete_ floor, kitchen, bathroom, flush_ toilet, septic, piped_ water home_ owner, f_ banca_ privada, rent_ usd.*, run panel fixed effect regressions on the treatment. Also, use proportion weights and cluster errors by polygon using *cve_ mun* variable. The export the results using the command `outreg2`.

[5] It is not necessary to do use the same colors or "style" off the figure, but the shape and content must be the same.