Econometrics I (Second Half):
Take-Home Exam

INSTRUCTIONS (READ CAREFULLY)

Please submit all of the files related to your solution to this exam via email to me (andrea.flores@fgv.br) and your TA Igor (igorbrito368@gmail.com) by **MONDAY, DECEMBER 16TH, 2024 at noon (Rio time)**.

Your solution to this exam should consist of **(1)** a **pdf** with your responses to the items in each of the two 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.

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 and Happy Holidays!

Question 1: Policy Evaluation

In this exercise, we will evaluate the impact of the rural implementation of *Progresa* on children's school enrollment using several of the policy evaluation methods covered during this second half of the course. The estimators are intended to exploit the fact that program eligibility was partly determined by a wealth index and a cutoff that varied from one municipality to the other and that treatment was randomized across municipalities. Thus, this exercise will serve as a way to test the extent to which non-experimental methods can help recover the treatment effects of interest within a randomized experiment.

Throughout this question, you will refer to the csv file `progresa_policy_evaluation.csv`. Notice that the dataset is formatted as a wide panel. The outcome variable of interest is y which captures school enrollment status of children aged 6-16. In the dataset, you can find the variable called *cut*, which captures the cutoff values c (it is okay to see variation in the values of these cutoffs since these vary by municipality). $yycali\_97$ contains the values of the wealth index used by the *Progresa* administration to assess eligibility such that households with $yycali\_97$ above the cutoff were assigned to the control group and households such that $yycali\_97$ are at or below the corresponding cutoff are treated (i.e. receive the cash transfer) [thus, *yycali_97* would be denoted as the running variable within a RD approach].

(a) Suppose that you trust that randomization was properly implemented by the program administration such that you can expect a simple comparison of outcomes after the receipt of the *Progresa* cash transfer would suffice to identify the effect of the program on children's school enrollment. For this, estimate the following:

$$s_i = \alpha + \beta Treat_i + \epsilon_i$$

Report and interpret your results. Do your results change once you add controls relating the characteristics of children at baseline? Explain.

(b) You now suspect that there might be some issues with the randomization implemented by the program and that potential selection of gains might not be fully controlled for in the specification estimated in **(a)**. Thus, estimate the following exploiting the fact that you observe outcomes for children both before and after the rollout of *Progresa* to capture the effect of the program on children's school attendance:

$$s_{i,t} = \alpha + \beta_1 Treat_i + \beta_2 Post_t + \beta^{DID}(Post_t \times Treat_i) + \epsilon_{i,t} \tag{1}$$

For this, report and interpret the results from implementing the specification described in 1. Do your results change once you add controls relating the characteristics of children at baseline? Explain.

(c) We now want to exploit our knowledge on non-parametric methods to estimate β^{RDD} using kernel-based local linear regression – letting $X_i = yycali\_97_i$ and using children's school enrollment after the rollout of *Progresa* as the outcome variable ($Y_i = y\_post_i$) – implement the following steps:

1. Define $\tilde{X}_i = \begin{bmatrix} 1 \\ X_i - c \end{bmatrix}$
2. For observations such that $Treat_i = 1$, compute
$$\widehat{\beta}_1 = \left(\sum_{i=1}^{n} K\left(\frac{X_i - c}{h}\right) \tilde{X}_i \tilde{X}_i' \mathbb{1}\{X_i - c \leq 0\} \right)^{-1} \left(\sum_{i=1}^{n} K\left(\frac{X_i - c}{h}\right) \tilde{X}_i Y_i \mathbb{1}\{X_i - c \leq 0\} \right)$$
3. For observations such that $Treat_i = 0$, compute
$$\widehat{\beta}_0 = \left(\sum_{i=1}^{n} K\left(\frac{X_i - c}{h}\right) \tilde{X}_i \tilde{X}_i' (1 - \mathbb{1}\{X_i - c \leq 0\}) \right)^{-1} \left(\sum_{i=1}^{n} K\left(\frac{X_i - c}{h}\right) \tilde{X}_i Y_i (1 - \mathbb{1}\{X_i - c \leq 0\}) \right)$$
4. Let $\widehat{\beta}^{RDD}$ be the first element of $\widehat{\beta}_1 - \widehat{\beta}_0$.

Use the Epanechnikov kernel function. Compute the bandwidth using Silverman's plug-in estimator for the Epanechnikov kernel (check lecture slides). Report your results in a table. Test the sensitivity of your results for β^{RDD} to (i) an increase in h and (ii) a decrease in h. Interpret your findings **and** compare with your estimates of β obtained in part **(a)** and of β^{DID} obtained in part **(b)**.

Question 2: Simulated Method of Moments (SMM)

Suppose we are interested in setting up a child schooling model that explicitly relates parental decisions on children's school attendance with income. Due to the complexity of the version of the model we ultimately want to use for counterfactual policy analysis, our estimation approach will involve a simulated method of moments estimator. For this, we consider a simplified static version of our model.

In our model, a family/household/parents choosing to send her child to school (i.e. $d_i = 1$) receives utility u_{i1} and receives utility u_{i0} if the family chooses to not send her child to school (i.e. $d_i = 0$) such that

$$u_{i0} = v_{i0}(x_i) + \epsilon_{i0}$$
$$u_{i1} = v_{i1}(x_i) + \epsilon_{i1}$$

The solution of the model is the following

$$d_i = \begin{cases} 1 & \text{if } u_{i1} = v_{i1}(x_i) + \epsilon_{i1} \geq v_{i0}(x_i) + \epsilon_{i0} = u_{i0} \\ 0 & \text{otherwise} \end{cases}$$

Letting parents' payoff be linear in income for both alternatives, then:

$$v_{i1}(x_i) = \gamma_{01} + \gamma_{11} x_i$$
$$v_{i0}(x_i) = \gamma_{00} + \gamma_{10} x_i$$

where ϵ_{i0} and ϵ_{i1} are drawn from an invertible $F(\cdot)$.

(a) *Identification:* Show that we can only identify $\beta_0 = \gamma_{01} - \gamma_{00}$ and $\beta_1 = \gamma_{11} - \gamma_{10}$ if we observe data on income x and parental decisions to send children to school d.

- *Hint:* You can derive the mapping between the model and the data using the decision rule of the family in terms of $\gamma_{01}, \gamma_{00}, \gamma_{11}, \gamma_{10}$, x_i, and individual unobserved heterogeneity.
- You **do not** need to impose a parametrization on the distribution of the unobserved heterogeneity to show identification.

(b) *Dataset Creation:* Suppose we first want to assess the extent to which an SMM estimator can recover true parameters values in the simplest case scenario and with "fake" data (in the sense that we will know the true data generating process).

For this, use the parent's decision rule to numerically simulate 100 data sets of size 500 and 1000. That is, we want to generate 100 datasets with 500 observations of income (x) and parental schooling decisions (d) in each dataset and another 100 datasets with 1000 observations of x and d in each dataset. *Note: You could interpret these 100 datasets for each sample size as corresponding to different municipalities/localities/states.*

Take the following steps to generate **each** dataset:

- Let the true parameter vector be $\beta = (\beta_0, \beta_1) = (0.5, 0.4)$
- Draw x from the log normal distribution with mean 3 and variance 1 for each observation. – This should yield a vector of size N.
- Assuming that ϵ_{i0} and ϵ_{i1} are independently and identically distributed according to the Type I Extreme Value Distribution, draw these two alternative-specific shocks for each individual observation. – This should yield two vectors of size N (one associated with $d_i = 0$ and the other one associated with $d_i = 1$).
- Substitute your true parameter values, the vector of x and the vectors of random shocks ϵ_0 and ϵ_1 into the decision rule derived in part (a) to generate the choice vector d – This should yield a vector of size N.

(c) *Implementation of SMM Estimator on each of the Generated Datasets:* You will implement a simulated method of moments estimator for β:

$$\hat{\beta}^{SMM} = \arg\min \left[\mathbf{m}(x) - \frac{1}{S} \sum_{s=1}^{S} \mathbf{m}(x(\beta)) \right]' W^{-1} \left[\mathbf{m}(x) - \frac{1}{S} \sum_{s=1}^{S} \mathbf{m}(x(\beta)) \right] \tag{2}$$

Using **30** simulation draws (i.e. $S = 30$) to approximate the choice probabilities in each estimation,

- Estimate β 100 times using the datasets of sample size 500 created in part (b). Compute and report the mean, variance, and mean square error of your estimates for these coefficients given that you know the true value.
- Estimate β 100 times using the datasets of sample size 1000 created in part (b). Compute and report the mean, variance, and mean square error of your estimates.

In each estimation procedure use two moments: the mean of d_i and the covariance of x_i and d_i. You can choose to let W be the identity matrix.

(d) Repeat part (c) with **100** simulation draws to approximate the choice probabilities.

(e) Compare your results obtained in parts (c) and (d). Make sure to address the following two questions in your comparison: (1) For a specific number of simulations, do you get closer to the true parameter values as you increase N? (2) For a specific number of observations, do you get closer to the true parameter values as you increase the number of simulation draws? Discuss.

(f) *Counterfactual Policy Experiment:* Take the 100 datasets with sample size of 1000 created in part (b). In each dataset, partition the sample into quartiles and keep only the observations from the bottom quartile. Randomly select 50 of the 100 datasets and assign them into an "experiment" in which the household will receive a 20% increase in household income, while the 50 datasets not selected in the randomization will not receive any additional income. That is, for those datasets randomly selected into the policy you are creating ($Treat_i = 1$), $x_i = x_i(1 + \Delta)$, where $\Delta = 0.15$ in this experiment, while for households not selected into the program ($Treat_i = 0$) $x_i = x_i$. Compute the "treatment effect" of this policy in the following ways:

- Among the households selected to treatment ($Treat_i = 1$), compare the average school attendance (d) after giving them the transfer equivalent to 15% of their income with their average school attendance without the transfer ($x_i = x_i$).
- Among the households in the bottom quintile of all the 100 datasets, compare the average school attendance (d) of households who receive the transfer ($Treat_i = 1$) with the average school attendance of households that were not selected for this policy experiment ($Treat_i = 0$).

Discuss your results.

Important Comments to Keep in Mind:

For Question 2:

- In part (b), make sure you seed when generating these datasets to ensure the replicability of this part.
- In each simulation used for the SMM, you will **not** be drawing new vectors of x since we are assuming x is observed and directly coming from each dataset generated in part (b), so the only vectors that need to be drawn in each simulation to predict choices are the ones associated with the ϵ shocks.
- It is completely normal for your SMM estimates to be sensitive to: (1) the choice of moments (you could also try checking what happens when you target the variance of d in the SMM), (2) the optimizer you use, and (3) your initial guess for β.
  - Since you know the true values of β, you can check what happens when your initial guesses

significantly deviate from the true values (this could give you some intuition of how well your optimizer is performing, especially since this is a very simple version of a discrete choice model).

- For parts (c) and (d), it is not a problem with your code if it takes some time to run – remember that you are implementing the estimator 100 times and each time, you are making 30 and 100 simulations (respectively).
- For parts (c) and (d), the mean, variance and mean squared errors relate to the 100 estimated parameters. Thus, the mean squared error provides a measure of how close you are getting to the true value of the parameters, on average.
- Given the simplicity of the model and the parametric assumptions made on the alternative-specific shocks (Type I EV), we then know that $F(\epsilon) = \exp(-\exp(-\epsilon))$ and – most importantly – that the difference between ϵ_{i0} and ϵ_{i1} has been shown to be logistic distributed.
  - It would be possible to also use a logit (i.e. maximum likelihood) to estimate the α as we would be able to analytically derive the likelihood function. It is not necessary to do this in this exam, but it is an exercise worth doing when making a decision between estimators (and when deciding the types of simplifying assumptions you are willing to impose to use a particular estimator).
- **Some optimizers might work better than others.** So do not hesitate to try multiple algorithms (such as simplex-based methods and gradient methods). One crucial way to detect problems with an optimizer is if the estimates obtained upon the optimization procedure do not move much from the initial guesses. In this case, the optimizer is not doing much in being able to get out of the initial region to search further away for an actual minimum.
- In parts (c) and (d), a good way to check that your code is working correctly is to check the behavior of your objective function around the estimates you obtain. For this, you can try to graph the objective function on a grid around each parameter estimate while fixing the rest. That is, take your estimates $(\hat{\beta}_0, \hat{\beta}_1)$ and try the following:
  - Define a grid around $\hat{\beta}_0$ and set $\beta_1 = \hat{\beta}_1$ and plot the objective function value on the grid defined when varying β_0.
  - Define a grid around $\hat{\beta}_1$ and set $\beta_0 = \hat{\beta}_0$ and plot the objective function value on the grid defined when varying β_1.

  Throughout these checks, you want to make sure that the objective function is concave on each of the grids over which you are plotting it.