Replication 5 - Discontinuities

Marina Merlo

14 de maio de 2019

1. and 2. Preparing the 2000 dataset

```
data2000 <- cepespR::get_elections(year=2000,
                                   position="Prefeito",
                                   regional_aggregation="Municipality",
                                   political_aggregation="Partido") %>%
  filter(NUM_TURNO == 1) %>%
  group_by(COD_MUN_IBGE)%>%
  mutate(totvotos00 = sum(QTDE_VOTOS)) %>%
  ungroup() %>%
  rowwise() %>%
  mutate(voteshare00 = QTDE_VOTOS/totvotos00) %>%
  group_by(COD_MUN_IBGE) %>%
  mutate(rank_2000 = rank(-QTDE_VOTOS)) %>%
  filter(rank_2000 %in% c(1,2)) %>%
  mutate(incumbency = ifelse(rank_2000 == 1, "1", "0")) %>%
  arrange(COD_MUN_IBGE, rank_2000) %>%
  mutate(first_rank_vote_pct = max(voteshare00),
         second_rank_vote_pct = nth(voteshare00,2),
         first_rank_vote = max(QTDE_VOTOS),
         second_rank_vote = nth(QTDE_VOTOS,2)) %>%
  mutate(win_margin = ifelse(rank_2000 == 1,
                             voteshare00 - second_rank_vote_pct,
                             voteshare00 - first_rank_vote_pct)) %>%
  filter(first_rank_vote != second_rank_vote) %>%
    ungroup()
```

3. Preparing the 2004 dataset

```
data2004 <- cepespR::get_elections(year=2004,
                                   position="Prefeito",
                                   regional_aggregation="Municipality",
                                   political_aggregation="Partido") %>%
  filter(NUM_TURNO == 1) %>%
  group_by(COD_MUN_IBGE)%>%
  mutate(totvotos04 = sum(QTDE_VOTOS)) %>%
  ungroup() %>%
  rowwise() %>%
  mutate(voteshare04 = QTDE_VOTOS/totvotos04) %>%
  dplyr::select(COD_MUN_IBGE,NUMERO_PARTIDO,voteshare04)
```

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?)

We do a left join, so we keep all the observations from 2000 and only add the 2004 vote share for each observation.

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.

Considering only the PMDB candidates, there's an increase of 1.7% in the vote share for the incumbents, significant at 5% level. This means that for the PMDB candidates there's a positive incumbent effect.

Table 1:

	Dependent variable:
	voteshare04
incumbency1	0.017**
	(0.007)
Constant	0.438***
	(0.005)
Observations	1,415
R^2	0.004
Adjusted R^2	0.003
Residual Std. Error	0.137 (df = 1413)
F Statistic	5.385** (df = 1; 1413)
Note:	*p<0.1; **p<0.05; ***p<0.01

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%.

The difference between treated (incumbent) and not treated in the number of voters is really close within the 1% margin - incumbents receive only 64 more votes than not incumbents, on average. However, within the 3% margin, this difference increase to 211 more votes for the incumbent, on average.

Table 2: +/-1% winning margin 2004 voters mean

incumbency	mean
0	5429.343
1	5493.155

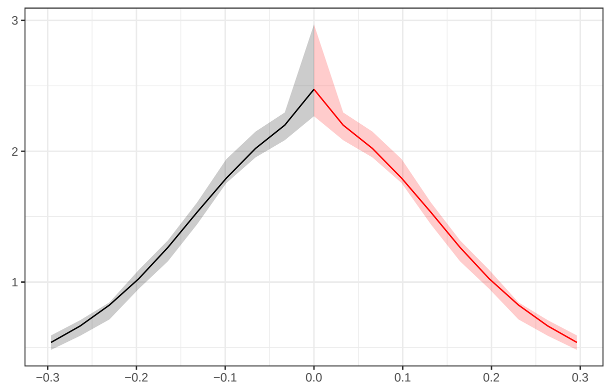
Table 3: +/-3% winning margin 2004 voters mean

incumbency	mean
0	5512.836
1	5723.098

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.

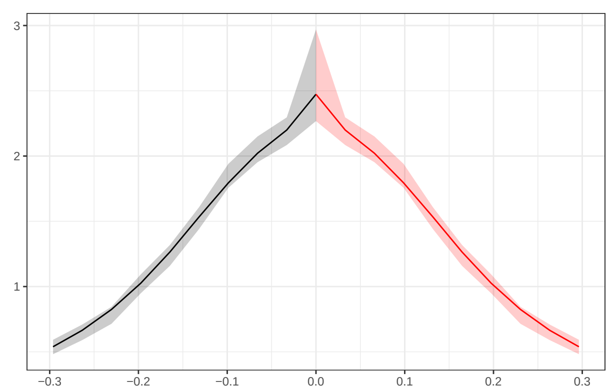
The number of observations is evenly distributed around the cutoff. The indicated bandwidth is at 0.099%. The plot shows that there were no manipulation from the units to be under or above the cutoff.

```
##
## RD Manipulation Test using local polynomial density estimation.
##
## Number of obs =       9754
## Model =               unrestricted
## Kernel =              triangular
## BW method =           comb
## VCE method =          jackknife
##
## Cutoff c = 0          Left of c           Right of c
## Number of obs         4877                4877
## Eff. Number of obs    2083                2083
## Order est. (p)        2                   2
## Order bias (q)        3                   3
## BW est. (h)           0.099               0.099
##
## Method                T                   P > |T|
## Robust                0                   1
```



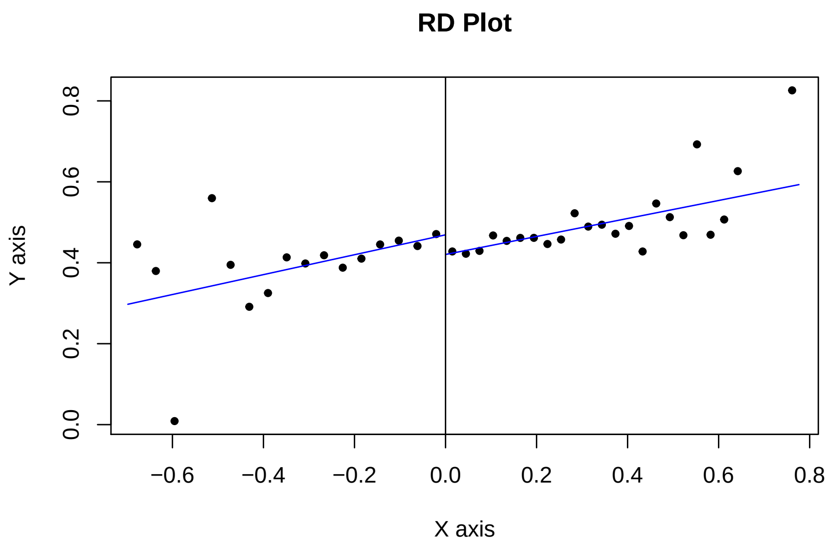

```
## $Estl
## Call: lpdensity
##
## Sample size                                      4877
## Polynomial order for point estimation    (p=)    2
```

```
## Order of derivative estimated           (v=)    1
## Polynomial order for confidence interval (q=)    3
## Kernel function                                  triangular
## Scaling factor                                   0.499948733722957
## Bandwidth method                                 user provided
##
## Use summary(...) to show estimates.
##
## $Estr
## Call: lpdensity
##
## Sample size                                      4877
## Polynomial order for point estimation    (p=)    2
## Order of derivative estimated            (v=)    1
## Polynomial order for confidence interval (q=)    3
## Kernel function                                  triangular
## Scaling factor                                   0.499948733722957
## Bandwidth method                                 user provided
##
## Use summary(...) to show estimates.
##
## $Estplot
```



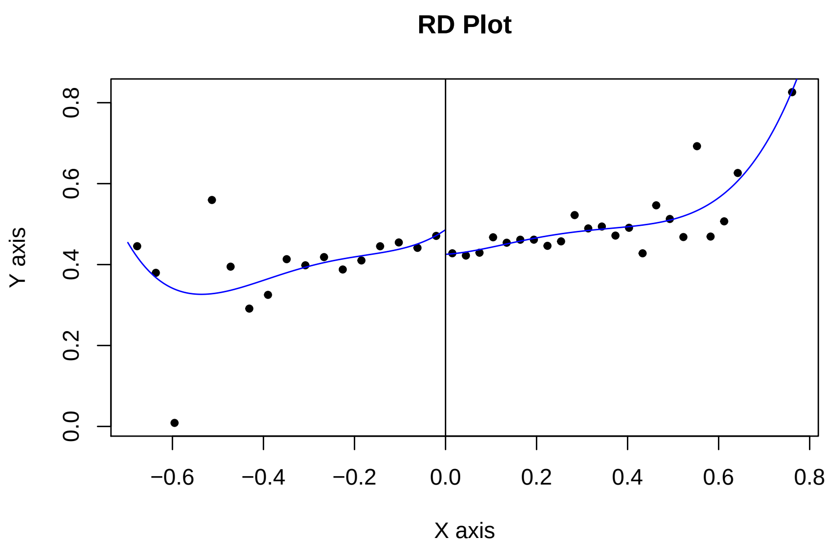
8. Before we run the analysis, let's construct a regression discontinuity plot to visually inspect the causal effect of incumbency at the threshold. Using a pre-packaged command like rdplot from the rdrobust package, create a regression discontinuity plot for the effect of incumbencyin 2000 on vote share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

There's a small discontinuity around the cutoff, with lower vote shares for the incumbents.



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

Even with the fourth-order polynomial regression line, the discontinuity around the cutoff remains. This means the difference seen 0 isn't due the data distribution.



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.

The difference in the vote share in 2004 between incumbents and challengers for PMDB candidates is different from zero at 1% significance level. Now that we limited our observations for those with +/-3% winning margin, the incumbent effect seen in Q6 becomes negative: incumbents receive 42,8% votes on average while challengers get 48,2%, a difference of 5,4%.

```
##
##  Welch Two Sample t-test
##
## data:  voteshare04 by incumbency
## t = 3.0003, df = 229.08, p-value = 0.002995
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.01859042 0.08972182
## sample estimates:
## mean in group 0 mean in group 1
##       0.4820535       0.4278974
```

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

The full RDD for the PMDB candidates show a negative incubency effect of 4,8% in the 2004 votshare. This is slightly smaller than the t.test because now we're accounting for the effect of the running variable on the outcome.

Table 4:

	Dependent variable:
	voteshare04
incumbency1	-0.048^{***}
	(0.011)
win_margin	0.231^{***}
	(0.028)
Constant	0.467^{***}
	(0.006)
Observations	1,415
R^2	0.048
Adjusted R^2	0.047
Residual Std. Error	0.134 (df = 1412)
F Statistic	35.734^{***} (df = 2; 1412)
Note:	*p<0.1; **p<0.05; ***p<0.01

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.

The negative effect of incumbency now it's even smaller, with a loss of 2.9% of vote share. Our observation number dropped to 236 when filtering for only those within the bandwidth of +/-3%, but we have less bias from the candidates that won or lost from a larger margin.

Table 5:

	Dependent variable:
	voteshare04
incumbency1	-0.029
	(0.037)
win_margin	-0.836
	(1.073)
Constant	0.468^{***}
	(0.022)
Observations	236
R^2	0.040
Adjusted R^2	0.031
Residual Std. Error	0.138 (df = 233)
F Statistic	4.820^{***} (df = 2; 233)
Note:	*p<0.1; **p<0.05; ***p<0.01

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?

The selected bandwidth was +/-13.5% winning margin. The results are closer to those of the t.test, estimating a negative effect of 5.6% for the incumbent

```
## Call: rdrobust
##
## Number of Obs.                 1415
## BW type                       mserd
## Kernel                   Triangular
## VCE method                       NN
##
## Number of Obs.                  649          766
## Eff. Number of Obs.             415          417
## Order est. (p)                    1            1
## Order bias  (p)                   2            2
## BW est. (h)                   0.135        0.135
## BW bias (b)                   0.216        0.216
## rho (h/b)                     0.624        0.624
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
```

```
##   Conventional    -0.056     0.020    -2.790     0.005    [-0.096 , -0.017]
##         Robust         -         -    -2.432     0.015    [-0.106 , -0.011]
## =============================================================================
```

14. Now let's try to adjust the functional form used to estimate the effect of the running 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.

```
## Call: rdrobust
##
## Number of Obs.                 1415
## BW type                       mserd
## Kernel                   Triangular
## VCE method                       NN
##
## Number of Obs.                 649         766
## Eff. Number of Obs.            516         560
## Order est. (p)                   2           2
## Order bias  (p)                  3           3
## BW est. (h)                  0.202       0.202
## BW bias (b)                  0.288       0.288
## rho (h/b)                    0.702       0.702
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional    -0.059     0.025    -2.370     0.018    [-0.107 , -0.010]
##         Robust         -         -    -2.122     0.034    [-0.115 , -0.005]
## =============================================================================
```

```
## Call: rdrobust
##
## Number of Obs.                 1415
## BW type                       mserd
## Kernel                   Triangular
## VCE method                       NN
##
## Number of Obs.                 649         766
## Eff. Number of Obs.            583         641
## Order est. (p)                   4           4
## Order bias  (p)                  5           5
## BW est. (h)                  0.275       0.275
## BW bias (b)                  0.351       0.351
## rho (h/b)                    0.784       0.784
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional    -0.062     0.037    -1.675     0.094    [-0.135 , 0.011]
##         Robust         -         -    -1.531     0.126    [-0.141 , 0.017]
## =============================================================================
```

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 theother two parties: the PFL and the PSDB. How similar are your results to those in Titiunik(2011) for the +/-3% window?

Table 6:

	Dependent variable:	
	voteshare04	
	PFL	PSDB
	(1)	(2)
incumbency1	-0.109^{**}	0.129^{**}
	(0.052)	(0.052)
win_margin	2.421	-3.818^{**}
	(1.503)	(1.604)
Constant	0.511^{***}	0.356^{***}
	(0.026)	(0.030)
Observations	129	110
R^2	0.036	0.056
Adjusted R^2	0.020	0.038
Residual Std. Error	0.143 (df = 126)	0.141 (df = 107)
F Statistic	2.324 (df = 2; 126)	3.173^{**} (df = 2; 107)
Note:		*p<0.1; **p<0.05; ***p<0.01