Econometrics I (Second Half): Take-Home Exam

INSTRUCTIONS

Please submit all of the files related to your solution to this exam via email to me (andrea.flores@fgv.br) and Yan Richard (yan.richard.oliveira@gmail.com) by **SUNDAY, DECEMBER 18TH, 2022 at 23:59h**.

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 were 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.

The solution to both questions in the exam count equally towards the perfect score. Good luck!

Question 1: Regression Discontinuity Design

In this exercise, we will evaluate the impact of the rural implementation of *Progresa* using an RD design method. The estimator is intended to exploit the fact that program eligibility was determined partly on the basis of an eligibility rule and a cutoff.

The data for this question is provided in the csv file `progresaRDD_exam_q1.csv`. The outcome variable (y) is the school enrollment status one year after the intervention of children who were aged 6-16 at baseline. In the dataset, the cutoff values c are captured in the variable *cutoff* (these cutoffs are varying at a local level). *xraw* contains the raw values of the marginality index used to assess eligibility (i.e. X). Importantly, *xnorm* captures $X_i - c$ (so that your analysis will require this variable). A household is treated if $X - c \leq 0$ and untreated otherwise. For graphing purposes throughout this question, you can use the variable *xbin* rather than *xnorm* as it bins *xnorm* values.

(a) Suppose we start by using a parametric model to estimate the treatment effect using RD such that

$$Y = \alpha_l + \phi_l(X - c) + \epsilon$$
$$Y = \alpha_r + \phi_r(X - c) + \epsilon$$

- Assume that ϕ_l and ϕ_r are both a quadratic function of $(X - c)$.
- Assume that ϕ_l and ϕ_r are both a cubic function of $(X - c)$.
- Assume that ϕ_l and ϕ_r are both a quartic function of $(X - c)$.

Note that using the specifications of the outcome equation presented above, we can obtain α^{RDD} using the following pooled regression equation:

$$Y = \alpha_l + \alpha_D \mathbb{1}\{X - c \leq 0\} + \phi_l(X - c) + \mathbb{1}\{X - c \leq 0\}\phi_r(X - c) + \epsilon \tag{1}$$

where $\mathbb{1}\{X - c \leq 0\}$ is an indicator variable such that it is equal to 1 if $X - c \leq 0$. Then, $\widehat{\alpha}^{RDD} = \widehat{\alpha}_D$ within this regression framework.

For each of the parametric models used to obtain the RD estimate, report your results in a table and graphically. For the graphs, first compute the fitted values generated by each parametric model and then generate a scatter plot of these fitted values in the y axis and X in the x-axis. Interpret your findings.

(b) We now want to exploit our knowledge on non-parametric methods to estimate α^{RD} using kernel-based local linear regression implementing the following steps:

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

Use the triangular kernel function. Start with a bandwidth of 15 and check the sensitivity of your results to increasing the bandwidth to 30. Report your results in a table. Interpret your findings **and** compare with your parametric RD results obtained in part **(a)**.

Question 2: Simulated Method of Moments (SMM) in a Static Discrete Choice Framework

Suppose we are interested in setting up a structural model of child schooling 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. However, 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, we consider a simplified static version of our model.

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

$$\begin{aligned} u_{i0} &= v_{i0}(x_i) + \epsilon_{i0} \\ u_{i1} &= v_{i1}(x_i) + \epsilon_{i1} \end{aligned}$$

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:

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

(a)
- Write down the decision rule of the parent in terms of $\gamma_{01}, \gamma_{00}, \gamma_{11}, \gamma_{10}$, x_i, and individual unobserved heterogeneity.
- Show that we can only identify $\alpha_0 = \gamma_{01} - \gamma_{00}$ and $\alpha_1 = \gamma_{11} - \gamma_{10}$ if we observe data on income x and parental decisions to send children to school d.

(b) 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.

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

- Let the true parameter vector be $\alpha = (\alpha_0, \alpha_1) = (0.8, 0.7)$
- Draw x from the log normal distribution with mean 5 and variance 3 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) You will implement a simulated method of moments estimator for α:

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

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

- Estimate α 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 α 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.

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 α.
  - Since you know the true values of α, 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).