Regression Discontinuity Design

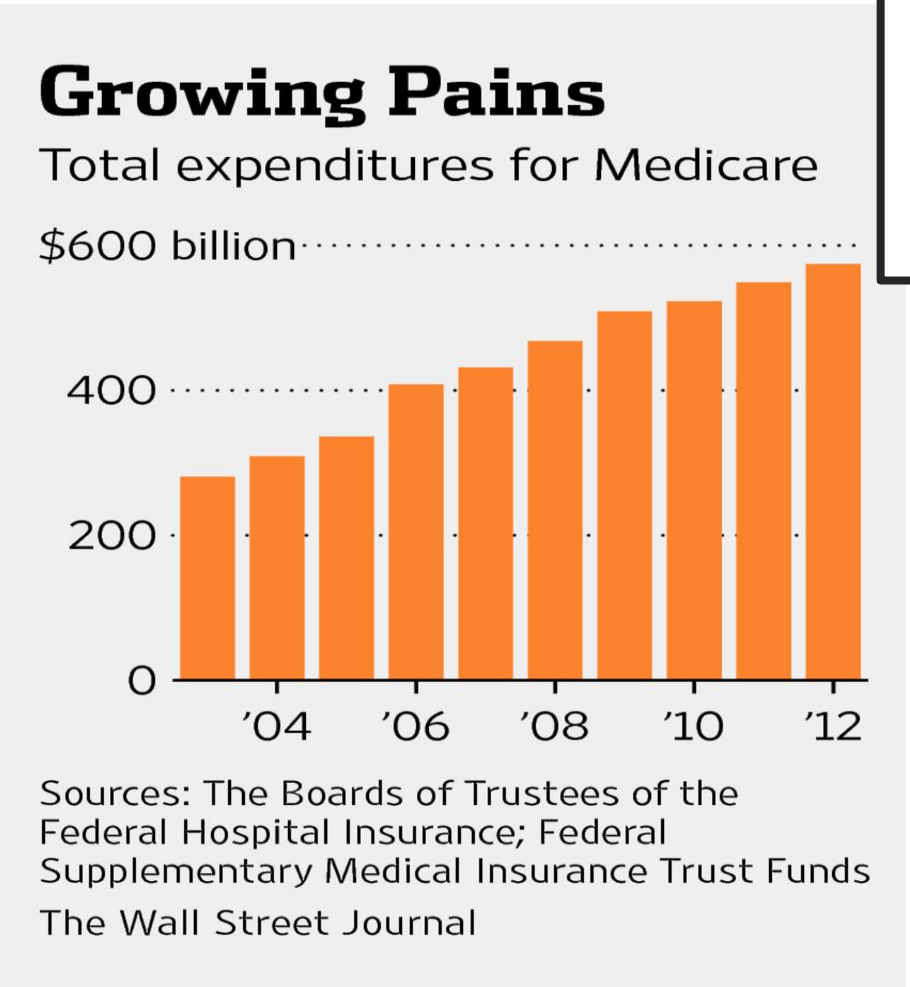
Econ 258 Data Analytics w/ R

Prepared by Ardina Hasanbasri

Outline

1) Conditional independence assumption and the Error Term
2) Regression Discontinuity Design
3) Coding Exercise Example with Politics!

Question Previews



1) What is the effect of a universal healthcare insurance program for the elderly? Card et al. (2008)

2) Can voters directly influence policies?
Or does the politicians you choose matter more?
Lee et al. (2004)

1.1 Conditional Independence and the Error Term

Let's Work More on Our Intuition of Conditional Independence

$$E[Y^1|D=1] - E[Y^1|D=0] = 0$$
$$E[Y^0|D=1] - E[Y^0|D=0] = 0$$

Recall, that we can rewrite the model (playful example). $\boldsymbol{D_i}$ is whether you use a cooking app.

$$cooking\ ability_i = \beta_0 + \alpha D_i + \beta_1 ingredients_i + \beta_2 travels_i + u_i$$

Conditional Independence Assumption:

$$E[\beta_0 + \beta_1 ingredients_i + \beta_2 travels_i + u_i | X, D = 1] -$$
$$E[\beta_0 + \beta_1 ingredients_i + \beta_2 travels_i + u_i | X, D = 0]$$

does this = 0?

$$E[u_i|X, D = 1] - E[u_i|X, D = 0] = 0$$

- Thus, another way to write conditional independence is:

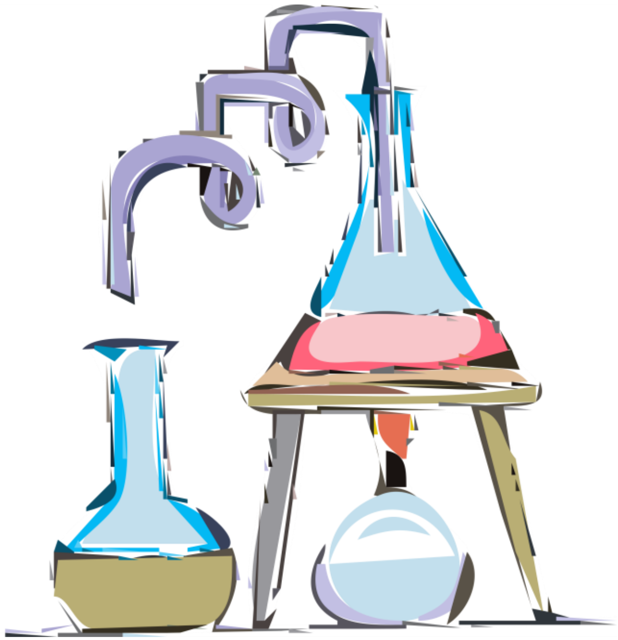
$$E[u_i|X, D = 1] - E[u_i|X, D = 0] = 0 \text{ or}$$

$$E[u_i|X, D = 1] = E[u_i|X, D = 0] = E[u_i] = 0$$

Intuition:

- Life is full of randomness. Each individuals have their own error term!
- On average, the randomness goes in both direction, and thus the average “shock” is zero.

1.2 Randomizing through Experiments



- By randomizing an experiment, we try to achieve this:

$$E[u_i|X, D = 1\,] = E[u_i|X, D = 0\,] = E[u_i] = 0$$

Intuition:

- Any other difference between the subjects of the control and treatment group that is not controlled for, such as ability and environmental background, on average will be 0.

- In fact, if you combined both group together, that uncontrolled ability or environmental background on average is 0.

- Remember how in OLS, $\boldsymbol{E[u_i] = 0}$ once controlled for everything.

1.3 Summary Takeaway

- If you can control for enough and make sure that the two groups satisfy

$$E[u_i|X, D = 1] = E[u_i|X, D = 0]$$

then you can get the average treatment effect (ATE)!

- For all the above equations, we are assuming no heterogenous effect of treatment.

α on the dummy for treatment or not has no subscript, does not depend on individual.

- But sometimes, we cannot get ATE using an experiment, if the experiment does not achieve randomization. (This is not an uncommon problem.)

2.1 Regression Discontinuity Design (RDD)

- One way to get two groups who are similar but are assigned randomly into to two different groups is to look at a **cutoff z.**

- Scholarship cutoffs example

A scholarship requires students to get a 3.8 GPA.

Students who won versus those who did not are different.

$$E[u_i|D_i = 0] \neq E[u_i|D_i = 1]$$

But how about students with GPA of 3.79 with GPA 3.8?

<u>2.2 Local Average Treatment Effect (LATE)</u>

- If people around the cutoff are similar to each other but are randomly assigned to a treatment and control group, we can get **<u>LATE</u>** by doing **<u>Regression Discontinuity Design.</u>**

- Conditional Independence Assumption needed to get LATE:

$$E[Y^0 \middle| D = 1, z^+] - E[Y^0 \middle| D = 0, z^-] = 0$$

Therefore, LATE = $E[Y^1 \middle| D = 1, z^+] - E[Y^0 \middle| D = 0, z^-]$

2.3 The Effect of Medicare (Card et. al 2008)

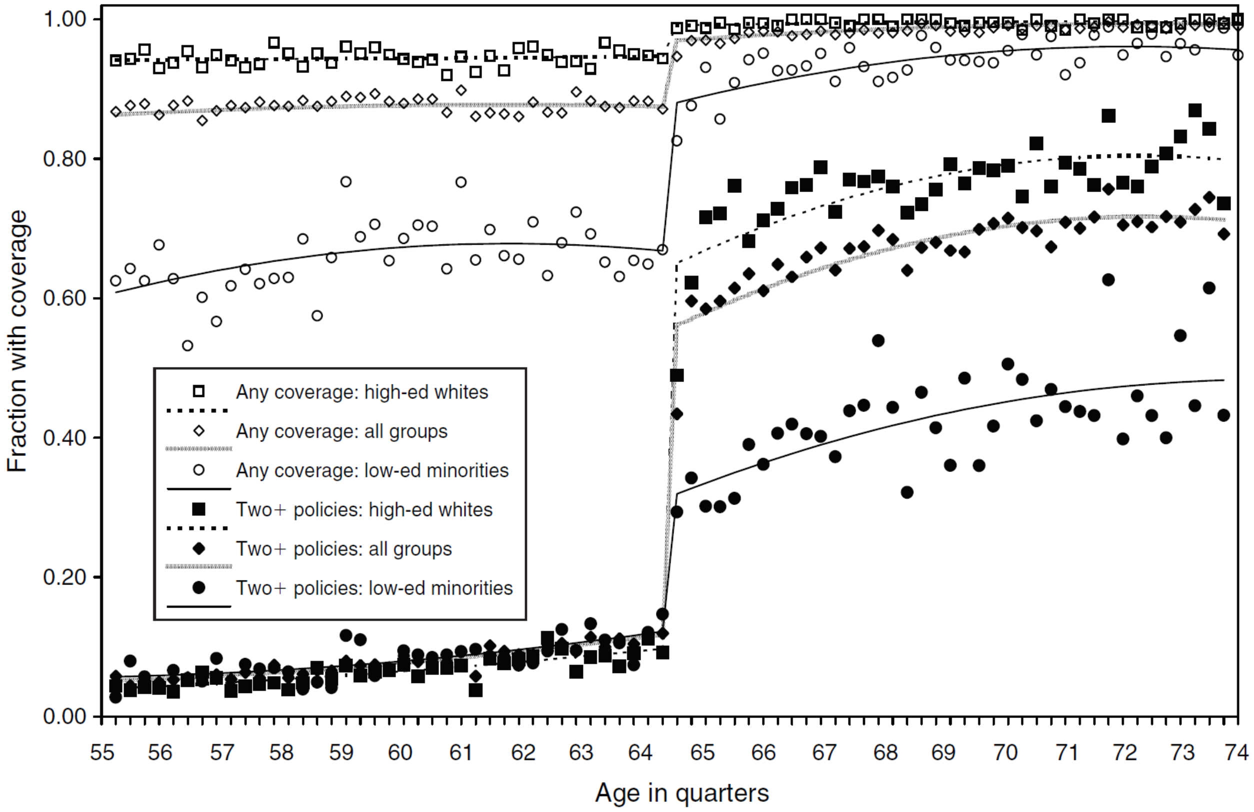


Figure 1. Coverage by Any Insurance and by Two or More Policies, by Age and Demographic Group

- What is the effect of universal healthcare insurance on the elderly? (Card et. al. 2008)
- Medicaid is available for people at the age of 65.
- People age 65 versus 64 is not necessarily that different.

Comparison of two different age groups.

TABLE 1—INSURANCE CHARACTERISTICS JUST BEFORE AGE 65 AND ESTIMATED DISCONTINUITIES AT AGE 65

	On Medicare		Any insurance		Private coverage		2+ Forms coverage		Managed care	
	Age 63–4 (1)	RD at 65 (2)	Age 63–4 (3)	RD at 65 (4)	Age 63–4 (5)	RD at 65 (6)	Age 63–4 (7)	RD at 65 (8)	Age 63–4 (9)	RD at 65 (10)
Overall sample	12.3	59.7 (4.1)	87.9	9.5 (0.6)	71.8	−2.9 (1.1)	10.8	44.1 (2.8)	59.4	−28.4 (2.1)
Classified by ethnicity and education:										
White non-Hispanic:										
High school dropout	21.1	58.5 (4.6)	84.1	13.0 (2.7)	63.5	−6.2 (3.3)	15.0	44.5 (4.0)	48.1	−25.0 (4.5)
High school graduate	11.4	64.7 (5.0)	92.0	7.6 (0.7)	80.5	−1.9 (1.6)	10.1	51.8 (3.8)	58.9	−30.3 (2.6)
At least some college	6.1	68.4 (4.7)	94.6	4.4 (0.5)	85.6	−2.3 (1.8)	8.8	55.1 (4.0)	69.1	−40.1 (2.6)
Minority:										
High school dropout	19.5	44.5 (3.1)	66.8	21.5 (2.1)	33.2	−1.2 (2.5)	11.4	19.4 (1.9)	39.1	−8.3 (3.1)
High school graduate	16.7	44.6 (4.7)	85.2	8.9 (2.8)	60.9	−5.8 (5.1)	13.6	23.4 (4.8)	54.2	−15.4 (3.5)
At least some college	10.3	52.1 (4.9)	89.1	5.8 (2.0)	73.3	−5.4 (4.3)	11.1	38.4 (3.8)	66.2	−22.3 (7.2)
Classified by ethnicity only:										
White non-Hispanic (all)	10.8	65.2 (4.6)	91.8	7.3 (0.5)	79.7	−2.8 (1.4)	10.4	51.9 (3.5)	61.9	−33.6 (2.3)
Black non-Hispanic (all)	17.9	48.5 (3.6)	84.6	11.9 (2.0)	57.1	−4.2 (2.8)	13.4	27.8 (3.7)	48.2	−13.5 (3.7)
Hispanic (all)	16.0	44.4 (3.7)	70.0	17.3 (3.0)	42.5	−2.0 (1.7)	10.8	21.7 (2.1)	52.9	−12.1 (3.7)

Note: Entries in odd-numbered columns are percentages of age 63- 64-year-olds in group with insurance characteristic shown in column heading. Entries in even-numbered columns are estimated regression discontinuities at age 65, from models that include quadratic control for age, fully interacted with dummy for age 65 or older. Other controls include indicators for gender, race/ethnicity, education, region, and sample year. Estimates are based on linear probability models fit to pooled samples of 1999–2003 NHIS.

TABLE V

REGRESSION DISCONTINUITY ESTIMATES OF CHANGES IN MORTALITY RATES

	Death rate in					
	7 days	14 days	28 days	90 days	180 days	365 days
	Estimated discontinuity at age 65 (×100)					
Fully interacted quadratic with no additional controls	−1.1	−1.0	−1.1	−1.1	−1.2	−1.0
	(0.2)	(0.2)	(0.3)	(0.3)	(0.4)	(0.4)
Fully interacted quadratic plus additional controls	−1.0	−0.8	−0.9	−0.9	−0.8	−0.7
	(0.2)	(0.2)	(0.3)	(0.3)	(0.3)	(0.4)
Fully interacted cubic plus additional controls	−0.7	−0.7	−0.6	−0.9	−0.9	−0.4
	(0.3)	(0.2)	(0.4)	(0.4)	(0.5)	(0.5)
Local linear regression procedure fit separately to left and right with rule-of-thumb bandwidths	−0.8	−0.8	−0.8	−0.9	−1.1	−0.8
	(0.2)	(0.2)	(0.2)	(0.2)	(0.3)	(0.3)
Mean of dependent variable (%)	5.1	7.1	9.8	14.7	18.4	23.0

Notes. Standard errors in parentheses. Dependent variable is indicator for death within interval indicated by column heading. Entries in rows (1)–(3) are estimated coefficients of dummy for age over 65 from models that include a quadratic polynomial in age (rows (1) and (2)) or a cubic polynomial in age (row (3)) fully interacted with a dummy for age over 65. Models in rows (2) and (3) include the following additional controls: a dummy for people who are within 1 month of their 65 birthdays, dummies for year, month, sex, race/ethnicity, and Saturday or Sunday admissions, and unrestricted fixed effects for each ICD-9 admission diagnosis. Entries in row (4) are estimated discontinuities from a local linear regression procedure, fit separately to the left and right, with independently selected bandwidths from a rule-of-thumb procedure suggested by Fan and Gijbels (1996). Sample includes 407,386 observations on patients between the ages of 60 and 70 admitted to California hospitals between January 1, 1992, and November 30, 2002, for unplanned admission through the ED who have nonmissing Social Security numbers. All coefficients and their SEs have been multiplied by 100.

3.1 Does your vote matter? Elect or Affect by Lee at al. (2004)

Question motivation:

- The median voter theorem is a famous theorem that says if people's preference is single peak and one dimensional, then due to political competition, political parties will try to get the most votes by running on a platform that resembles the median vote.

- Is this theory true? If we see a region where two parties are getting approximately half the vote, then the theorem is true if the two parties have similar platforms. It will not matter whether a democrat or a republican win. Thus, voter AFFECT policies.

- If after winning, the parties just vote on policies they like, then voters ELECT politicians, and this is evidence against the median voter theorem.

- How can we answer this with a regression discontinuity design?

This example is in page 186 Cunningham's textbook.

3.2 Outcome of Interest and Model

Note: This is a very simplified version of the model. Focus on the intuition.

The outcome of interest here is roll call votes on a policy (RC). The higher the number, the more liberal the policy position.

$$RC_t = D_t x_t + (1 - D_t) y_t$$

D is a democrat winning. P* is the popularity of democratic party. In regression form:

$$\begin{aligned} RC_t &= \alpha_0 + \pi_0 P_t^* + \pi_1 D_t + \varepsilon_t \\ RC_{t+1} &= \beta_0 + \pi_0 P_{t+1}^* + \pi_1 D_{t+1} + \varepsilon_{t+1} \end{aligned}$$

Note: This is actually SDO. The SDO will give us the treatment effect as long independence assumption holds in our sample. We will be estimating LATE here.

$$\begin{aligned}
\underbrace{E[RC_{t+1}|D_t = 1] - E[RC_{t+1}|D_t = 0]}_{\text{Observable}} &= \pi_0[P^{*D}_{t+1} - P^{*R}_{t+1}] && \text{Affect} \\
&+ \underbrace{\pi_1[P^{D}_{t+1} - P^{R}_{t+1}]}_{\text{Observable}} && \text{Elect} \\
&= \underbrace{\gamma}_{\text{Total effect of initial win on future roll call votes}} && (80) \\
\underbrace{E[RC_{t}|D_t = 1] - E[RC_{t}|D_t = 0]}_{\text{Observable}} &= \pi_1 && (81) \\
\underbrace{E[D_{t+1}|D_t = 1] - E[D_{t+1}|D_t = 0]}_{\text{Observable}} &= P^{D}_{t+1} - P^{R}_{t+1} && (82)
\end{aligned}$$

3.3 Pure Intuition Based on Regression

$$E[RC_{t+1}|D_t = 1] - E[RC_{t+1}|D_t = 0]$$

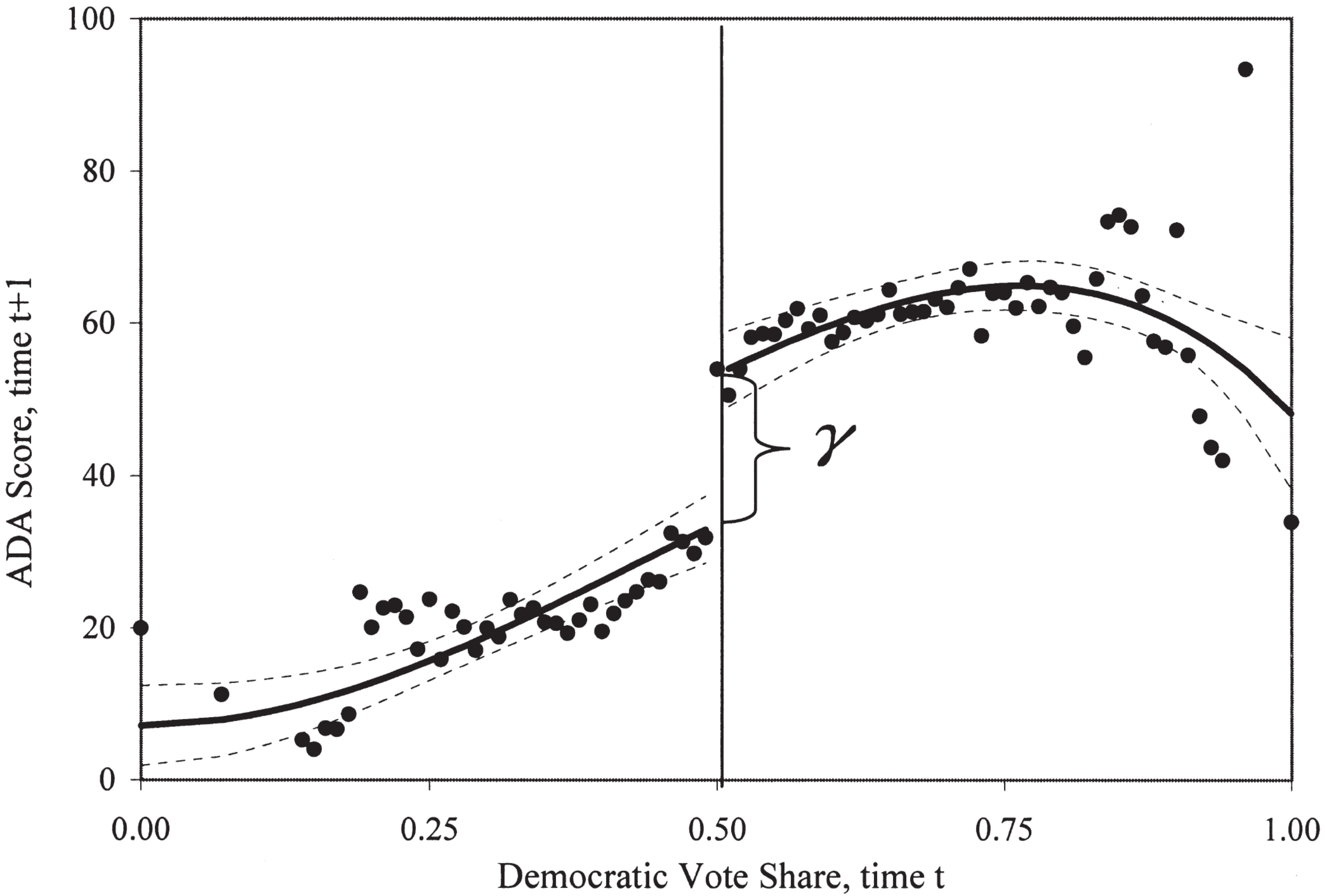
is difference between how liberals the policies are at t+1 given who was elected the previous period. Regress roll call score with who won last period.

$$E[RC_t|D_t = 1] - E[RC_t|D_t = 0]$$

is the difference between how liberals the policies t+1 are given who was elected the current period. This is estimated using t+1 instead of t in regression.

$$E[D_{t+1}|D_t = 1] - E[D_{t+1}|D_t = 0]$$

is the difference in probability of having a democrat winning depending on who wins the previous period.



100
80
60
40
20
0
ADA Score, time t+1
γ
0.00
0.25
0.50
0.75
1.00
Democratic Vote Share, time t

3.4 Replicate Results from Paper

TABLE I

RESULTS BASED ON ADA SCORES—CLOSE ELECTIONS SAMPLE

	Total effect			Elect component	Affect component
	γ	π_1	$(P^D_{t+1} - P^R_{t+1})$	$\pi_1[(P^D_{t+1} - P^R_{t+1})]$	$\pi_0[P^{*D}_{t+1} - P^{*R}_{t+1}]$
Variable	ADA_{t+1}	ADA_t	DEM_{t+1}	(col. (2)*(col. (3))	(col. (1)) − (col. (4))
	(1)	(2)	(3)	(4)	(5)
Estimated gap	21.2	47.6	0.48		
	(1.9)	(1.3)	(0.02)		
				22.84	−1.64
				(2.2)	(2.0)

Standard errors are in parentheses. The unit of observation is a district-congressional session. The sample includes only observations where the Democrat vote share at time t is strictly between 48 percent and 52 percent. The estimated gap is the difference in the average of the relevant variable for observations for which the Democrat vote share at time t is strictly between 50 percent and 52 percent and observations for which the Democrat vote share at time t is strictly between 48 percent and 50 percent. Time t and $t + 1$ refer to congressional sessions. ADA_t is the adjusted ADA voting score. Higher ADA scores correspond to more liberal roll-call voting records. Sample size is 915.

3.5 Coding Exercise

Download data "voting_data_for_rdd_example.dta"

1) Regress score on lag democrat for when lagdemvoteshare is between 0.48 to 0.52 (should be able to do this in one line after importing data).

2) Regress score on democrat for when lagdemvoteshare is between 0.48 to 0.52

3) Regress democrat on lagdemocrat, again when lagdemvoteshare is between 0.48 to 0.52.

4) How would the results change when using the entire sample instead of with restrictions?

5) So do voters elect or affect policy?

Extra: See if you can replicate figure on slide 17 using the command Rdestimate. (Will need to install package "rdd").

Reference

"Causal Inference: The Mixtape (V.1.7)" by Scott Cunningham.