Regression Discontinuity Design

Lee's (2008) paper: "Randomized experiments from nonrandom selection in U.S. House elections".

Lee (2008) aims at estimating the incumbency advantage in elections to the U.S. House of Representatives at the Congressional district level by using a sharp RD design.

An OLS regression of incumbency status on election success is likely to be biased because of unobserved differences (i.e. incumbents have already won an election so they may just be better). Thus, a sharp RD is a nonexperimental technique that proves useful to determine a causal link.

The identification strategy relies on comparing districts where Democratic party barely won an election and hence barely became an incumbent with districts where Democratic party barely lost (and a Republican party won). The outcome variable of interest is the party's vote share in subsequent election. All variables are defined for the Democratic party only.

In the main dataset (TP8rdd.dta), one finds the following variables:

mov: Democratic vote share margin of victory, election t
demsharenext: Democrat vote share, election $t+1$
demshareprev: Democrat vote share, election $t-1$
demwinprev: indicator, 1 if Democratic party won, election t
demofficeexp: Democrat political experience, as of election t
othofficeexp: Opposition political experience, as of election t
othelecexp: Opposition electoral experience, as of election t
statedisdec: State, district, decade clusters

Note: Two other datasets are use in the appendix of this TP, which are the individual candidate's data and the democratic group (i.e. party) data (without clustering by state-district-decade). These two datasets are use to recreate the RD graphs in Lee's paper for illustrative reasons.

The main idea of this paper is: What is the likelihood of winning next election given that the democratic party is incumbent?
For the non-parametric estimations please install rdrobust, rdlocrand and rddensity commands.

1 Definition of treatment

We use the aggregated group democratic dataset clustered at the state, district and decade level. Please open the *TP8rdd.dta* dataset.

1.1) First, lets generate the treatment indicator variable, *victory*, i.e. a dummy for whether the Democrats won the election. Assume that there are only two parties Democrats and Republicans. Explain the assignment mechanism.

The variable *victory* takes the value one if the variable of margin of victory, i.e. mov, is above zero. We are interested in studying the causal impact of a Democratic electoral victory in a U.S. Congressional District race on the probability of future Democratic electoral success.

Recall that in a Sharp RD design, you exploit the fact that the treatment status is a deterministic and discontinuous function of a covariate x_i. In the case of this exercise, it is victory around the 50% vote threshold, or $mov = 0$. That is, Democratic electoral victory is a deterministic function of the vote share, and within a certain threshold, those districts where Democrats won with a victory slightly above 50%, are comparable (as if there was a random assignment) to those where democrats lost by a close margin.

2 RD graph using local averages (bins), and graphs using polynomial fit with all data

2.1) Plot local averages of the outcome variable against generated cells of the assignment variable on either side of the threshold using bins that calculate the average value of the outcome variable in each cell (e.g. 60 bins).[*Hint*: Use *binscatter* command]. What do you observe?

With the data in bins, we clearly see a jump in the regression at the cutoff value of zero of the running variable *mov*.

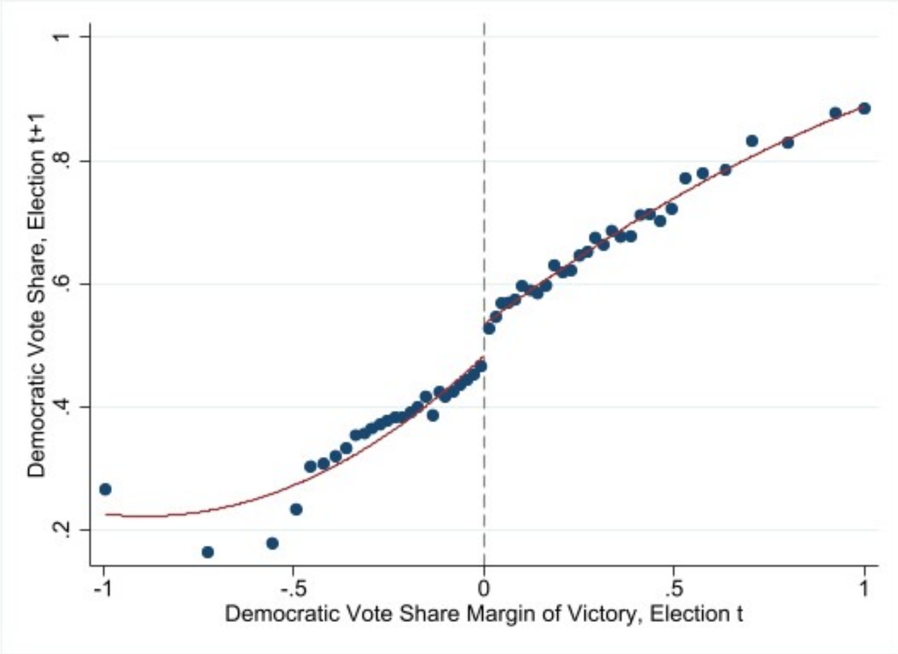

Figure 1: Binscatter Plot

2.2) Plot polynomial fits (e.g. 1st degree) using all data (i.e. no bins) of the outcome variable against the assignment variable. [*Hint*: Use a two-way graph with local polynomial smoothing]. Do you still observe the jump in the regression?

Likewise, with a local polynomial, we still observe the jump in the regression at the cutoff value of zero of the running variable, *mov*.
(Note: For more on how the graphs are generated with bins, see class slides 29-33 of Ch 7).

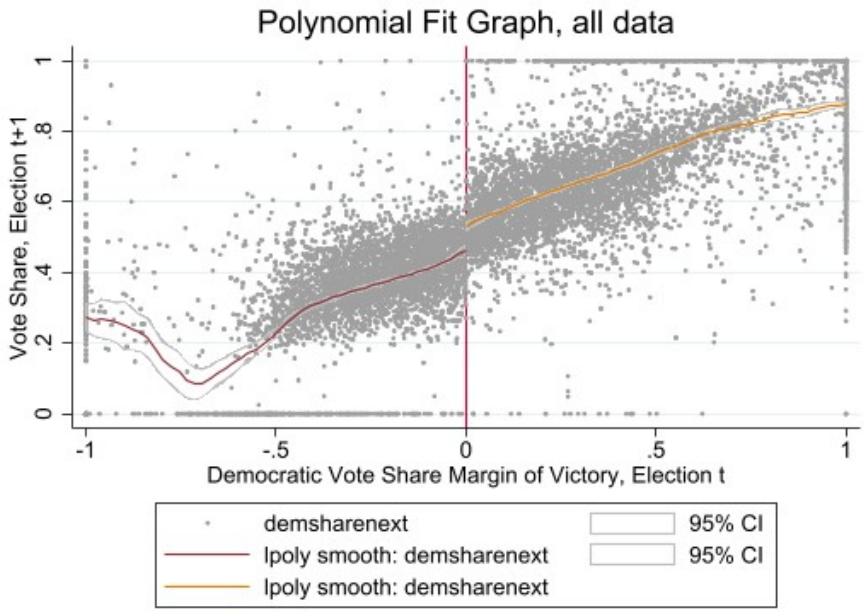

Figure 2: Polynomial Reg Plot

3 RD estimation of party incumbency advantage

3.1) Use polynomial approximations to generate estimates of the discontinuity gap. That is, estimate the incumbency effect by regressing the outcome variable (i.e. Democrat vote share in subsequent election), -not local average of this variable-, on a constant, the victory indicator, and a fourth-order polynomial in the assignment variable (i.e. *mov*), and their interactions. [*Hint*: You have to generate second-, third-, and fourthorder polynomial in *mov*]. Interpret the results.
The sharp RD Regression equation that we are estimating is of the following form:

$$Y_i = \alpha + \beta_{01}\tilde{x}_i + \beta_{02}\tilde{x}_i^p + \cdots + \beta_{0p}\tilde{x}_i^p + \rho D_i + \beta_1^* D_i\tilde{x}_i + \beta_2^* D_i\tilde{x}_i^2 + \cdots + \beta_p^* D_i\tilde{x}_i^p + \eta_i \quad (1)$$

What is the magnitude of the treatment effect coefficient that we are interested in our case? [*Hint*: Recall from your slides, that the treatment effect X_0 is ρ.]

Table 1: Results for polynomial 4th degree

	(1)
	demsharenext
mov	0.524***
	(0.146)
mov2	1.530*
	(0.753)
mov3	4.221**
	(1.369)
mov4	3.046***
	(0.770)
rmov	0.0341
	(0.196)
rmov2	-2.290*
	(0.956)
rmov3	-2.909
	(1.627)
rmov4	-3.811***
	(0.911)
victory	0.0755***
	(0.0114)
Constant	0.454***
	(0.00798)
Observations	6559
R^2	0.684

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The ρ coefficient, i.e. the victory dummy coefficient, is 0.076 with a std.error of 0.011. That is, the difference in electoral outcome in the next election between a democrat winner and a looser in t at the cutoff value is 7.5% measured in terms of vote share margin. This result is significant at a 1% level.

3.2) Using a non-parametric kernel estimation (i.e. local linear regression), estimate the effect of incumbency advantage. That is, use a non-parametric estimation with polynomial of order 1. What regression equation are we estimating? [*Hint*: You have this equation in your slides.] Plot the RD regressions. What do you observe compared to the results of the parametric regression in 3.1?

With the Kernel method, we estimate the following model using local linear regression:

$$Y_i = \alpha + \rho D_i + \beta_{01}\tilde{x}_i + \beta_1^* D_i \tilde{x}_i + \eta_i \tag{2}$$

Where $\tilde{x}_i = X_i - X_0$ and X_0 is the cutoff value of the running variable (in our case mov). From the lecture slides (Ch 7, slides Non Parametric Model), we know that we can estimate the average effect directly in a single local linear regression, as depicted in equation (2). To do so, one minimizes the square of the residuals. That is, one solves:

$$\min_{\alpha,\rho,\beta_{01},\beta_1} \sum_{n=1}^{N} \left[Y_i - \alpha + \rho D_i + \beta_{01}\tilde{x}_i + \beta_1^* D_i \tilde{x}_i \right]^2 \times \mathbf{1}[X_0 - h \leq X_i < X_0 + h] \tag{3}$$

While estimating this equation in a given window of width h around the cutoff is straightforward, the difficulty relies in choosing this bandwidth. We then face essentially a trade-off between bias and efficiency.

Going back to the results, the Local Linear Regression RD estimate (i.e. non-parametric RD of degree 1) is 0.062 with std. errors of 0.012, which is similar to the parametric estimate in 3.1 (i.e. 0.075 with std.error of 0.011).

Likewise, as in the parametric case, we observe a jump in the regression at the cutoff value of $mov = 0$.

When we use the covariates in the regression, the value of the RD estimate is 0.061 with a std. error of 0.011.

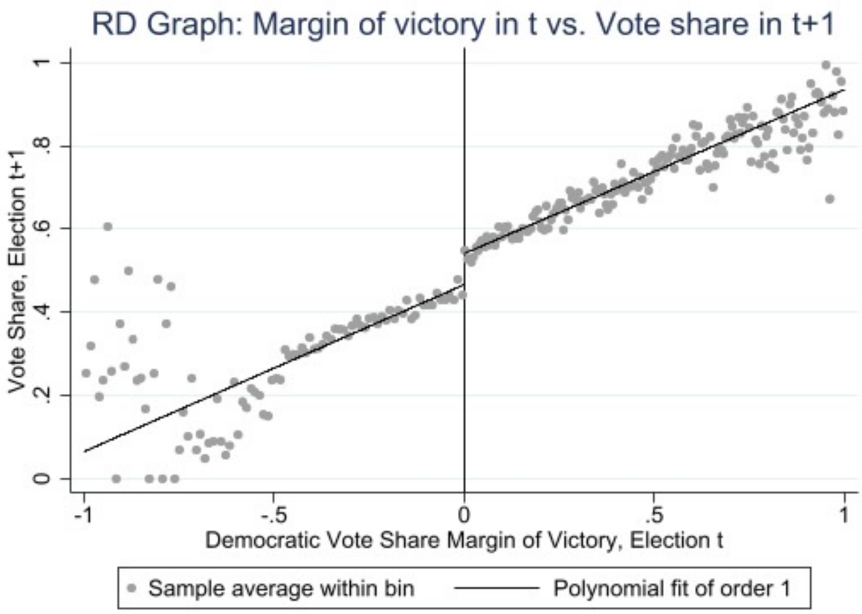

Figure 3: Local Linear Reg Kernel Estimation

Table 2: Results for non parametric

	(1)	(2)
	demsharenext	demsharenext
RD_Estimate	0.0616***	0.0616***
	(0.0118)	(0.0113)
Observations	6559	6559
R^2		

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

3.3) Use the model specification from 3.1 and explore the sensitivity of the results by including the following baseline covariates: *demshareprev, demwinprev, demofficeexp, demelectexp, othofficeexp, othelectexp*. How does the estimated effect on party incumbency change?

The coefficient of Victory changes very slightly: it is now 0.077 with a sd. error of 0.011.

Table 3: Results for polynomial 4th degree with controls

	(1)
	demsharenext
mov	0.425**
	(0.138)
mov2	1.229
	(0.700)
mov3	3.235*
	(1.259)
mov4	2.304**
	(0.701)
rmov	-0.0357
	(0.187)
rmov2	-1.607
	(0.898)
rmov3	-2.631
	(1.510)
rmov4	-2.676**
	(0.837)
victory	0.0766***
	(0.0109)
demshareprev	0.298***
	(0.0171)
demwinprev	-0.00659
	(0.00703)
demofficeexp	-0.000165
	(0.00273)
othofficeexp	-0.000234
	(0.00368)
demelectexp	-0.00291
	(0.00263)
othelectexp	0.00333
	(0.00353)
Constant	0.313***
	(0.0111)
Observations	6559
R^2	0.712

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

4 Sensitivity of RD estimates to Different order of Polynomials

4.1) Estimate the incumbency effect by regressing Democrat vote share in subsequent election on a constant, the victory indicator, a second order polynomial in *mov* and their interactions. Do the same for a third-order, 4th order, 5th order and 6th order polynomials.

Table 4: Results for different polynomial specifications

	(1)	(2)	(3)	(4)	(5)	(6)
	demsharenext	demsharenext	demsharenext	demsharenext	demsharenext	demsharenext
victory	0.118***	0.0515***	0.111***	0.0755***	0.0417**	0.0650***
	(0.00600)	(0.00732)	(0.00970)	(0.0114)	(0.0134)	(0.0147)
mov	0.297***	0.632***	-0.0971	0.524***	1.427***	0.266
	(0.0177)	(0.0370)	(0.0876)	(0.146)	(0.233)	(0.334)
rmov	0.0463*	-0.153***	0.458***	0.0341	-0.641*	0.598
	(0.0200)	(0.0465)	(0.110)	(0.196)	(0.318)	(0.478)
mov2		0.380***	-1.718***	1.530*	8.685***	-4.305
		(0.0437)	(0.267)	(0.753)	(1.757)	(3.520)
rmov2		-0.509***	1.898***	-2.290*	-11.15***	1.025
		(0.0511)	(0.320)	(0.956)	(2.344)	(4.843)
mov3			-1.463***	4.221**	25.50***	-32.12*
			(0.196)	(1.369)	(5.217)	(15.58)
rmov3			1.260***	-2.909	-19.43**	41.63*
			(0.224)	(1.627)	(6.584)	(20.35)
mov4				3.046***	29.16***	-88.87**
				(0.770)	(6.454)	(32.42)
rmov4				-3.811***	-35.45***	75.87
				(0.911)	(8.154)	(41.97)
mov5					11.13***	-100.4**
					(2.789)	(31.28)
rmov5					-8.874**	108.8**
					(3.387)	(38.78)
mov6						-39.31***
						(11.25)
rmov6						37.24**
						(14.02)
Constant	0.433***	0.478***	0.428***	0.454***	0.481***	0.456***
	(0.00468)	(0.00538)	(0.00700)	(0.00798)	(0.00923)	(0.0101)
Observations	6559	6559	6559	6559	6559	6559
R^2	0.671	0.678	0.683	0.684	0.685	0.685

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

4.2) Plot the different RD coefficients of the regression equations in the previous question (4.1) to see how the estimates vary according to a different polynomial degree. Comment the results.

Polynomial order changes the RD estimate, and seems to be a bit more stable around $p = 4$ (the order used in the paper).

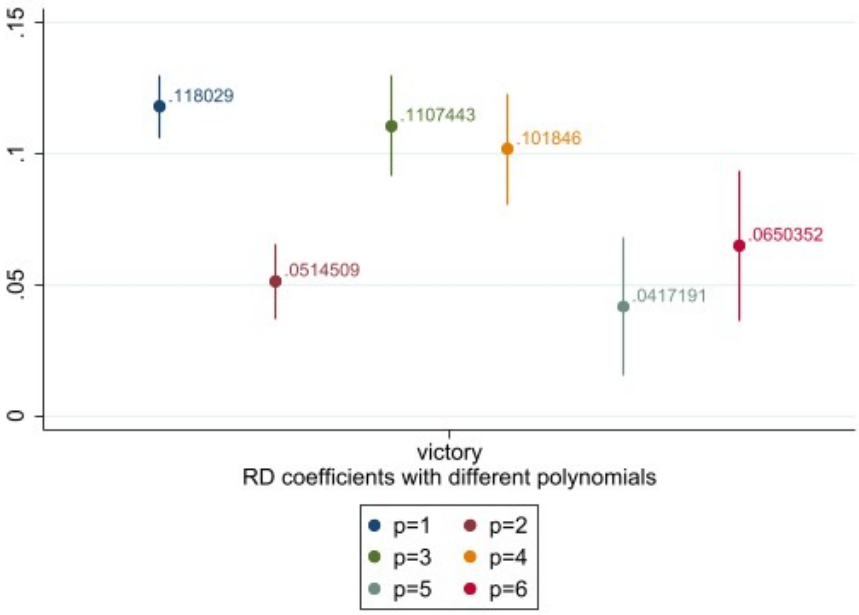

Figure 4: Coefficient Plot for parametric estimation of RD coefs with different Polynomial order

4.3) Conduct the same exercise as in 4.1, but now using a non-parametric kernel estimation with polynomial of order 2. Comment.

Using an order 2 polynomial for the non-parametric regression, we find a similar effect, but slightly higher: 0.067 with st. error of 0.001

Table 5: Results for non parametric

	(1)
	demsharenext
RD_Estimate	0.0647***
	(0.0115)
Observations	6559
R^2	

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

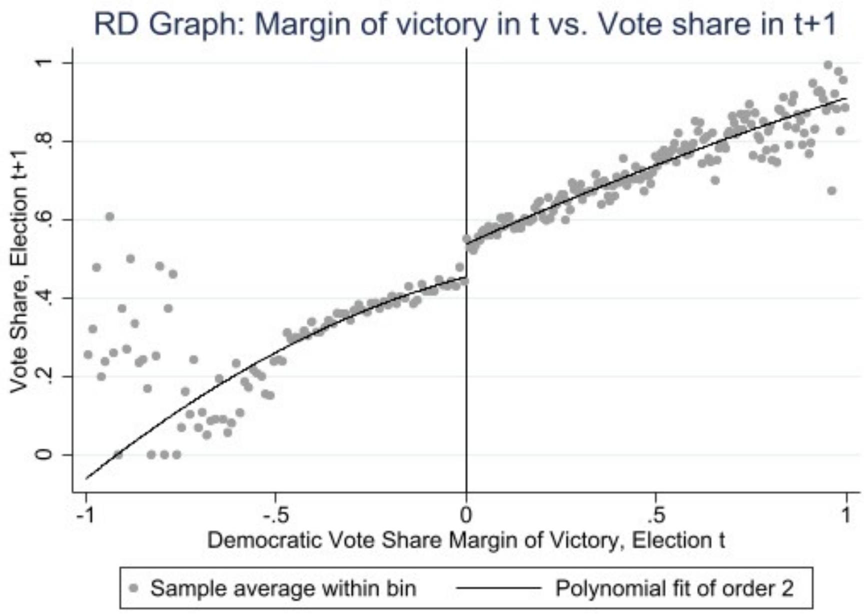

Figure 5: Kernel Estimation order 2

5 Validation and Placebo cutoffs

5.1) **Validation**. Conduct a parallel RDD analysis on the baseline covariates from question 3.1 (i.e. using a fourth-order polynomial in *mov*). What would you expect? What are the results?

We expect that the baseline covariates, that is that pre-determined characteristics (like candidate's political experience) are balanced in a neighborhood of the discontinuity threshold. This means that when we regress each of the covariates with respect to the polynomial of order 4 in *mov*, we should not find a significant effect of the dummy variable victory. When we run the loop to regress each covariate with respect to "fitpoly4", indeed, the treatment dummy is not significant in any of the regressions.

Table 6: Results for validation of controls

	(1) demshareprev	(2) demwinprev	(3) demofficeexp	(4) othofficeexp	(5) demelectexp	(6) othelectexp
victory	-0.00521	-0.0114	0.0302	0.125	0.00845	0.115
	(0.0133)	(0.0422)	(0.267)	(0.205)	(0.271)	(0.211)
Observations	6559	6559	6559	6559	6559	6559
R^2	0.640	0.684	0.305	0.352	0.286	0.335

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.2) Conduct the same test as in 5.1 using a non-parametric kernel estimation (i.e. regressing the outcome variable with respect to other covariates, and estimate original regression establishing placebo cutoffs).

As expected, and in the same manner as in the case of the parametric estimation, none of the covariates when regressed against the margin of victory variable, have any significant effect.

Table 7: Results for validation of controls in RD

	(1) demshareprev	(2) demwinprev	(3) demofficeexp	(4) othofficeexp	(5) demelectexp	(6) othelectexp
RD_Estimate	0.00622	0.0377	0.218	-0.0219	0.223	-0.0113
	(0.0107)	(0.0463)	(0.188)	(0.146)	(0.190)	(0.149)
Observations	9177	9177	11795	11795	11795	11795
R^2						

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.3) **Placebo cutoffs**. Generate two different cut-offs variable for victory (i.e. +0.25 and +0.15), and run the same model specification from question 4 (i.e. fourth-order polynomial in the assignment variable *mov*). Comment.

For our placebo cutoffs of +0.25, our pseudo treatment variable Victory25 displays a coefficient of 0.217, but it is insignificant. Likewise, for our placebo cutoff of +0.15, our Victory15 pseudo treatment variable displays a coefficient of 0.087, but it is insignificant.

Table 8: Results for placebo estimations

	(1) demsharenext	(2) demsharenext
victory25	0.217	
	(0.226)	
victory15		0.0872
		(0.0630)
Observations	6559	6559
R^2	0.678	0.680

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.4) Plot the local averages and the polynomial fits (e.g. 1st degree) of the outcome variable against generated cells of the assignment variable using these placebo cutoffs (i.e. +0.25 and +0.15). What do you observe?

In both graphs (i.e. binscatter and polynomial graph using all data), there is no significant jump at these placebo cutoffs values of +0.25 and +0.15).

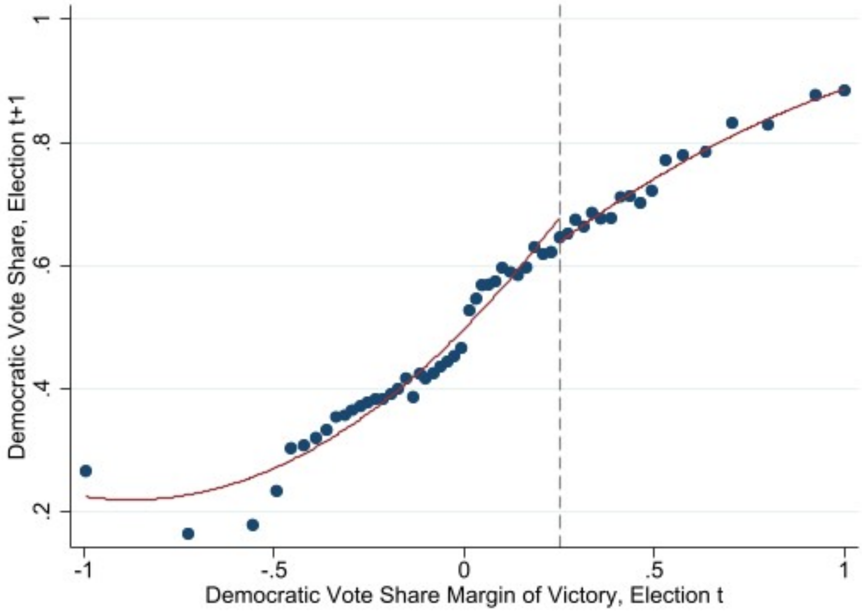

Figure 6: Binscatter plot with Placebo cutoff at +0.25

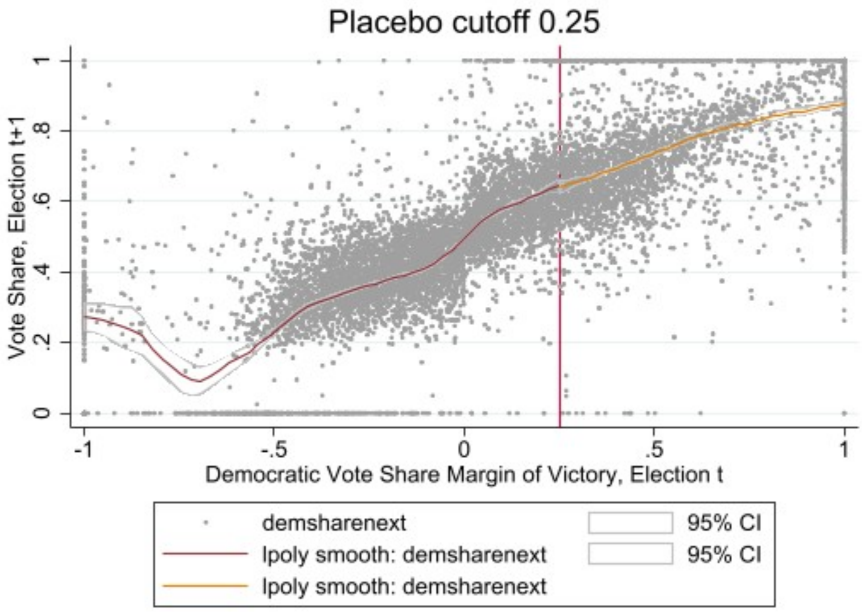

Figure 7: Polynomial plot with Placebo cutoff at +0.25

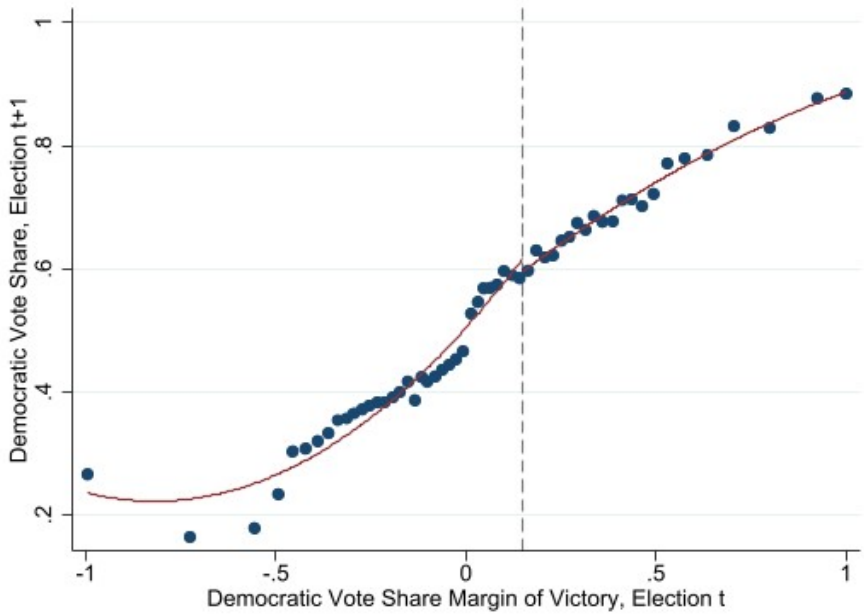

Figure 8: Binscatter plot with Placebo cutoff at +0.15

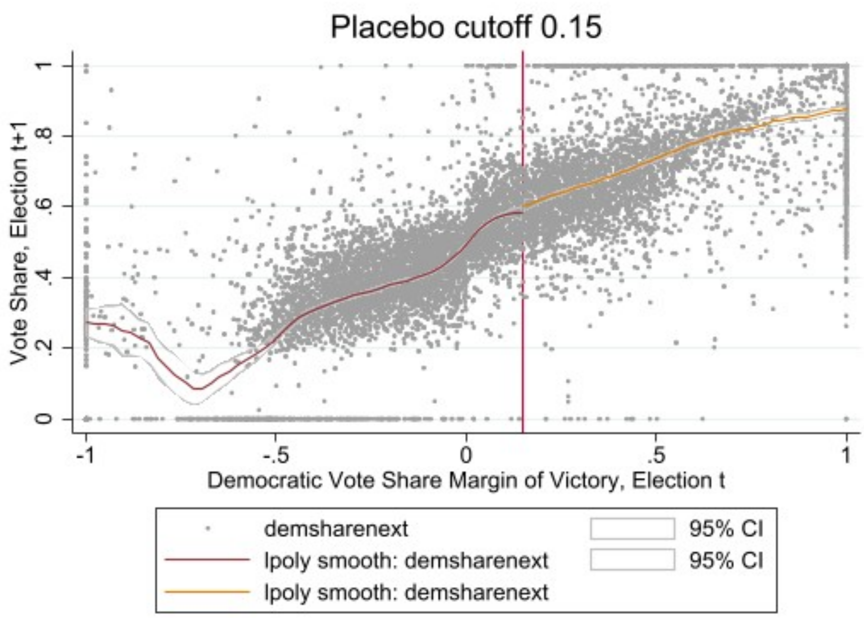

Figure 9: Polynomial plot with Placebo cutoff at +0.15

5.5) Conduct the same placebo test as in 5.3 using a non-parametric kernel estimation.

With regards to the Placebo cutoffs, when we conduct the RD nonparametric estimation using these Placebo cutoffs of +0.25 and +0.15, the treatment variable is not significant. Therefore, as we can see in the plots,there are no jumps in the regression at the placebo cutoffs we established.

Table 9: Results for non parametric placebo

	(1)	(2)
	demsharenext - cutoff 15	demsharenext - cutoff 25
RD_Estimate	-0.0114	-0.00601
	(0.00976)	(0.0140)
Observations	6559	6559
R^2		

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

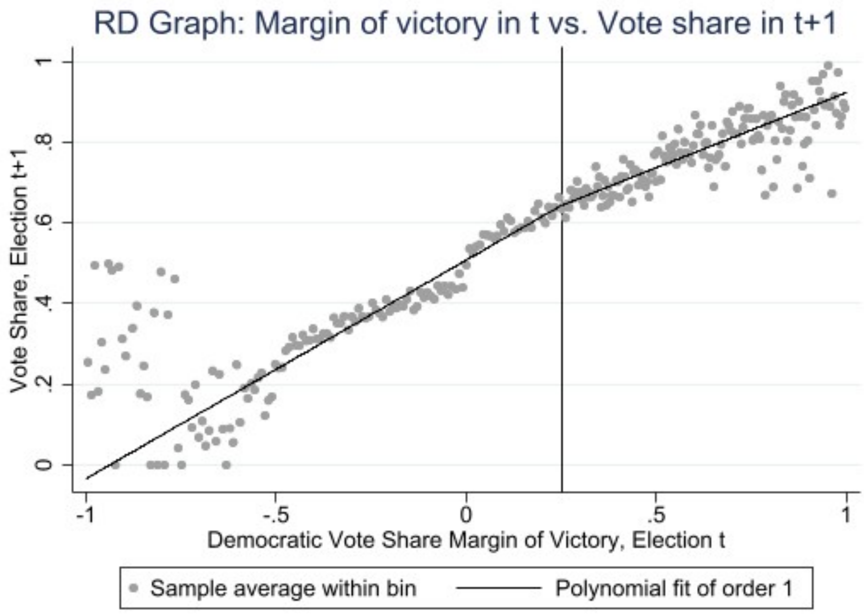

Figure 10: Non parametric estimation with placebo cutoff +0.25

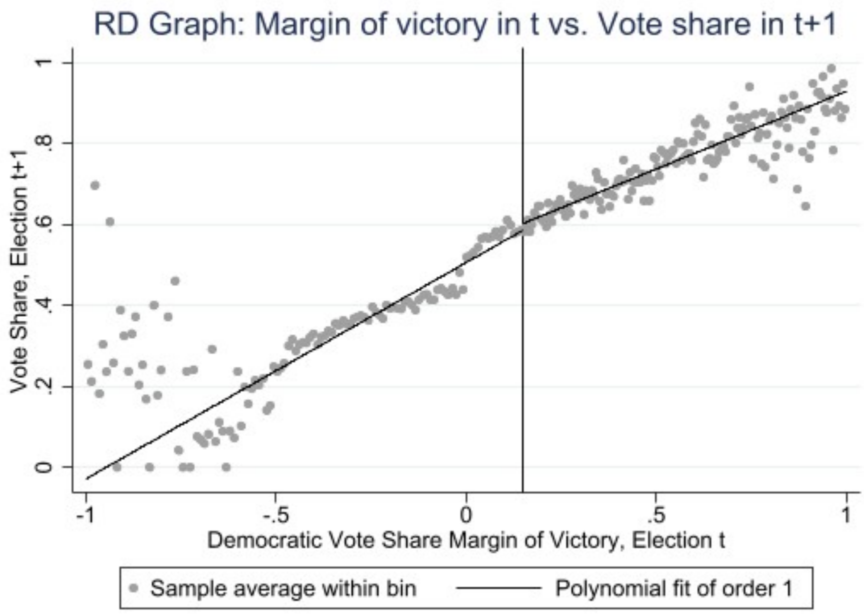

Figure 11: Non parametric estimation with placebo cutoff +0.15

5.6) Check sensitivity to bandwidth and polynomial order of the nonparametric estimation, and plot the different RD coefficients according to these different specifications.

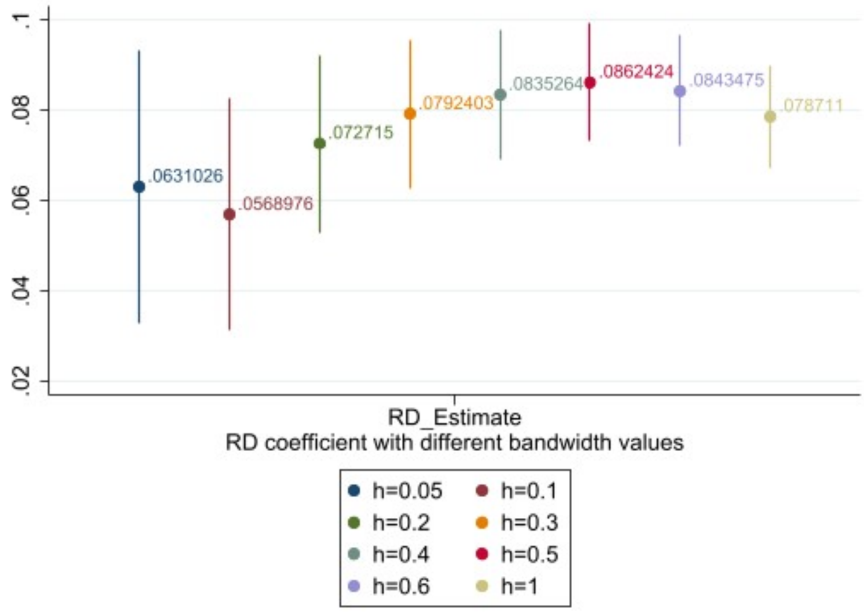

Figure 12: RD non parametric coefficients according to different bandwidths

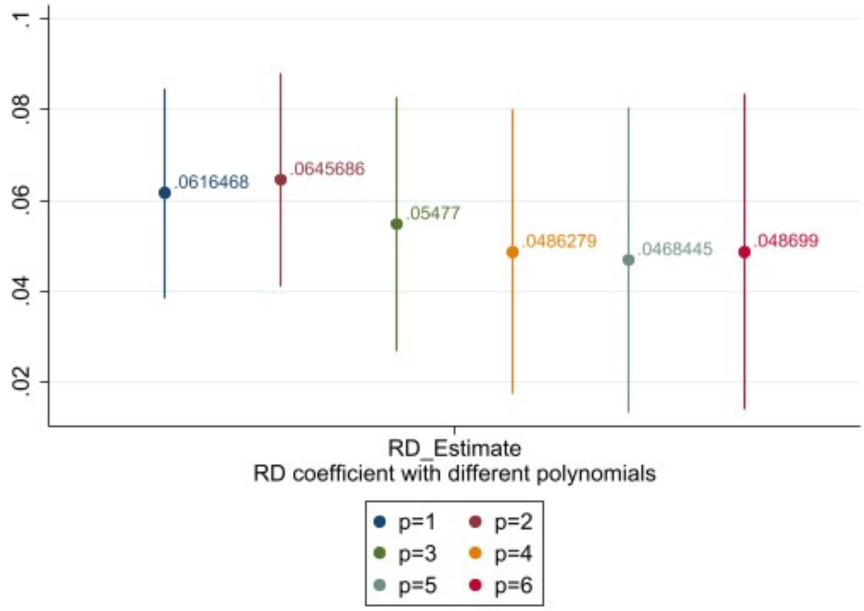

Figure 13: RD non parametric coefficients according to different bandwidths

References

- Lee, D.S. (2008). *Randomized experiments from nonrandom selection in U.S. House elections*. Journal of Econometrics, Vol. 142, pp. 675697

- Angrist, J. D. and J. Pischke (2009). *Mostly Harmless Econometrics*. Princeton University Press, New Jersey. Chapter 6.

- Lee, David S., and Thomas Lemieuxa, (2010). *Regression discontinuity designs in economics*. Journal of economic literature 48.2: 281-355.

A Replicating Lee (2008) RD Graphs using individual and group Democratic Party data

In this appendix, we are going to visualize the motivation of Lee's paper by replicating the four initial graphs in the paper. We will use Democratic party House Representatives data at an **individual candidate level**, and at a **group/party level**. Both data-sets vary in a yearly basis.

Replicating Figure 2 (a and b) of Lee (2008).

0.1) Using individual data of democratic house representatives (i.e. the *individfinal.dta* dataset), plot the estimated probability of a Democrat both running and winning election in $t+1$ (variable *mmyoutcomenext*) as a function of the Democratic vote share margin of victory in election t.
(*Hint*: use variable *difshare* which measures the Democratic vote share minus the vote share of the Democrats' strongest opponent, virtually always a Republican). Comment.

Each point is an average of the indicator variable for running and winning election $t+1$ for each interval. To the left of the dashed vertical line, the Democratic candidate lost election t; to the right, the Democrat won. As apparent from the figure, there is a striking discontinuous jump, right at the 0 point. Democrats who barely win an election are much more likely to run for office and succeed in the next election, compared to Democrats who barely lose. The causal effect seems enormous: about 0.45 in probability. Nowhere else is a jump apparent, as there is a well behaved, smooth relationship between the two variables, except at the threshold that determines victory or defeat.

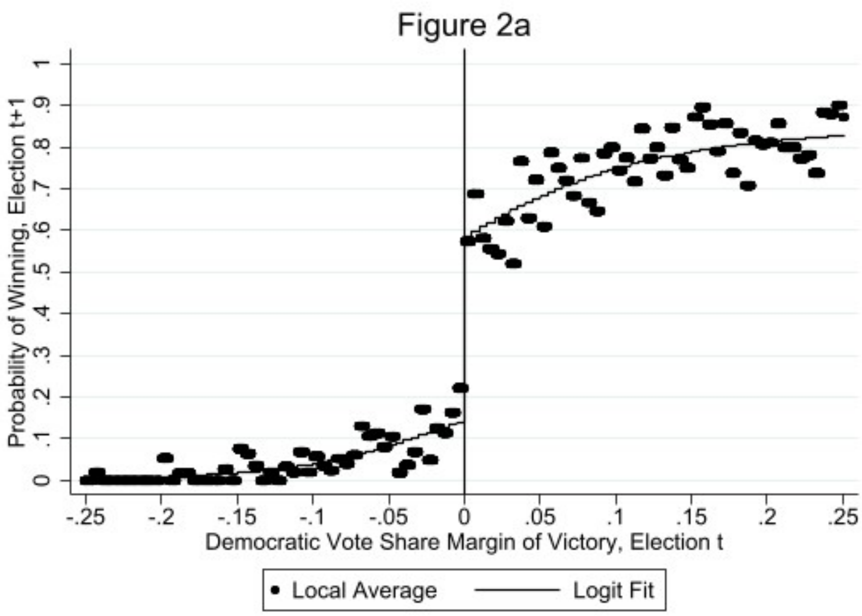

Figure 14: Lee (2008) - Figure 2.a

0.2) Using the same dataset, plot the Candidate's accumulated number of past election victories (i.e. variable *mofficeexp*) as a function of the Democratic vote share margin of victory in election t. What do you observe?

This figure corroborates the hypothesis that pre-determined characteristics (like candidate's political experience) are balanced in a neighborhood of the discontinuity threshold.

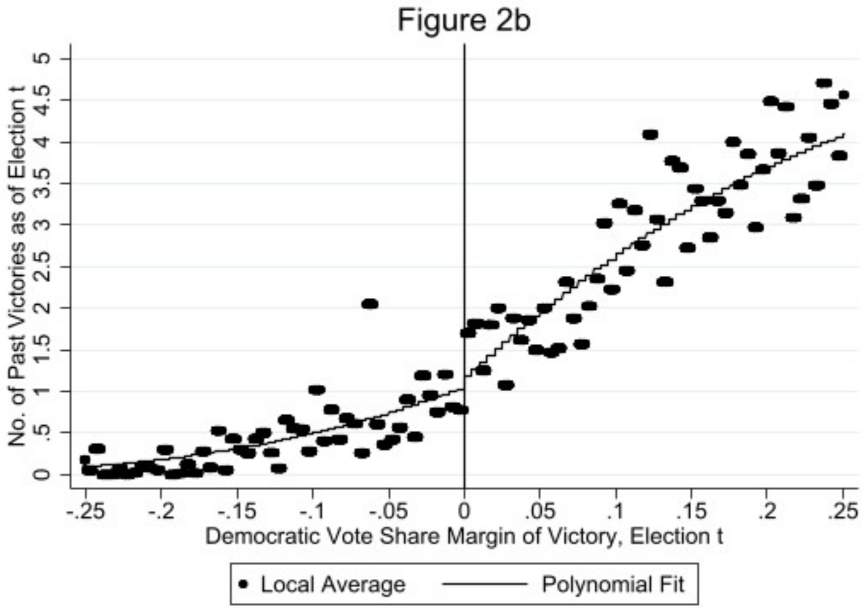

Figure 15: Lee (2008) - Figure 2.b

Replicating Figures 4 (a and b) of Lee (2008)

0.3) Using group data of democratic house representatives (i.e. the *groupfinal.dta* dataset), plot the Democratic party's vote share in election $t+1$ (i.e. *mdemsharenext*) as a function of the margin victory in election t (i.e. *difdemshare*). What do you observe?

In both figures (2a and 4a), there is a positive relationship between the margin of victory and the electoral outcome. For example, as in Fig. 4a, the Democratic vote shares in election t and $t+1$ are positively correlated, both on the left and right side of the figure. This indicates selection bias: a simple OLS comparison of means of Democratic winners and losers would yield biased measures of the incumbency advantage.

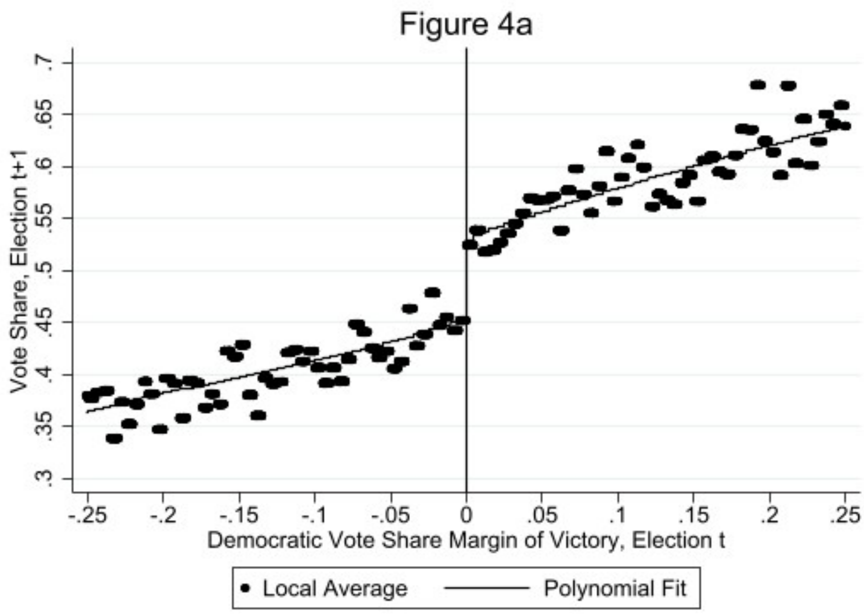

Figure 16: Lee (2008) - Figure 4.a

0.4) Using the same democratic group data, plot the Democratic party vote share in the previous election $t-1$ (i.e. *mdemshareprev*), by margin of victory in election t. Comment.

The main proposition (prop 3) of Lee's paper is that in the limit [of the threshold], there is randomized variation in treatment status. Figs. 2b and 4b corroborate this finding: all pre-determined characteristics (i.e. covariates) are balanced in a neighborhood of the discontinuity threshold (implication (c) of prop 3).

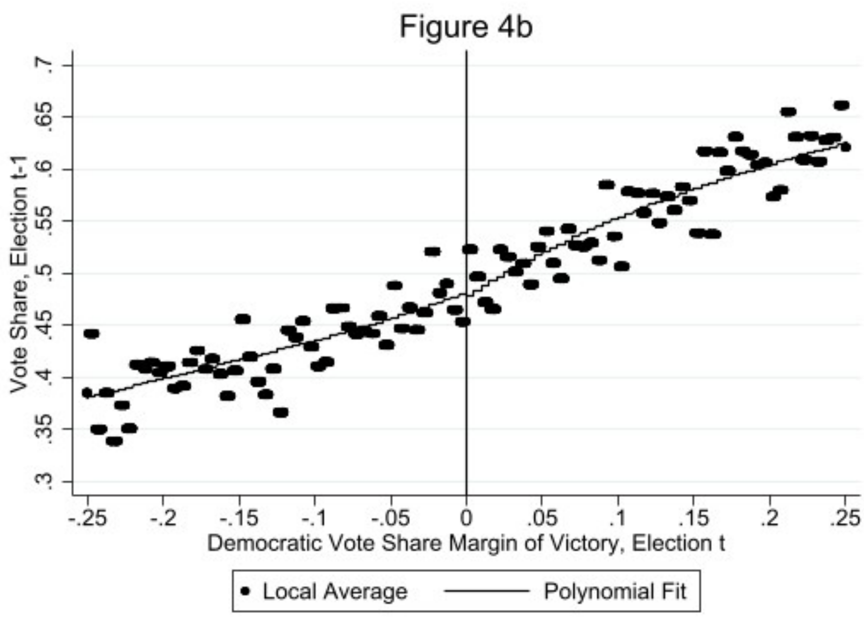

Figure 17: Lee (2008) - Figure 4.b