Microeconometrics II — Problem Set 03 — RDD

Vinícius de Almeida Nery Ferreira

October 29, 2024

1 Implementation of Regression Discontinuity Design

We are interested in the impact of being a incumbent senator on the probability of being reelected in the US. Our outcome is the vote share of Democrats in the following election for the same seat, while our running (or forcing) variable is the margin of victory/defeat in the current election, with the cut-off being zero (winning or losing the current election).

We have information on which class of election the seat belongs to and on the experience of the Democratic candidate, both in the US House and in the US Senate. We also have data on the amount of people that seat represents.

1.1 Summary Statistics

We first show some summary statistics. We have 1,390 observations in total, but only the first 1,297 have information on `vote`. Information with complete information ranges from 1914 to 2004.

A few interesting observations: the median Democratic candidate for the senate has already served 4 terms, with no one with no experience in the Senate.

Table 1: Summary Statistics of Numeric Variables

	Vote	Margin	Terms in House	Terms in Senate	Population
Minimum	0	-100	0	1	78,000
1st Quartile	42.67	-12.206	0	1	997,056
Median	50.55	2.166	0	4	2,554,388
Mean	52.67	7.171	1.437	4.556	3,827,919
3rd Quartile	61.35	22.766	2	7	4,663,226
Maximum	100	100	16	20	37,253,956
NAs	93	0	282	282	0

Table 2: Frequency of Most Common Values of Categorical Variables

State	Year	Class
52:33	2010:34	1:455
61:33	1974:33	2:448
67:33	1978:33	3:487
53:32	1980:33	
25:30	1984:33	
37:30	2002:33	

Note: Table read as *value:frequency.*

1.2 Graphical Analysis

Using the `rdplot` function of the `rdrobust` **R** package, we plot the conditional expectation of `vote` as a function of `margin` using bins in Figure 1. We include 95% confidence intervals for each bin and a global regression function (dashed red line). We see a clear discontinuity at the cut-off $c = 0$.

The default kernel of the package is the uniform, which we keep. Bandwidth is chosen using the optimal Integrated Mean Squared Error (IMSE) of Calonico, Cattaneo and Titiunik (2015) based on spacing estimators (`binselect = "es"`).

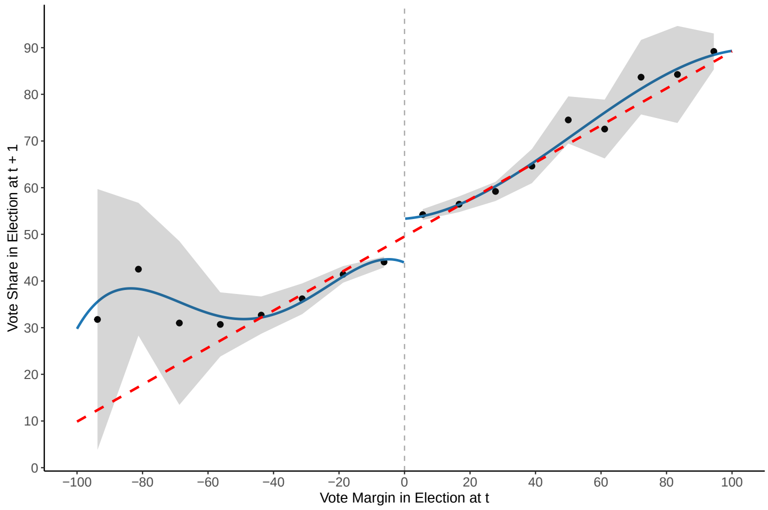

Figure 1: RD Plot with Evenly Spaced, IMSE-Optimal Bins

1.3 Treatment Effect

Note that we are in a sharp RDD setting: only those with $margin > 0$ get elected, which is our treatment. To estimate the treatment effect, we will use the `rdrobust` function. We assume that the RD assumption holds, namely, that we have a quasi-experiment around the threshold.

We use `rdrobust` by specifying `y = vote` and `x = margin`, with `c = 0`. We use the Epanechnikov kernel and a local-polynomial of order `p = 2` to construct the point-estimator and `q = p + 1 = 3` to construct the bias-correction (which are not the default). Bandwidth is automatically selected based on a MSE-optimal procedure. The variance-covariance matrix is estimated using a heteroskedasticity-robust nearest neighbor estimator (the default option).

Results are shown in Table 3. We see that having won the previous election has a significant effect on the chance of winning the next: a Democratic incumbent with a positive margin close to zero has a 7.86 p.p. higher voter share than a "loser" with a negative margin close to zero at the previous election. Results are significant with both conventional and robust confidence intervals.

Table 3: Sharp RDD Results

	Coef.	Std. Err.	z	$\mathbb{P} > \|z\|$	95% C.I.
Conventional SEs	7.863	1.917	4.101	0.000	[4.105, 11.621]
Robust SEs	-	-	3.839	0.000	[3.999, 12.343]

Note: Epanechnikov kernel, $p = 2$, $q = 3$.

1.4 Density Tests

We implement McCrary (2008)'s Density Test using the package and function `rddensity`. The null of the test is of no-manipulation of the running variable.

We use a Epanechnikov kernel and a local-polynomial of order `p = 2` to construct the density-estimator and `q = 3` to construct the bias-correction. Using jackknife standard errors, we get a test-statistic of -0.89 and a p-value of 0.37.

Thus, we don't reject the null of no-manipulation. This makes sense in our context, given that the margin variable is a result of a de-centralized decision of voters, and not of the candidates themselves, although they can try and influence it with political promises.

1.5 Density Plots

Using the same parameters as the previous item, the density plot of *margin* does not a show significant discontinuity at the threshold $c = 0$. This further confirms the evidence of no manipulation and can be seen in Figure 2.

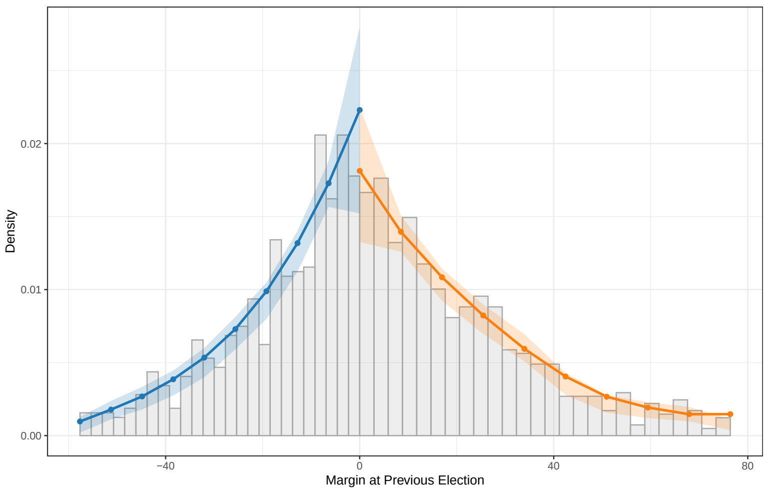

Figure 2: Density Plot of the Margin of Votes

1.6 Balance Test of Covariates Around the Threshold

We will use the permutation test proposed by Canay and Kamat (2017). We will use the `RDperm` function from the `RATest` package. This procedure can be seen as an informal placebo test of the Continuity assumption: if covariates aren't discontinuous around the cut-off, potential outcomes are likely to be continuous as well. In turn, if covariates are discontinuous, it suggests that something is off.

Note that `class`, although numerically coded, indicates a qualitative information. For this reason, we create three indicator columns and use these for the tests.

For `population` and $\mathbb{1}$ `{class}` $= i$, $i = 1, 2, 3$, we get a joint p-value of 0.43, so we have no evidence of discontinuity in our covariates. Furthermore, all individuals p-values are well above any usual significance level.[1]

To include `termshouse` and `termssenate`, we remove all 282 observations with missing values, leaving us with 1,108 observations. Including these covariates alongside the previous two, we get a joint p-value of 0.10: we still don't have strong evidence against the null that covariates are continuous.[2]

[1]Conclusions are unchanged when treating `class` as numeric in the permutation test, with a joint p-value of 0.51.
[2]When treating `class` as numeric, we get a joint p-value of 0.28.