FLS 6415: Replication 5 - Discontinuities

April 2019

To be submitted (code + answers) by midnight, Wednesday 15th May.

First read the paper by Titiunik (2011) on the class website. However, we will not use her pre-prepared dataset - we will start the analysis by constructing our own. As with all regression discontinuities, 90% of the work is in preparing the dataset.

1. We need data from 2000 and 2004 Mayors. We can download this from cepespdata.io. Choose "Consultar resultados eleitorais" and we want prefeito data at the municipal level for *parties* in the 2000 elections first. Then 'Selecionar Colunas' to add the `COD_MUN_IBGE` variable. Then export this dataset to 'CSV'. Finally, make the same selection for 2004 and download this as a separate CSV. Details for the description of each variable can be found on cepesp.io (see adicionar colunas).

2. First, prepare the 2000 dataset: a. Filter the data to include only the first round and to remove municipalities where only one party ran. b. Calculate the total number of votes in each municipality. c. Calculate the percentage vote share for each party in each municipal contest d. Calculate which position the party came in the municipal election (their `rank`). e. Filter the dataset to focus on only the first and second-placed parties. f. Make a binary variable that is equal to '1' for the winning party that becomes the incumbent. g. Add a column for the vote share of the winning party, and a second column for the vote share of the second-placed party. *(potentially tricky, email me if you're stuck)* h. Remove two annoying cases where the election result was tied (so we dont know who became the incumbent)! i. Calculate the winning margin of each party. - For the winning party, this is the vote share of the winner minus the vote share of the second place party. - For the second-place parties, this is the vote share of the second place party minus the vote share of the winning party.

3. Next, prepare the 2004 dataset: a. Filter for the first round. b. Calculate our outcome measure: the vote share (*not* the winning margin) of each party in each municipal contest. c. Select only the Municipality Code, Party and Vote Share variables

4. Join the two datasets (2000 and 2004 for all parties) based on the municipality (`COD_MUN_IBGE`) and party (`NUMERO_PARTIDO`) so that for every party that ran in both 2000 and 2004 we know what vote share they got in 2004. (What type of join do we want here? Left, Right, Inner?)

5. If we did not know about regression discontinuity we might run the observational OLS regression of 2004 vote share on incumbency in 2000. For the next set of questions we will focus only on the PMDB. Subset the data so it includes only the PMDB, run and interpret this regression.

6. Before implementing any regression discontinuities, let's check for balance around the discontinuity. Within a +/-1% winning margin in 2000 check the balance of the total number of voters in treated and control municipalities in 2000 (we created this variable in Q2). Compare this to the balance for a winning margin of +/-3%.

7. Next, check for sorting and manipulation of the threshold with the McCrary density test using the `rddensity` function. Interpret the results and produce a density plot using the `rdplotdensity`.

8. Before we run the analysis, let's construct a regression discontinuity plot to visually inspect the causal effect of incumbency at the threshold. Using a pre-packaged command like `rdplot` from the `rdrobust` package, create a regression discontinuity plot for the effect of incumbency in 2000 on vote share in 2004 for the PMDB. Use linear regression lines. Interpret the results.

9. Create a second regression discontinuity plot with fourth-order polynomial regression lines.

10. We will now implement four alternative specifications of the same regression discontinuity. For the first version of the analysis, implement a simple difference-in-means test comparing the average vote share received by the PMDB in 2004 within a bandwidth of +/-3% winning margin in 2000. Interpret these results and compare to the observational regression in Q5.

11. For the second version, implement the full-data regression discontinuity analysis. Interpret this regression and compare it to your results in Q10.

12. For the third version, implement the limited-bandwidth regression discontinuity analysis for a bandwidth of +/-3%. Interpret this regression and compare it to your results in Q10 and Q11.

13. Fourth, let's implement the optimal-bandwidth linear regression discontinuity using the `rdrobust` command. What bandwidth was selected? How do the results compare to the other methodologies?

14. Now let's try to adjust the functional form used to estimate the effect of the rnuning variable. Implement the optimal-bandwidth regression discontinuity but with a second-order polynomial (quadratic) trend. Also try a third-order polynomial (cubic) trend and assess the sensitivity of the results.

15. The Mayor of a small municipality calls you for political advice. He wants to know what vote share his party (the PMDB) is likely to receive in the next election. He is very confident because at the last election he won easily with a winning margin of 30%. Based on the evidence you have recorded above from the regression discontinuities, how would you advise the Mayor about his likely performance in the next election?

16. Choose your preferred specification and implement the regression discontinuity for the other two parties: the PFL and the PSDB. How similar are your results to those in Titiunik (2011) for the +/-3% window?