Policy Evaluation – Regression Discontinuity Analysis

Alexandre Kogioumtzis

KU Leuven

Replication Analysis of “The Incumbency Curse: Weak Parties, Term Limits, and Unfulfilled Accountability” by Marko Klasñja & Rocio Titiunik

1. Introduction

Reproduction of results in the field of scientific research is a growing topic of concern among the scientific community. By analysing and reproducing some of the results of the paper “The Incumbency Curse: Weak Parties, Term Limits, and Unfulfilled Accountability”, one can hope to reach a deeper understanding of the Regression Discontinuity method and also develop one’s awareness of the need to create reproducible results when producing an academic paper.

Furthermore, as a complement to the existing results of the study, a covariate matching was used to obtain an estimate of the impact of the margin of victory of an election on the margin of victory of the next one.

1. Literature

The method of Regression Discontinuity has been developed in 1960 (Thistlethwaite & Campbell, 1960) and has since then been implemented in most statistical software. Three researchers have been working on implementing it in Stata, the software used in the following paper. They are Calonico, Cattaneo and Titiunik. They have been working consistently on describing technical and methodological details of the Regression Discontinuity method has well and the following work will be based both upon material seen in class as well as their paper describing the foundations of Regression Discontinuity in R and Stata (Cattaneo, Idrobo, & Titiunik, 2019).

Notably, Professor Titiunik is one of the authors of the paper whose results are reproduced hereafter.

1. Methods & Data

*Explain the identification problem in a context-specific way (why not with and without comparison?)*

The topic of the paper is the analysis of mayoral elections results in Brazil over a period ranging from 1996 to 2012. The researchers want to analyse whether “[…] the presence of term limits and weak political parties affect the incentives and behaviour of individual politicians such that the parties suffer systematic losses”.

The authors could not assume that the treatment assignment of winning an election was randomly assigned as elections in Brazil are not an ideal experimental setting. In an ideal environment, one could observe both the margin of victory at time t+1 of a party having won an election in a municipality but also have the exact same party lose the same election in an exact replicate of the municipality for each factor considered in the experiment. This is something obviously not observable in practice and no two municipalities are the same. Therefore, they used a regression discontinuity model on the municipalities where a party barely lost and compared it to municipalities where they barely won so to “[…] isolate causal effects of winning office from the spurious correlation between current and future electoral success”. Regression Discontinuity models are used in quasi-experimental settings such as is the case when analysing political victories.

Regression Discontinuity (RD) designs have three key characteristics defining them.

The first is a score assigned to each individual unit considered, the second is a cut-off value that determines whether or not the units considered have the treatment effect considered assigned to them. The third element is the treatment effect in itself.

There are several important considerations to keep in mind when considering RDs, the value of the information provided by them is (usually) local and only evaluable right before and right after the cut-off point, they are local by construction.

In our case, the three fundamentals elements are well defined as each municipality is a unit, the cut-off is at a margin of win/loss of 0 and is strictly respected, indeed, winning an election gives access to mayoral powers without any exceptions.

An assumption of RD designs is that there exist conditions that allow to assume that units near the cut-offs only differ in the treatment effect considered. Another assumption is continuity in the functions of the independent variables both before and after the cut-off value.

In traditional RD designs, the treatment effect is defined as the difference in value of the dependent variable’s mean right before the cut-off with its evaluation right after it.

The dataset used for the study comes from the merging of “[…]a municipality-level dataset of demographic and socioeconomic variables obtained from the Institutito Brasileiro de Geofrafia e Estatistica (IBGE), with election returns and characteristics of individual candidates, parties and coalitions for mayoral and municipal legislature elections for 1996,2000,200,2008 and 2018, obtained from Brazil’s Tribunal Superior Eleitoral”. It contains 27 455 municipality-year observations and 5564 unique municipalities.

There were multiples variables present in the dataset, some such as GDP per municipality, population per municipality which were used for descriptive analysis purposes and others such as the margin of win (expressed in percentage) as well as whether each considered party had run or not.

The political parties considered were different in regard to the hypothesis posed by the authors. More precisely, the PT (Workers Party) was deemed big and cohesive enough to have an impact on their politician’s actions given their future power over them. Then, there are a multitude of smaller parties that have little power over their own politicians.

Notably, most of the analysis focused on a subset of the dataset above-mentioned as the particularities of RD designs lead to focus on close races, meaning those situated around the cut-off point of 0 with a range of +-2 points.

In contrast to Regression Discontinuity methods, a matching method will also be used to compare two estimates of the treatment effect. Matching allows to estimate a counterfactual even when using non-experimental data, which is the case in mayoral elections in Brazil. It matches (or pair up) two observations that differ as little as possible, according to a predefined measure of similarity, but with distinct treatment outcomes. Many types of matching exist, two were extensively covered in class and the one used hereafter, covariate matching with Nearest-Neighbour Matching (NNM), was chosen because the original paper does not use fuzzy treatment assignment and it would therefore not make sense to use a matching using a probability score like Propensity Score Matching (PSM) does.

NNM covariate matching was used with two covariates available in the dataset, the first one is the *population* and the second one is the *Municipal GDP per capita*. They were the only variables available in the dataset that would allow for such a matching in a meaningful manner. “Matching estimators use an average of the outcomes of the nearest individuals to impute the missing potential outcome for each sampled individual. The difference between the observed outcome and the imputed potential outcome is an estimate of the individual-level treatment effect” **INSERT REFERENCE TEFFECTS STATA**.

Matching required the creation of a dummy variable called **dummy\_win**, based on the cut-off value of 0 where if the incumbent’s margin of victory was positive, meaning the party won, the value of the dummy variable was set to 1 and 0 otherwise.

Some assumptions are needed in order to be able to use accurately a matching technique.

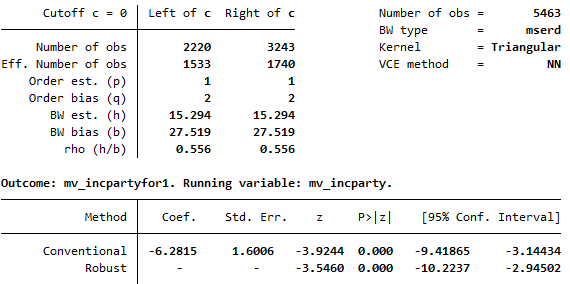
First, it must be assumed that the treatment groups depend on factors that are observed. Here, we assume that winning at time t+1 depends indeed on whether a party or not has won at time t. This assumption is strong and could lead to less reliable results than in the RD design since we do not have a very large number of data to draw conclusions from.

Second, compared to RD designs that are parametric, matching is a method that requires a lot of data as a trade-off to not being parametric and thus not needing parametric tuning.

Third, a “Common support assumption” requires similar units for much of the treatment and control group. This assumption will be verified in the following **Results** section.

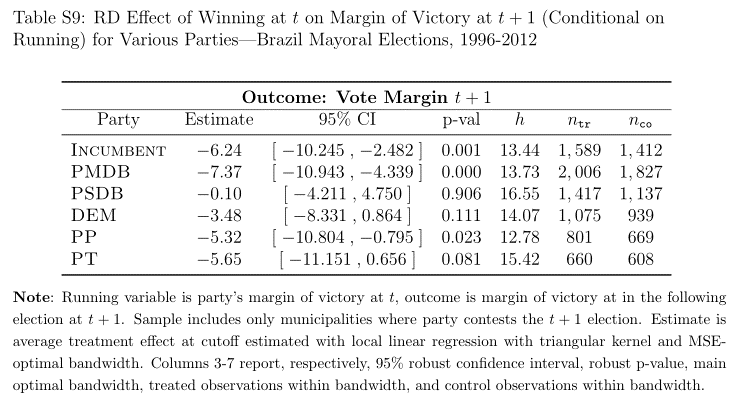
1. Results

The method of RD design was applied using STATA as previously studied in class with the main use of two command. The first one being **rdrobust** with mv\_incparty1 as the outcome variable, mv\_incparty as the running variable and a cut-off value set a 0. An interesting aspect of regression designs is that different orders of regression can be used to estimate the treatment effect and the regression curve. For the reproduction of results, the same parameters were used as in the paper. The order of the regression was set to 1, meaning thus that a linear regression was plotted. Another important parameter was the optimal Mean Squared Error which created bins for the analysis that were similar to those used in the study as well as the Triangular kernel, the kernel defines the weights to assign to observation based on their distance from the cut-off, values closer to the cut-off are more important in RD designs as the results are evaluated locally around the cut-off value. The results of the replication are in Figure 1.



1. Main results of the Replication analysis

The coefficient has a value of -6.2815 and represents the difference between the intercepts of the two separate regressions. The coefficient was evaluated on 5463 observations, a statistic not disclaimed in the appendix of the original paper where this estimation is made. In it, the value of the coefficient was of -6.24, there is therefore a percentage of difference between the results of 0.665%. The difference between those two results can maybe be explained by an update in the STATA library used in 2019 compared to the one that was used in 2015. Furthermore, we observe a different 95% confidence interval, the one in the replication is narrower but fully contained by the one from the study. The results are now more contained than before.



2. Results table from the 1st Online Appendix of the study

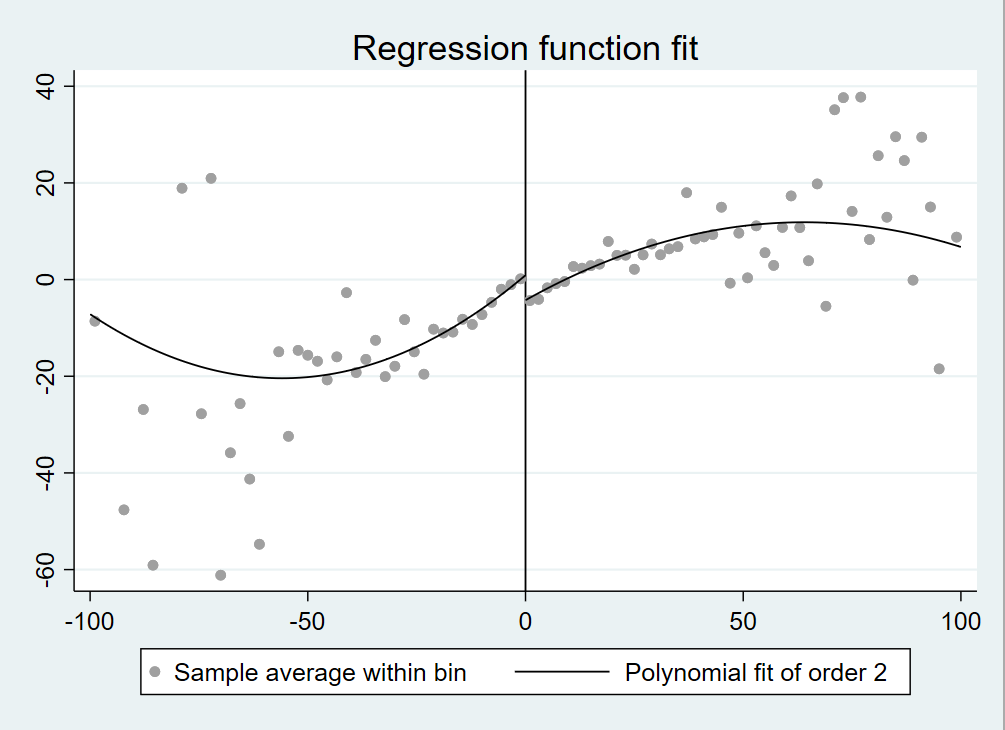
The second command used in Regression Discontinuity design in STATA is **rdplot** and displays the plot corresponding to the RD analysis.



3. Plot of the RD design - Linear Regression

To further the existing analysis, different polynomial orders were considered in order to evaluate the coefficients that would stem from them. Usually, no polynomial of order zero is used as it doesn’t do well at boundary points, which are the main focus of RD designs. Increasing the polynomial order can increase the accuracy of approximation but can also lead to overfitting. In the paper, mostly 4th order polynomial fits were used but the main regression functions cannot be reproduced as the data doesn’t contain the needed variables.

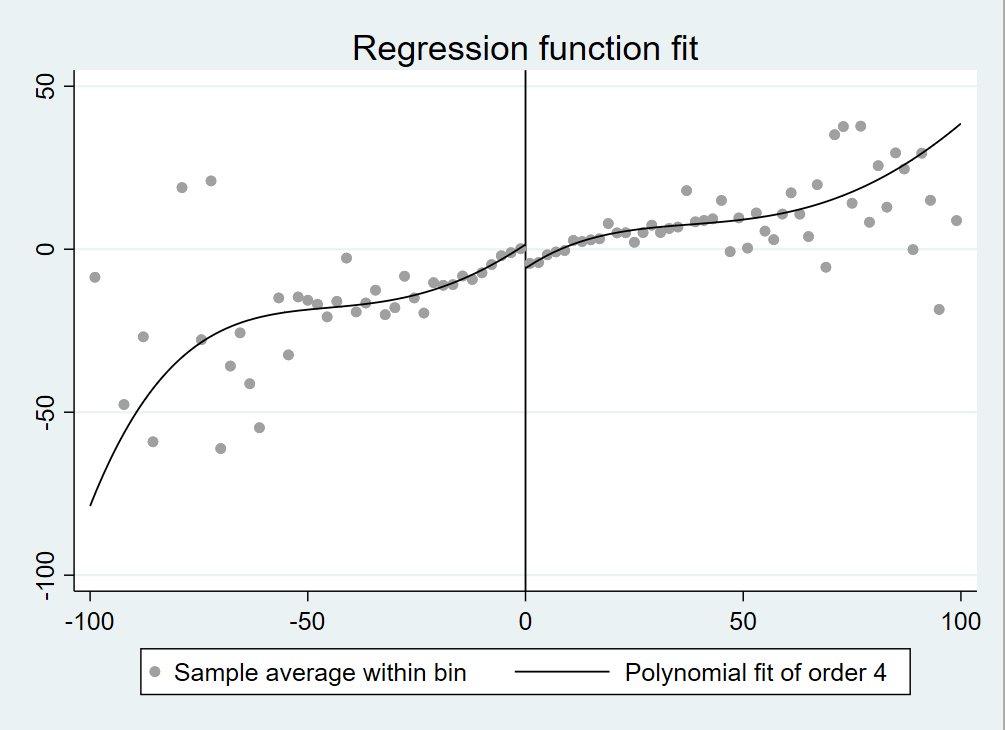
Hereafter are three plots with respectively a polynomial fit of order 2,3 and 4.



4. Plot of the RD design - Quadratic Regression

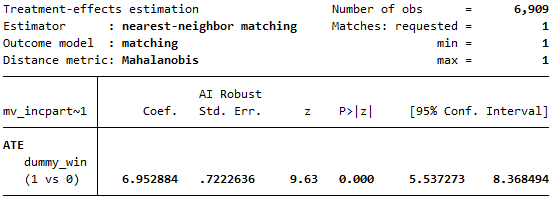


5. Plot of the RD design - Cubic Regression



6. Plot of the RD design - 4th order Regression

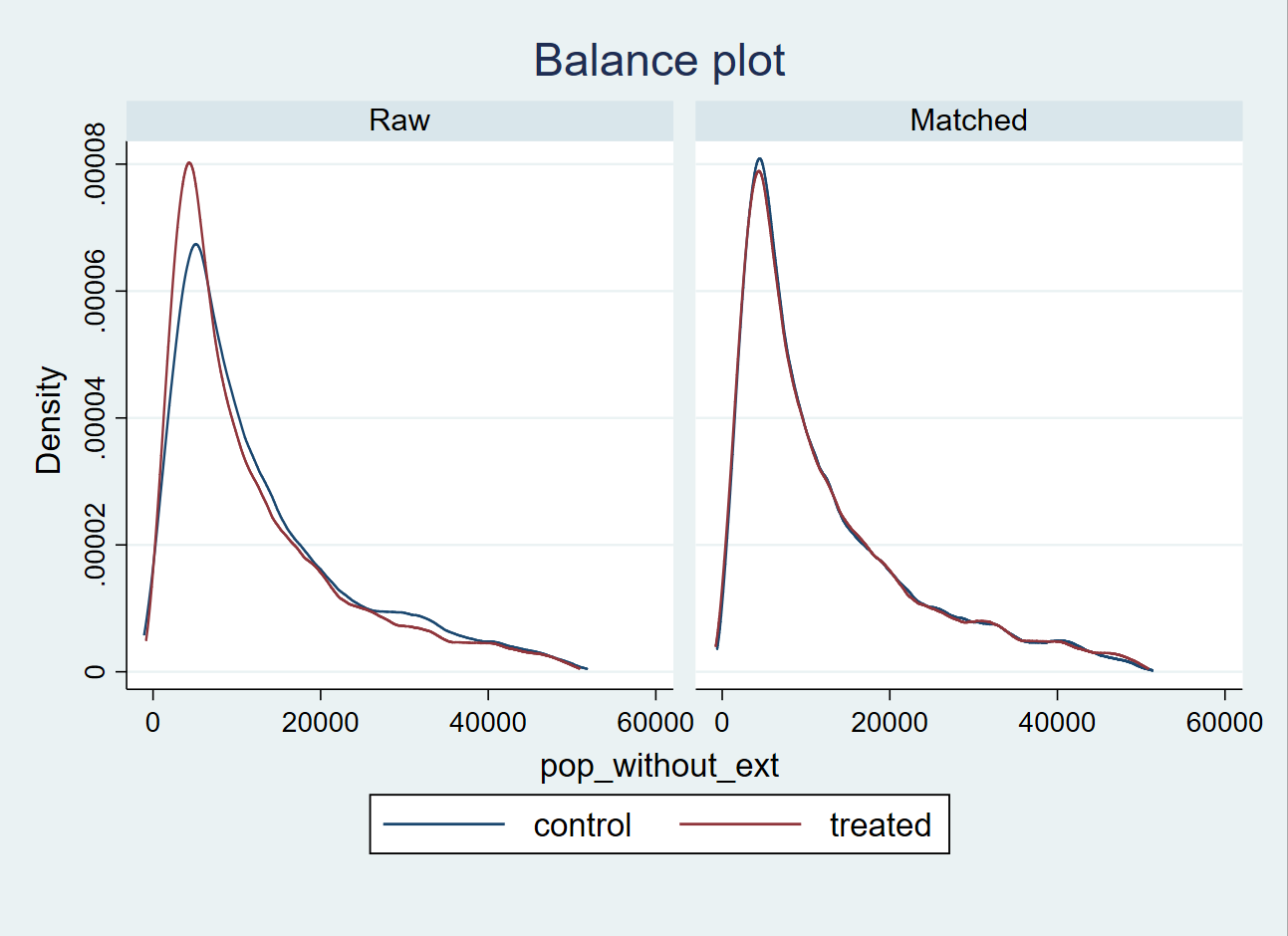
As discussed in the methodology, a covariate matching method using NNM was created in contrast to the RD design. The estimate for the coefficient was estimated using a Mahalanobis distance and returned a value of 6.953 on 6909 observations.



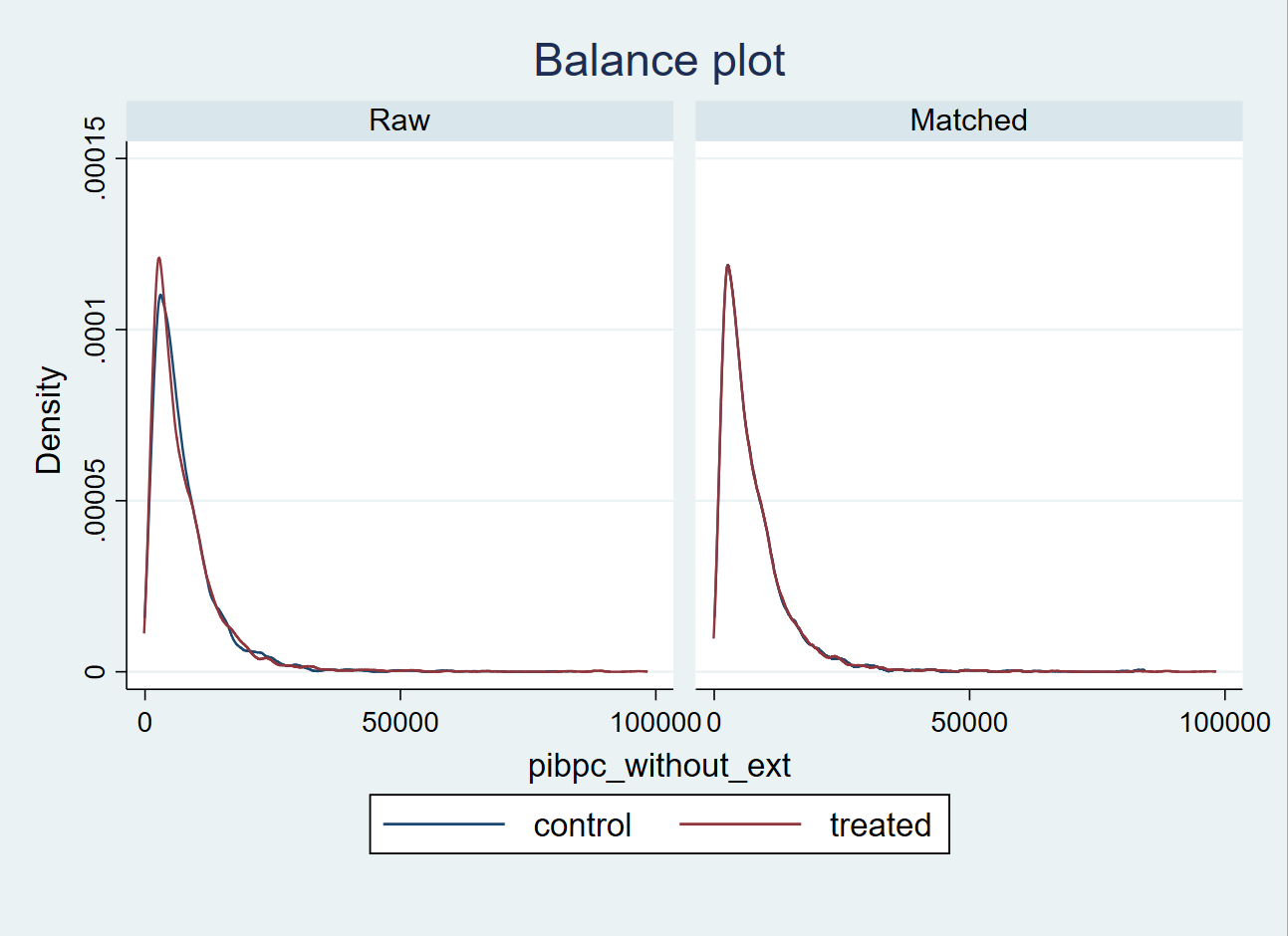
7. Covariate matching method with NNM

The difference of sign compared to the RD design can be attributed to the different way STATA applies the difference between the values of treatment and control groups in matching and RD techniques. In terms of relative difference between the two methods, one can observe a 10.68% increase in the value of the estimate.

The matching method’s assumption of “Common support” can be verified visually for NNM with the following command **tebalance density** and is done for both the population and the GDP per person in each communality. Both variables have to have their extreme values removed in order to obtain readable and pertinent x-axis. For the population, it is limited to values of 100 000 people and for the GDP, to values of 100 000 money units.



8. Common support assumption for population without values exceeding 100 000



9. Common support assumption for GDP per capita without values exceeding 100 000.

1. Conclusion

In conclusion, this paper is first made of a brief reference to scholar papers that represent the most commonly used background for Regression Discontinuity used with STATA and R. Secondly, an overview of the Regression Discontinuity method was made containing an explanation of its different parameters followed by a brief description of the dataset. The results that were reproduced evaluated the impact of winning an election on the following election. As a continuation to the analysis, a Nearest-Neighbour Matching technique was also applied in the same setting and returned a larger coefficient estimate but still close to the value of the RD.

Given the parametric properties of the RD, one can probably trust its estimate more than the estimate obtained on the few data used in this paper compared to a typical matching setting which is probably one of the reasons that led the researchers not to rely on matching techniques.

It must be noted that scope of the replication of the paper has been more limited than expected due to missing variables. However, the results replicated using the regression discontinuity design are similar to the ones presented in the study.

1. References

Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2019). *A Practical Introduction to Regression Discontinuity Designs*. https://doi.org/10.1017/9781108684606

Thistlethwaite, D. L., & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, *51*(6), 309–317. https://doi.org/10.1037/h0044319

Slides from the chapters on matching and on Regression Discontinuity (2019) - "Policy Evaluation" by Nick Deschacht - KU Leuven