Econometrics I (Second Half)
Problem Set 2

Instructions (Read Carefully)

Please submit all of the files related to your solution to this problem set via email to me (andrea.flores@fgv.br) and your TA Igor (igorbrito368@gmail.com) by **Sunday, December 8th at 11:59pm**.

Your solution to this problem set should consist of (1) a pdf with your responses to the items in each of the questions as if it was a report (for items that just require coding with no discussion, include a reference to a particular part of your code), and (2) the files containing the code used in each question that requires estimation.

Go as far as you can. Even if you don't manage to complete all items in a question, partial credit will be applied generously if you clearly describe the issues you faced in approaching that particular question and intuitively explain how you think these issues could be addressed.

Good Luck!

Question 1: Difference-in-Differences and Matching Methods

Throughout this question, we will focus on the impact of the urban implementation of *Progresa/Oportunidades* on children's school attendance. For this, you will be using the data file named `oportunidades_encelurb_ps2.dta` containing information on school-aged children from the official external quantitative evaluation of *Oportunidades*. The file contains information on current school attendance of children in the household as well as socioeconomic status indicators at the household level. The relevant variables for outcomes will include: current school attendance (1 if attending school, 0 otherwise), weekly hours spent in school, total education-related expenditures made on the child, whether the child has received help from parents when doing homework, grade failure (repetition), weekly hours spent in housework, and weekly hours spent in child care.

(a) Compute the difference-in-differences estimator manually by *computing and reporting* – in each of the cases below – the following: $\bar{y}^1_{t_1}$ (mean outcomes for children in households receiving the cash transfer after the disbursements were made), $\bar{y}^1_{t_0}$ (mean outcomes for children in households receiving the cash transfer before the disbursements were made), $\bar{y}^0_{t_1}$ (mean outcomes for children in households **not** receiving the cash transfer after the

disbursements were made), $\bar{y}_{t_0}^0$ (mean outcomes for children in households receiving **not** the cash transfer before the disbursements were made)

C1: Include all the sample of children, report $\widehat{\alpha}_{All}^{DID} = (\bar{y}_{t_1}^1 - \bar{y}_{t_0}^1) - (\bar{y}_{t_1}^0 - \bar{y}_{t_0}^0)$.

C2: Restrict the analysis to include only observations for girls, report $\widehat{\alpha}_{G}^{DID} = (\bar{y}_{t_1}^1 - \bar{y}_{t_0}^1) - (\bar{y}_{t_1}^0 - \bar{y}_{t_0}^0)$.

C3: Restrict the analysis to include only observations for boys, report $\widehat{\alpha}_{B}^{DID} = (\bar{y}_{t_1}^1 - \bar{y}_{t_0}^1) - (\bar{y}_{t_1}^0 - \bar{y}_{t_0}^0)$.

(b) For each of the cases (C1-C3) covered in part (a), run the following linear regression

$$Y_{it} = \beta_0 + \beta_1 Treat_i + \beta_2 Post_t + \alpha(Post_t \times Treat_i) + \epsilon_{it} \tag{1}$$

where $Post_t = \mathbb{1}[wave > 1]$. Using your results from part (a), verify that your OLS estimates for α coincide with the estimates for α^{DID} you calculated manually and interpret the following in terms of $\bar{y}_{t_1}^1, \bar{y}_{t_0}^1, \bar{y}_{t_1}^0, \bar{y}_{t_0}^0$:

- $\widehat{\beta}_0$
- $\widehat{\beta}_1$
- $\widehat{\beta}_2$

(c) We have enough reasons to suspect that significant pre-treatment differences between untreated and treated households – potentially driving households' program participation decision – might be contaminating our estimates of the ATT of *Oportunidades*. We then want to implement the Matching DID (MDID) estimator, which explicitly accounts for the participation decision by exploiting information on each households' probability to participate in the program. Before considering the matching algorithm, you must estimate the propensity score. This propensity score then captures the probability that a household participates in *Oportunidades* as a function of different set of variables captured at baseline. Then estimate the propensity score p using a logit model including household characteristics you suspect affect this decision.

Note: The propensity score should be estimated at the household level, so that two children living in the same household would have the same propensity score associated with them.

(d) Using p, compute and report $\widehat{\alpha}_{All}^{MDID}$, $\widehat{\alpha}_{G}^{MDID}$, and $\widehat{\alpha}_{B}^{MDID}$ (for each of the cases covered in part (a)) implementing the following matching algorithms:

- 2 nearest neighbor matching (with replacement)
- Epanechnikov kernel matching using Silverman's rule of thumb for bandwidth selection

Interpret your results.

Question 2: Event Study Design

In this question, we will use an extract from the PSID to quantify the effect of first childbirth on both mothers' and fathers' labor market outcomes. For this, refer to the file PSID_event_study.dta. In this dataset, you have a variable defined as *yearkid1* which captures the year of birth of the first child of each individual in the panel. There is also information on age (*age*), education (*educr*), race(*white* and *black*), marital status (*ms*), year-specific dummy variables (*Dyear1968, Dyear1969, ... , Dyear2001, Dyear2002*), age-specific dummy variables (*Dage20, Dage21, Dage22, ... , Dage44, Dage45*) .

Throughout this question, we consider the following specification:

$$Y_{istk} = \sum_{j=-3}^{-2} \alpha_j \mathbb{1}[j = k] + \sum_{j=0}^{10} \alpha_j \mathbb{1}[j = k] + \sum_{l \in [20,45]} \gamma_l \mathbb{1}[age_{istk} = l] + \beta X_{it} + \eta_s + \eta_t + \epsilon_{istk} \qquad (2)$$

X_{it} denotes a vector of controls, including education (*educr_new*), a quadratic polynomial on education (*educrsq_new*), race (*black, white*), and a categorical variable capturing marital status (*ms*), and η_s and η_t denote state and year fixed effects. The age indicators are included in the dataset as *Dage20, Dage21, Dage22, ... , Dage44, Dage45*

(a) Using the variable *yearkid1* as the year in which the event occurred, define an event-time variable centered around this year so that $event = year - yearkid1$ for each individual in the panel. To check that this has been defined correctly, make sure that these coincide with the pre-constructed dummy variables called *Deventm3, Deventm2, ... Devent0, ..., Devent10*.

(b) Using the specification described in 2, run an event study on the following outcomes

- Employment rate (*dn*)
- Hours worked (*wrkhrs*)
- Earnings (*lbrx*)

As a normalization, leave out the dummy variable corresponding to the year just before the birth of the first child. Make a plot of the coefficients associated with the event dummies (α_j's) with confidence intervals for men and women separately. Interpret your results.

(c) Re-do the estimation and plots from part **(b)** for the two sub-samples described below (again, for men and women separately):

1. **Sub-sample 1:** individuals **not exposed** to a job-protected leave policy at the time their first child is born (*policy* = 0).

2. **Sub-sample 2:** individuals **exposed** to a job-protected leave policy at the time their first child is born (*policy* = 1). *In the specification for this sub-sample, omit the dummy variables for the years 1968 and 1969 as this would generate an unbalanced panel in event times.*

Interpret your results.

(d) Discuss the differences (if any) between the estimates you obtained in parts **(b)** and **(c)**. Interpret your results. In your discussion, carefully address the following:

- Are there any significant differences in the labor market outcomes of parents exposed to job-protected leave and those not exposed to these policies **before** first childbirth?

- Are there any significant differences in the labor market outcomes of parents exposed to job-protected leave and those not exposed to these policies **after** first childbirth?