Escola de Economia de São Paulo - Fundação Getulio Vargas

Course: Microeconometria 2

Instructor: Bruno Ferman and Vitor Possebom

Problem Set: Lectures 3-4

Total = 250 points

Question 1 (Lecture 3: Regression Discontinuity Design - 100 points)

We will learn how to use three packages in R:

1. `rdrobust`*: Use a local linear regression to estimate the CATE at "X = threshold" in a sharp setting or the LATE at "X = threshold" in a fuzzy design. In this exercise, we will focus on the sharp case.*

2. `rddensity`*: Use a local polynomial density estimator to implement the density test proposed by McCrary (2008)*

3. `RATest`*: Implement the covariate balance test proposed by Canay and Kamat (2017).*

To do so, we will analyze a classic RDD example: the impact of being a incumbent senator on the probability of being reelected in the US. The dataset `rdrobust_senate.csv` *contains 1.390 observations, where the observation is an election in state i and year t. It contains the following variables:*

1. `vote`*: The outcome variable is* `vote`*, which ranges from 0 to 100 and records the Democratic vote share in the following election for the same seat (that is, six years later). The cutoff is normalized to 0.*

2. `margin`*: The running variable is* `margin`*, which ranges from -100 to 100 and records the Democratic party's margin of victory in the statewide election for a given U.S. Senate seat, defined as the vote share of the Democratic party minus the vote share of its strongest opponent.*

3. `class`: *This covariate identifies the electoral class each Senate seat belongs to (this indicates which of the possible three electoral cycles each seat is in).*

4. `termshouse`: *This covariate captures the experience of the Democratic candidate by recording the cumulative number of terms previously served in the U.S. House.*

5. `termssenate`: *This covariate captures the experience of the Democratic candidate by recording the cumulative number of terms previously served in the U.S. Senate.*

6. `population`: *This covariate records the population of the Senate seat's state.*

Answer the following questions:

1. *(10 points) Report summary statistics for all variables in the dataset.*

2. *(20 points) We start with a graphical analysis. Using the function* `rdplot`*, plot the conditional expectation of* `vote` *as a function of* `margin` *using bins. Include confidence intervals and a global regression function. To choose the bandwidth, use IMSE-optimal evenly-spaced method based on spacings estimators.*

3. *(20 points) Now, estimate the desired treatment effect parameter using the function* `rdrobust`*. Report the conventional point estimate and the robust confidence interval.*

4. *(20 points) Implement McCrary's Density Test using* `rddensity`*. Use the robust test statistic and p-value. Do you reject the null of no-manipulation?*

5. *(10 points) Visually inspect the last test using* `rdplotdensity`*.*

6. *(20 points) Using* `rdperm`*, jointly test whether the variables* `class` *and* `population` *present discontinuities at the threshold value.*[1] *Report the p-value of the joint test. Do you reject the null of covariate balance?*

~~Question 2~~ (Lecture 4: Fuzzy RDD with Multiple Cutoffs - 100 points)

[1] To include the covariates `termshouse` and `termssenate`, we need to eliminate their missing values. If you want some extra fun, feel free to include these variable in our covariate balance test.

Cattaneo, Keele, Titiunik and Vazquez-Bare (2016) also analyze the case of the Fuzzy Regression Discontinuity Design with Multiple Cutoffs. We will go through their work to have a better understanding of this scenario.

1. *(5 points) State and interpret Assumption 6.*
2. *(5 points) State and interpret the pooled fuzzy estimand* $\left(\tau_{FRD}^{P}\right)$.
3. *(5 points) State and interpret Assumption 7.*
4. *(5 points) State and interpret Assumption 8.*
5. *(30 points) State and interpret Lemma 2.*
   (a) *Double average of what?*
   (b) *How do you interpret each conditional treatment effect parameter within the double average?*
   (c) *First term of the numerator of the weight.*
   (d) *Second term of the numerator of the weight.*
   (e) *Third term of the numerator of the weight.*
   (f) *What is the trade off between the local and global interpretation of this term?*
6. *(50 points) Prove Lemma 2.*

Question 3 (Lecture 4: Estimating Treatment Effect Derivatives - 50 points)

Propose a consistent estimator for the treatment effect derivative (Theorem 1 from Lecture 2) in a Sharp RDD.