

Causal inference week 5: Regression Discontinuity Design

Andy Eggers

Oxford DPIR

HT 2018

Introduction

Motivation: a non-RDD study of turnout and electoral systems

Regression discontinuity design: uses and assumptions

- Examples in political science

- The (quasi-)randomization approach to RDD

- The continuity approach to RDD

- RDD Implementation

Evaluating RDD assumptions

Example: effect of naturalization on political integration in Switzerland

Wrap-up

Appendix: diagrams

Overview

Strategies for estimating effects of treatments so far:

- ▶ Randomize treatment and take the DIGM
- ▶ Identify and control for confounding variables such that the CIA holds
- ▶ Identify an instrumental variable and use two-stage-least-squares to estimate average treatment effect for compliers

Today: make use of special situations where treatment is applied based on a cutoff

Introduction

Motivation: a non-RDD study of turnout and electoral systems

Regression discontinuity design: uses and assumptions

- Examples in political science

- The (quasi-)randomization approach to RDD

- The continuity approach to RDD

- RDD Implementation

Evaluating RDD assumptions

Example: effect of naturalization on political integration in Switzerland

Wrap-up

Appendix: diagrams

Electoral systems in Swiss municipalities

Ladner & Milner (1999) study effect of electoral system on turnout in Swiss municipalities.

Some Swiss municipalities use PR, others use a majoritarian system:

Table 3

Voting system for the communal executive

Cantons where the majority system is used in all communities:

Zurich, Lucerne, Uri, Schwyz, Obwalden, Nidwalden, Glarus, Basel-Stadt, Schaffhausen, Appenzell-Ausserrhoden, Appenzell-Innerrhoden, St. Gallen, Aargau, Vaud, Neuchâtel, Geneva

Cantons where the proportional system is used in all communities:

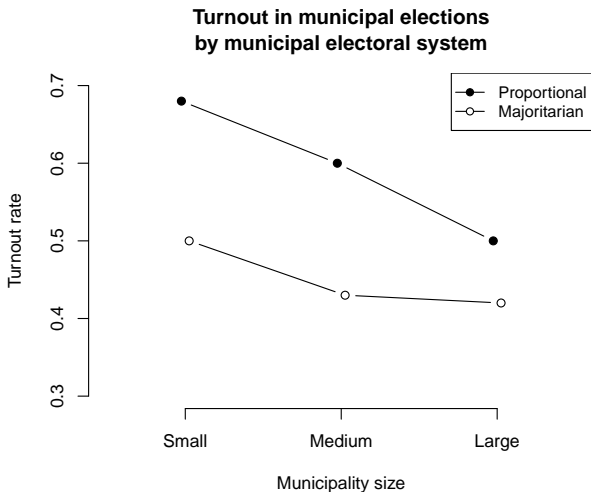
Zug, Tessin

Cantons with both systems (in per cent):

	Majority system	PR system	Responding communities (<i>N</i>)
Wallis	21	79	123
Solothurn	26	74	104
Jura	42	58	71
Fribourg	47	53	191
Berne	60	40	341
Basel-Land	83	17	65
Thurgau	95	5	106
Graubünde	99	1	151

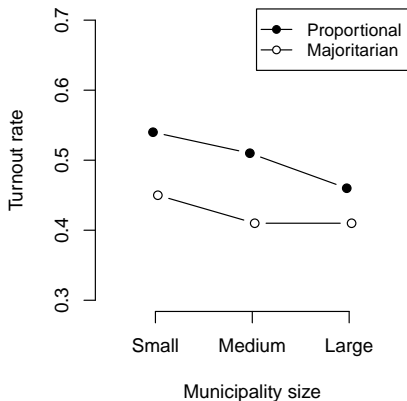
Source: Communal Secretary Survey (1988), see Ladner (1991a).

Municipal electoral systems and municipal turnout

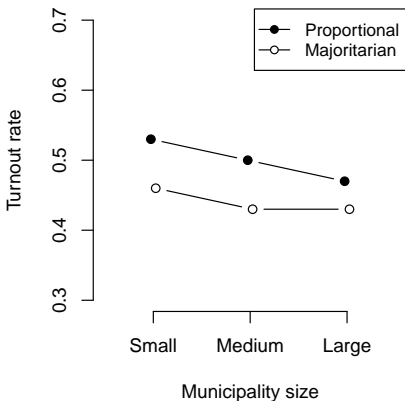


Municipal electoral systems and turnout in other elections

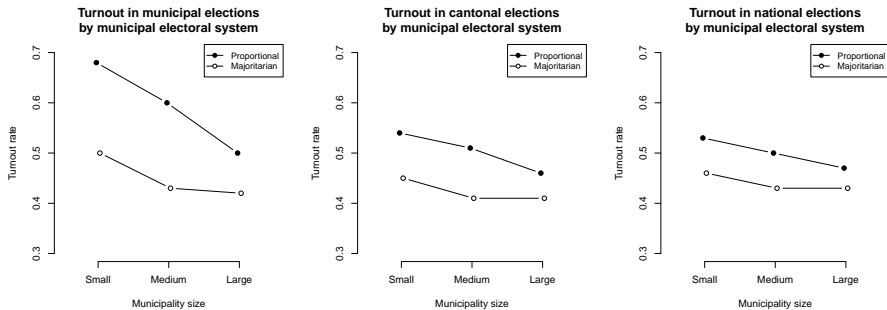
Turnout in cantonal elections by municipal electoral system



Turnout in national elections by municipal electoral system



Interpreting Ladner & Milner (1999)



Two interpretations:

1. PR is adopted in places that are “culturally predisposed toward higher political participation” (**confounding, omitted variable bias**)
2. voting in municipal elections affects voting in higher-level elections (**spillovers**)

Going beyond Ladner & Milner (1999)

How can we disentangle the **determinants** of the electoral system from its **effects**?

- ▶ **additional covariates** to capture predisposition toward higher political participation (weeks 2&3)
- ▶ is there an **instrumental variable**? (e.g. the party in local control at a certain point in history) (week 4)
- ▶ is there another context where the electoral system depends on a cutoff (**RDD**)? (week 5)
- ▶ can we exploit changes in the electoral system with a diff-in-diff/panel strategy? (weeks 6&7)

Introduction

Motivation: a non-RDD study of turnout and electoral systems

Regression discontinuity design: uses and assumptions

- Examples in political science

- The (quasi-)randomization approach to RDD

- The continuity approach to RDD

- RDD Implementation

Evaluating RDD assumptions

Example: effect of naturalization on political integration in Switzerland

Wrap-up

Appendix: diagrams

French municipal election systems

In France, small municipalities use a majoritarian system while larger ones use PR.

- ▶ “Small”: population 1-3,499 (until recently)
- ▶ “Large”: population 3,500+ (until recently)

How can we use this to measure the effect of the electoral system on turnout?

Why is this better than a situation where the municipalities choose their electoral system?

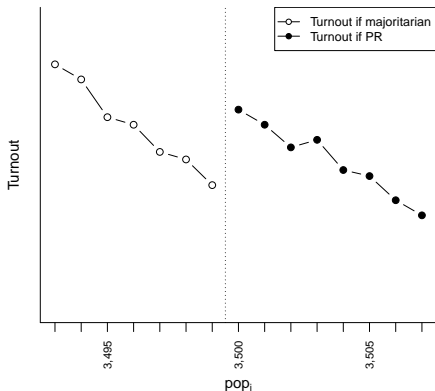
The appeal of RDD: we almost observe both potential outcomes

Recall: The ideal is the ATE:
 $E[Y_i(1) - Y_i(0)]$.

Just above the threshold c , we observe $E[Y_i(1)|X_i = c + \epsilon]$.

Just below the threshold c , we observe $E[Y_i(0)|X_i = c - \epsilon]$.

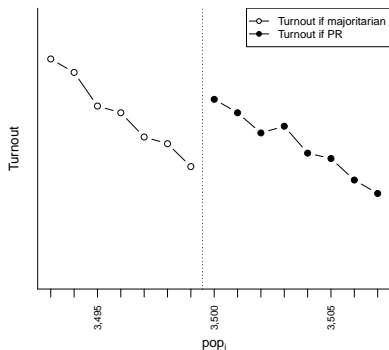
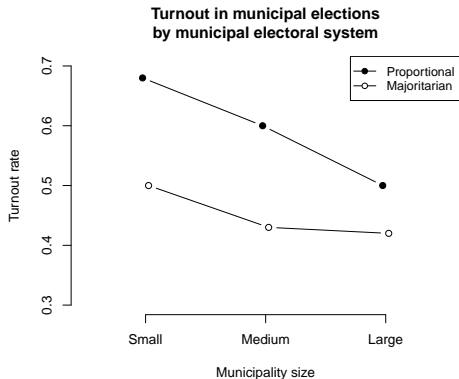
So with a bit of **extrapolation**, we can estimate $E[Y_i(1) - Y_i(0)|X_i = c]$.



The appeal of RDD: credible CIA

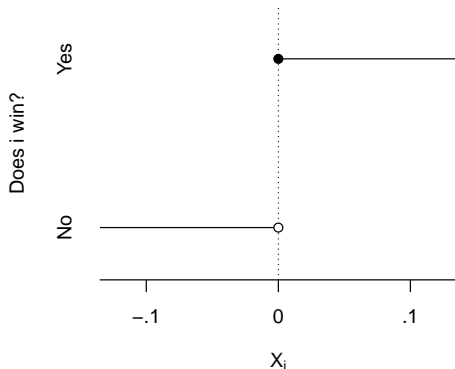
In the usual “selection on observables” study, we hope CIA holds, i.e. treatment independent of potential outcomes conditional on a bunch of covariates.

In an RDD study, CIA pretty credible conditional on $X_i = c$!



What treatments depend on a threshold?

Election outcomes in plurality systems: define running variable X_i as difference between i 's vote share and the best vote share among all other candidates. Then i wins if and only if $X_i > 0$.

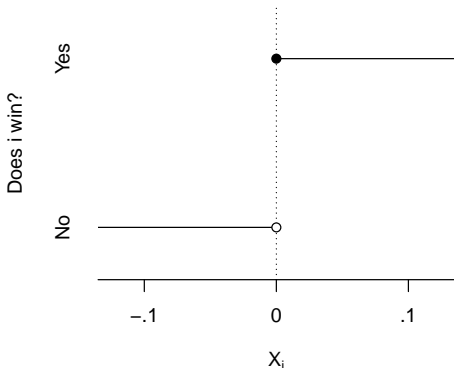


Used to study the effect of: party control on policy outcomes; incumbency status on election outcomes; election outcomes on candidate positioning; winning office on wealth, health, divorce; representation by women, minorities, more educated politicians on policy; etc

What treatments depend on a threshold? (2)

Election outcomes in other electoral systems:

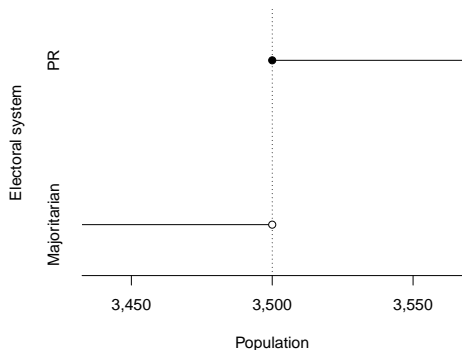
- ▶ minimum vote threshold in PR systems
- ▶ allocation of “last seat” in PR systems
- ▶ thresholds determining who advances in two-round systems



Used to study effect of small party representation on policy outcomes; effect of incumbency status on wealth, future election outcomes; strategic voting

What treatments depend on a threshold? (3)

Policies in municipal governments in many countries depend on population thresholds.



This has been used to study effects of: mayor salary, gender quotas, electoral rules, direct democracy, fiscal transfers, council size

What treatments depend on a threshold? (4)

Various others:

- ▶ Journalists report a recession following two quarters of “negative growth”
- ▶ Eligibility to vote depends on an age threshold
- ▶ Qualifying for university, gymnasium, etc depends on achieving a passing mark on an exam
- ▶ Many social programs depend on cutoffs, e.g. household income below a certain amount, age above a certain amount

RDD interpretation 1: quasi-random experiment

The official population of a city depends on

- ▶ its “true” population, which might be related to potential outcomes (i.e. turnout in each system)
- ▶ random factors that are not related to potential outcomes

If

- ▶ the error is large enough, and
- ▶ we focus on cities close enough to the threshold

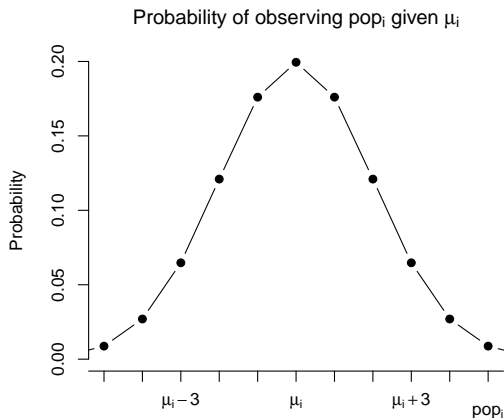
then treatment status (PR or not) is close to random (as-if random).

RDD interpretation 1: quasi-random experiment (2)

Suppose official population pop_i is

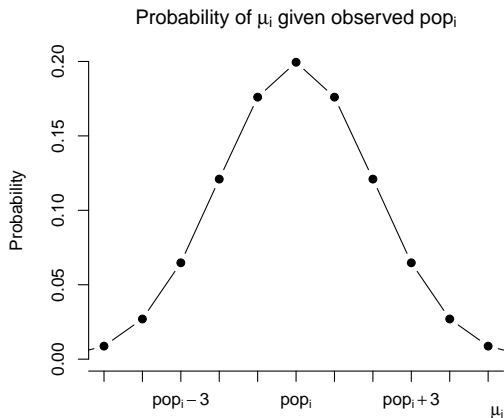
$$pop_i = \mu_i + \epsilon_i$$

where μ_i is the true population and ϵ_i is random error, with mean 0.



RDD interpretation 1: quasi-random experiment (3)

The distribution of true population μ_i given observed population pop_i mirrors the distribution of random error ϵ_i (according to Bayes' Law, assuming uniform μ_i).



RDD interpretation 1: quasi-random experiment (5)

So given a city's observed population, we can calculate the (ex ante) probability it would be treated; the larger the random error, the more similar this is for cities just above and below the threshold.

Ex ante probability of PR elections as function of official population pop_i

	Official pop. (pop_i)			
	3,498	3,499	3,500	3,501
sd=2	0.23	0.4	0.6	0.77
sd=5	0.38	0.46	0.54	0.62
sd=10	0.44	0.48	0.52	0.56
sd=20	0.47	0.49	0.51	0.53

More randomness in **running variable** → closer to randomized experiment.

DIGM to analyze RDD?

If the running variable were **completely** random near the threshold, DIGM would be an unbiased estimator of ATE near the threshold.

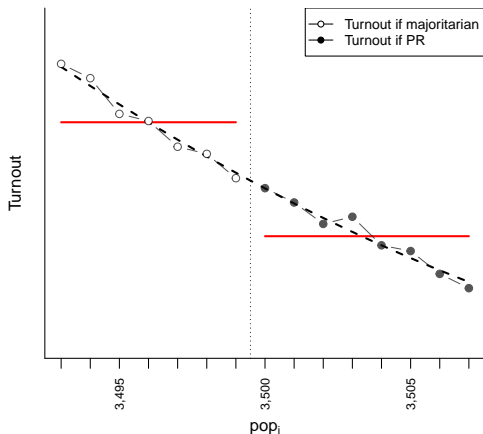
For example, you might run this regression

$$\text{Turnout}_i = \beta_0 + \beta_1 \text{PopOver3500}_i$$

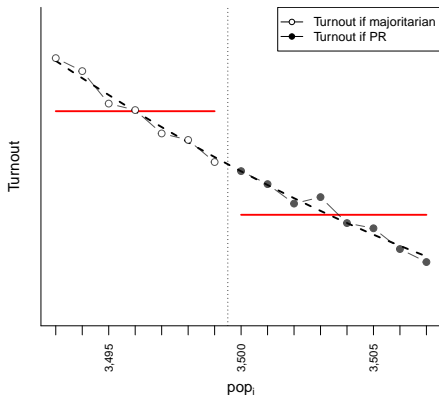
for cities with $\text{Pop}_i \in [3, 450, 3, 550]$.

DIGM to analyze RDD? (2)

Usually the running variable is not completely random, and the running variable is related to the outcome, so DIGM will be biased.



DIGM to analyze RDD? (3)



Solution: control for running variable, e.g. run regression

$$\text{Turnout}_i = \beta_0 + \beta_1 \text{PopOver3500}_i + \beta_2 \text{Pop}_i + \beta_3 \text{Pop}_i^2$$

for cities with $\text{Pop}_i \in [3,450, 3,550]$.

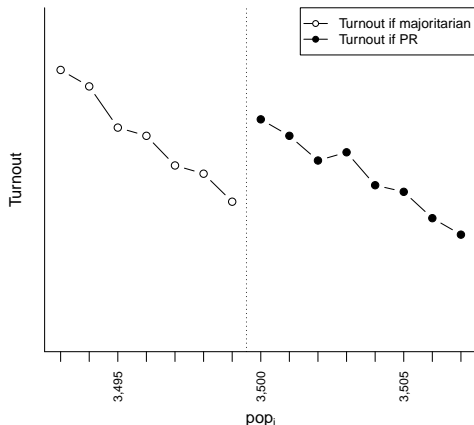
RDD interpretation 2: continuity

Suppose there is actually no random error in population measurements \rightarrow no “quasi-randomization”. Does RDD still work?

We want to estimate

$$E[Y_i(1) - Y_i(0) \mid \text{Pop}_i = 3,500].$$

Under what assumptions/conditions can we do this?

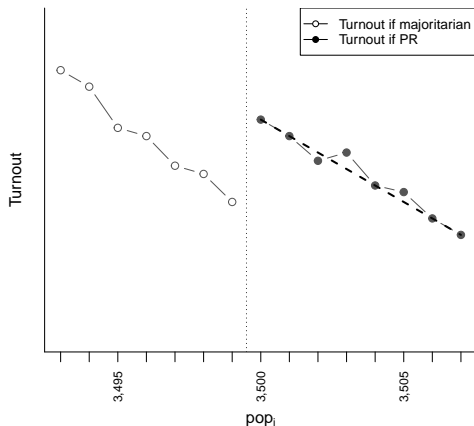


Component 1: unbiased estimate of the CEF on the right

To estimate $CATE_{3,500} = E[Y_i(1) - Y_i(0) \mid Pop_i = 3,500]$, we need:

(1) A good (unbiased, precise) estimate of $E[Y_i(1) \mid Pop_i = 3,500]$, using observations with $Pop_i \geq 3,500$.

Requirement: lots of data and/or “well-behaved” conditional expectation function (CEF)



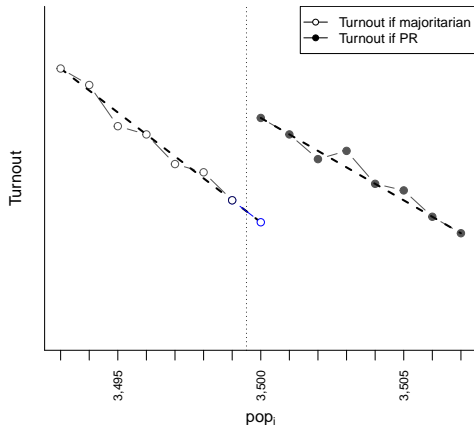
Component 2: extrapolation of CEF across the threshold

To estimate $\text{CATE}_{3,500} = E[Y_i(1) - Y_i(0) | \text{Pop}_i = 3,500]$, we need:

(2) A good estimate of $E[Y_i(0) | \text{Pop}_i = 3,500]$, using observations with $\text{Pop}_i < 3,500$.

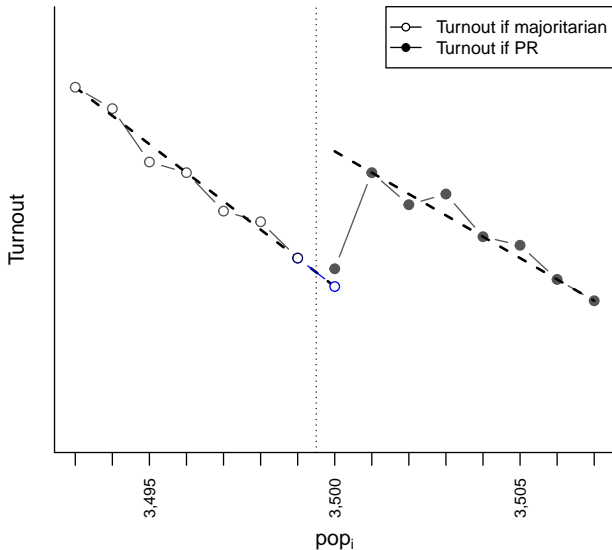
Requirement:

“well-behaved” conditional expectation function (CEF), allowing for some **extrapolation**

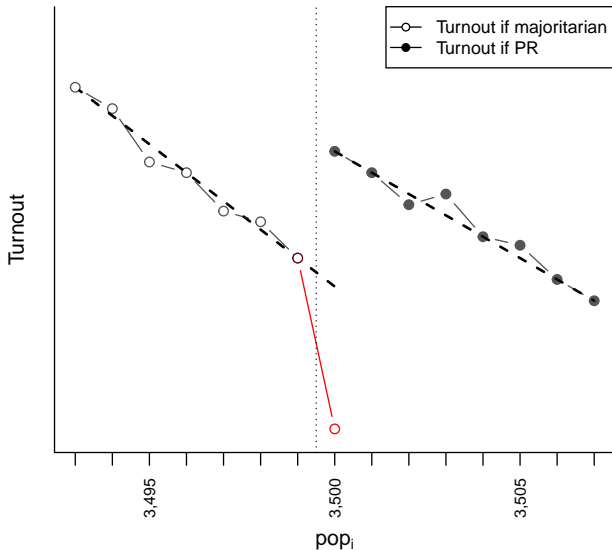


More generally, need minimal extrapolation from both sides to estimate $\text{CATE}_{X_i=c}$.

RDD gone wrong: poorly-behaved CEF on right

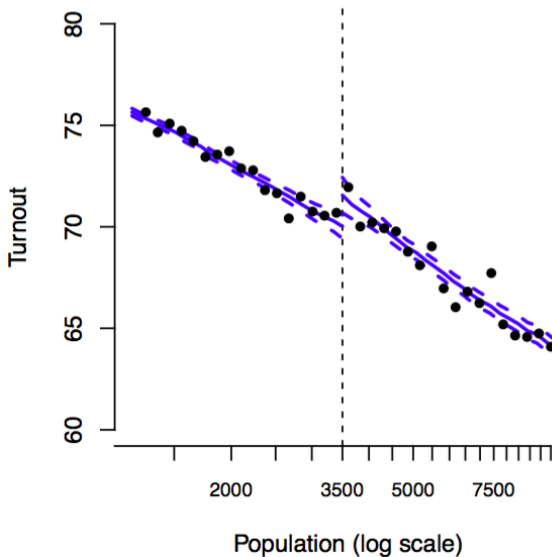


RDD gone wrong: poorly-behaved CEF across threshold



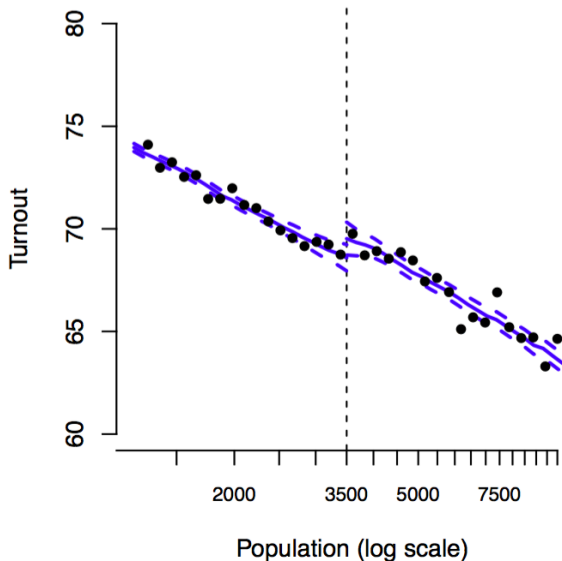
Results of RDD analysis in French municipalities (2001)

Turnout in 2001 elections as function of population



Results of RDD analysis in French municipalities (2008)

Turnout in 2008 elections as function of population



Results of RDD analysis in French municipalities

Effect of crossing 3,500 threshold on turnout in various elections

Outcome	Mean turnout	Effect		
		(1)	(2)	(3)
Municipal, 2001	70.73	0.989 (0.778)	1.537** (0.538)	1.525*** (0.433)
Municipal, 2008	69.14	0.763 (0.765)	0.929† (0.523)	1.476*** (0.423)
Municipal, 2001 & 2008	69.96	0.878 (0.71)	1.242** (0.481)	1.502*** (0.385)
Presidential, 2002	74.95	-0.04 (0.413)	-0.189 (0.29)	-0.038 (0.241)
Regional, 2004	63.38	-0.448 (0.583)	-0.7† (0.414)	-0.241 (0.341)
Presidential, 2007	86.33	-0.248 (0.326)	-0.439† (0.224)	-0.253 (0.185)
Window:	25%	25%	50%	75%

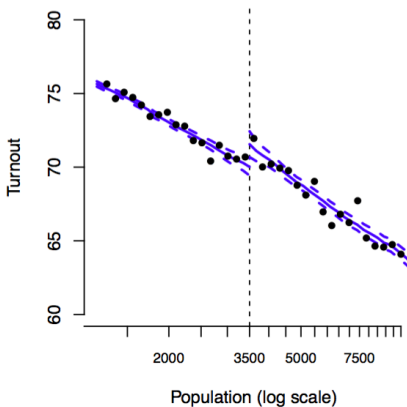
Factors to consider in implementing RDDs

- ▶ Do you use all the data, or just data close to the threshold?
- ▶ How do you decide what “close” means? (**bandwidth selection**)
- ▶ Do you weight the data close to the threshold more strongly? (**kernel selection**)
- ▶ Do you run a linear regression, or do you include polynomials of the running variable?
- ▶ Do you allow the functional form to be different above and below the threshold?

How to do RDDs

- ▶ Show robustness to alternatives, as **MM** recommend, but don't cherry-pick
- ▶ Bias-corrected approach in `rdrobust` package (Calonico, Cattaneo, Titiunik) is currently the standard, but that may change
- ▶ Show the RD plot (also available in `rdrobust`)

Turnout in 2001 elections as function of population



Introduction

Motivation: a non-RDD study of turnout and electoral systems

Regression discontinuity design: uses and assumptions

- Examples in political science

- The (quasi-)randomization approach to RDD

- The continuity approach to RDD

- RDD Implementation

Evaluating RDD assumptions

Example: effect of naturalization on political integration in Switzerland

Wrap-up

Appendix: diagrams

RDD assumptions: recap

RDD works if you can estimate the conditional expectation function (CEF) $E[Y_i | X_i]$ **near** the threshold and extrapolate a little **to** the threshold.

When might this not be true?

The key RDD pitfall: precise sorting

Suppose politically entrenched mayors can manipulate the census results and prefer to have majoritarian elections.

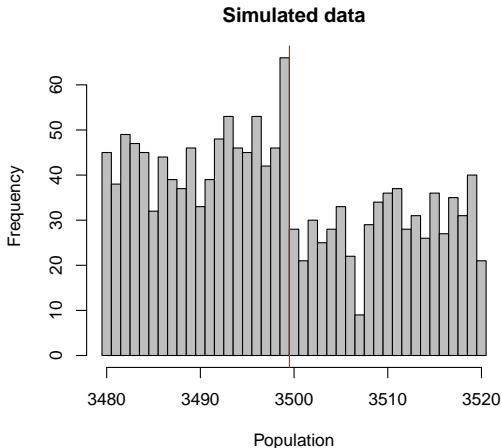
After seeing an initial census count, these mayors can add or subtract up to 10 inhabitants.

Question: How would this affect your RDD?

Effect of precise sorting (1)

You might expect:

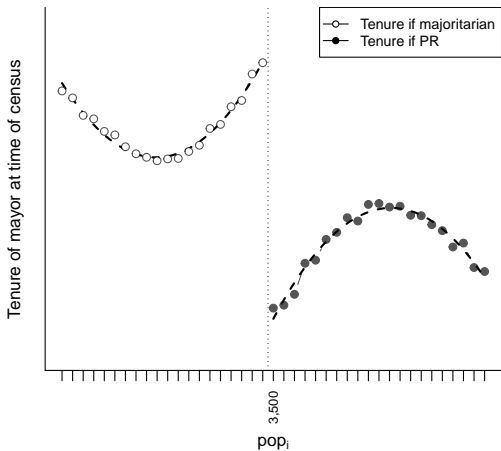
Too many cities with populations between 3,490 and 3,499, too few between 3,500 and 3,509



Effect of precise sorting (2)

You might expect:

An “effect” of crossing the threshold on covariates related to political entrenchment, e.g. mayor’s tenure at time of census

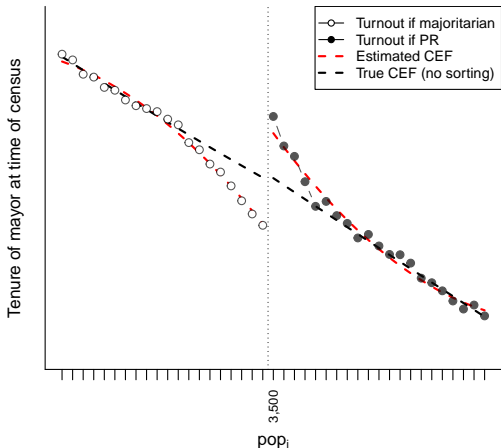


Effect of precise sorting (3)

You might expect:

A biased estimate in the RDD
(though you won't know it).

e.g. if turnout is low in places with politically entrenched mayors, you may find a jump crossing the threshold even if PR actually has no effect on turnout.



Testing for sorting

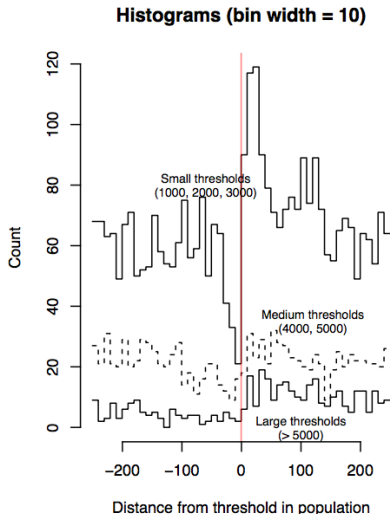
- ▶ McCrary (2008) test for discontinuity in density of running variable (see also newer versions)
- ▶ RDD analysis where the outcomes are pre-treatment covariates (placebo outcomes)

Note: Only sorting could lead to violation of continuity assumption, but

- ▶ You might have sorting but no discontinuity in the potential outcomes
- ▶ You might have sorting and discontinuity in potential outcomes but no discontinuity in the density

Some evidence on sorting

- ▶ Quite a lot of evidence of sorting where population cutoffs affect municipal policies in Italy, France, Spain, Brazil (but not at 3,500 in France!)
- ▶ Some evidence of sorting in close elections to the U.S. House of Representatives since WWII (Caughey and Sekhon 2011) but anomalous and probably a fluke (Eggers et al 2015)



Pooled histograms from Italian censuses,
1961-2001 (Eggers et al 2017)

Introduction

Motivation: a non-RDD study of turnout and electoral systems

Regression discontinuity design: uses and assumptions

- Examples in political science

- The (quasi-)randomization approach to RDD

- The continuity approach to RDD

- RDD Implementation

Evaluating RDD assumptions

Example: effect of naturalization on political integration in Switzerland

Wrap-up

Appendix: diagrams

Overview

Hainmueller, Hangartner, and Pietrantuono (2015): “Naturalization Fosters the Long-Term Political Integration of Immigrants” PNAS

Broad question: does receiving citizenship (naturalization) make immigrants more engaged in politics (e.g. voting, knowledge, perceived efficacy)?

Setting: Switzerland, where some naturalization cases were *voted on* in referendums by secret ballot

Strategy: locate and interview winners and losers of naturalization referendums, compare via IV and **fuzzy RDD**

Sharp and Fuzzy RDD

- ▶ In a **sharp RDD**, probability of treatment changes from 0 to 1 at a threshold
 - ▶ electoral system is PR if and only if municipal population $\geq 3,500$
 - ▶ Democrat wins if and only if Democrat-Republican margin is greater than 0
- ▶ In a **fuzzy RDD**, probability of treatment changes discontinuously at a threshold, but not from 0 to 1
 - ▶ towns in New England (USA) below a certain size **may** have a town meeting (but some do not)
 - ▶ **here**: some people whose naturalization referendum is unsuccessful later become citizens anyway

Fuzzy RDD is a form of IV

You can use RDD to estimate

- ▶ the effect of winning a naturalization referendum on the probability of becoming a citizen (ITT_D , first-stage)
- ▶ the effect of winning a naturalization referendum on e.g. turnout (ITT_Y , reduced form)

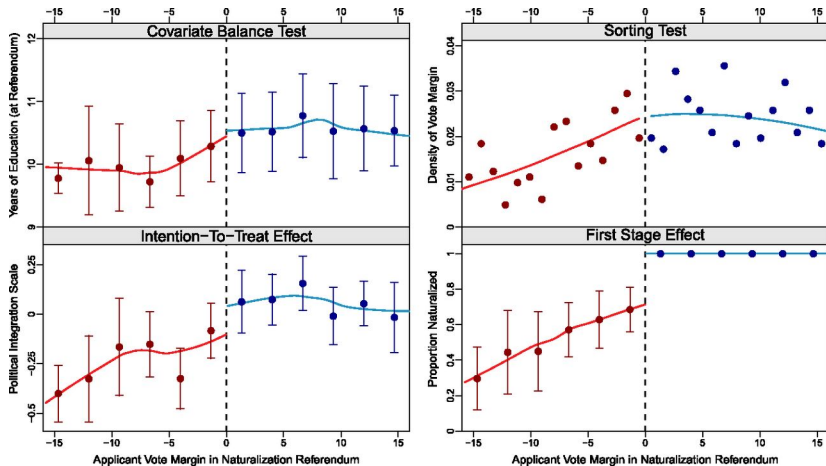
If you are willing to assume

- ▶ the outcome of a naturalization referendum only affects the person's political attitudes by affecting citizenship (**exclusion**)
- ▶ there is no one who becomes a citizen if and only if her naturalization referendum fails (**no defiers/monotonicity**)

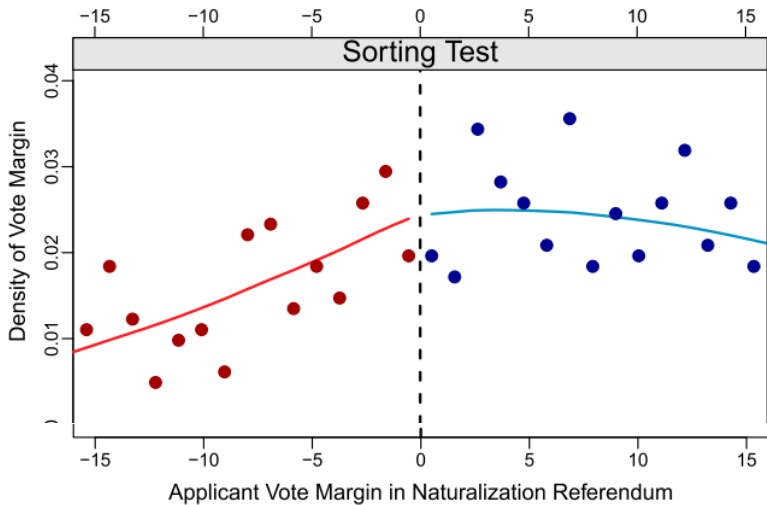
then you can estimate the effect of citizenship on e.g. turnout, using logic of Wald estimator.

Fuzzy RDD is an IV in which Z_i is based on RDD.

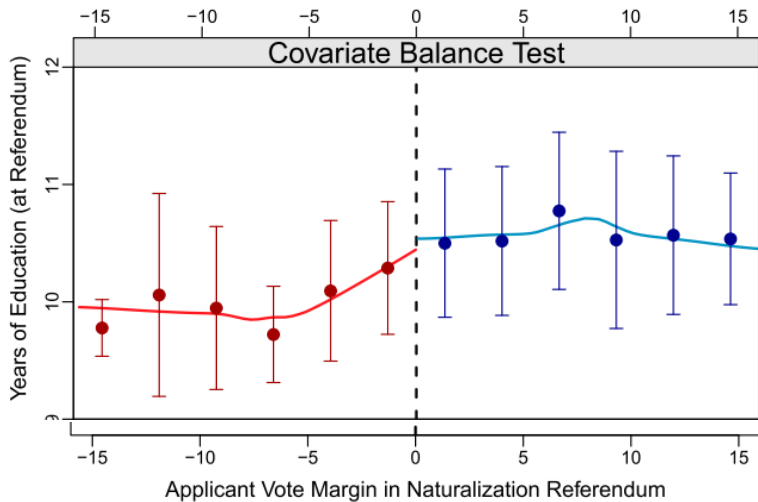
Hainmueller et al 2015 fuzzy RDD: key graphical results



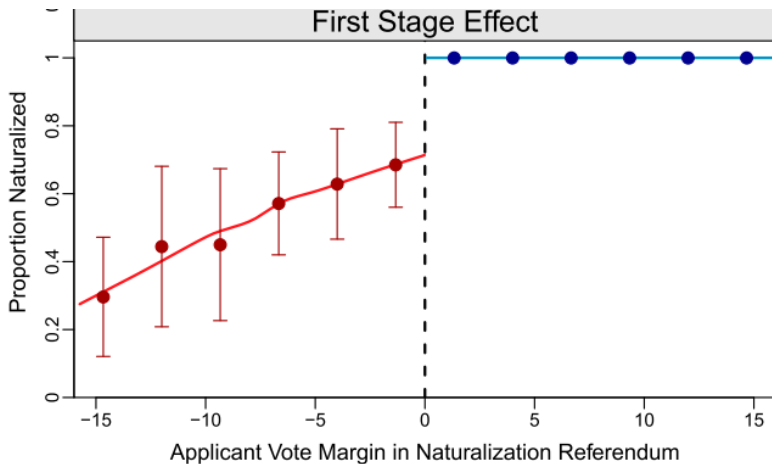
Hainmueller et al 2015: sorting test



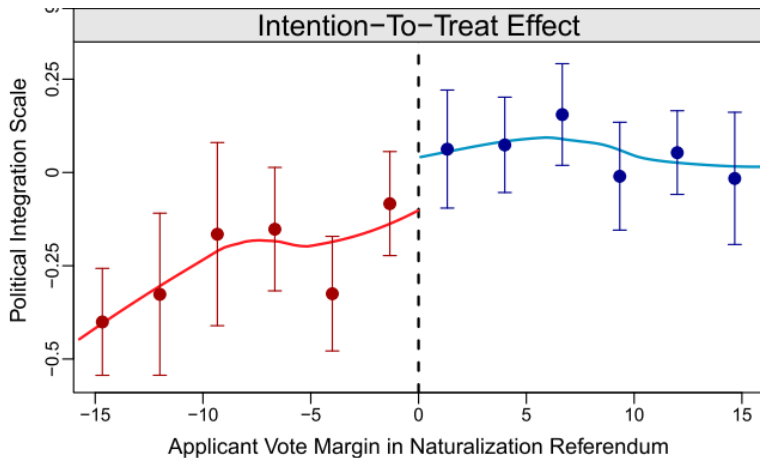
Hainmueller et al 2015: covariate balance test



Hainmueller et al 2015: first stage



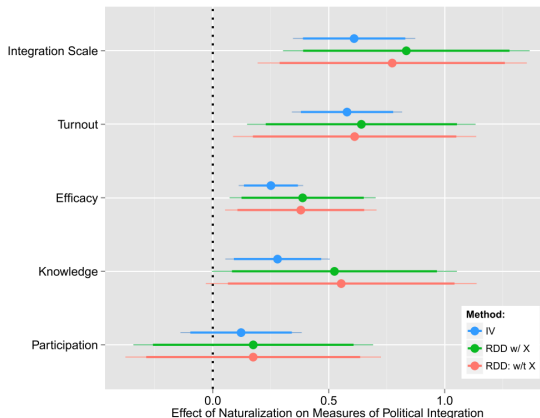
Hainmueller et al 2015: ITT



Hainmueller et al 2015: IV vs fuzzy RDD

They report two sets of analyses:

- ▶ “IV” uses **many** covariates to make instrument (“winning”) exogenous
- ▶ “fuzzy RDD” uses **one** covariate (running variable in RDD) to make instrument (“winning”) exogenous



Introduction

Motivation: a non-RDD study of turnout and electoral systems

Regression discontinuity design: uses and assumptions

- Examples in political science

- The (quasi-)randomization approach to RDD

- The continuity approach to RDD

- RDD Implementation

Evaluating RDD assumptions

Example: effect of naturalization on political integration in Switzerland

Wrap-up

Appendix: diagrams

Beware bad RDDs!

Would this ever make sense?

Some political scientists define democracy as being 6 or higher in Polity score. I will study the effect of democracy on redistribution with an RDD using a 6 on Polity as the cutoff.

Concluding thoughts

- ▶ like IV, RDD requires special circumstances – most things don't depend on cutoffs (but many do)
- ▶ appealingly simple, transparent technique

Introduction

Motivation: a non-RDD study of turnout and electoral systems

Regression discontinuity design: uses and assumptions

- Examples in political science

- The (quasi-)randomization approach to RDD

- The continuity approach to RDD

- RDD Implementation

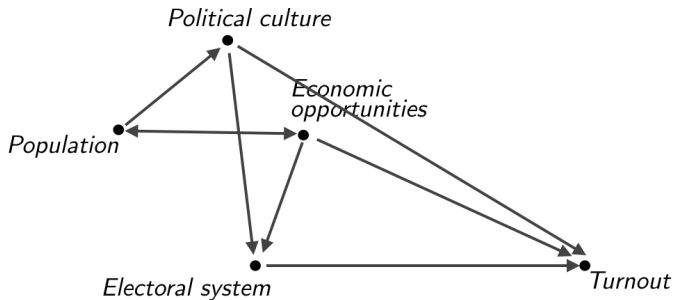
Evaluating RDD assumptions

Example: effect of naturalization on political integration in Switzerland

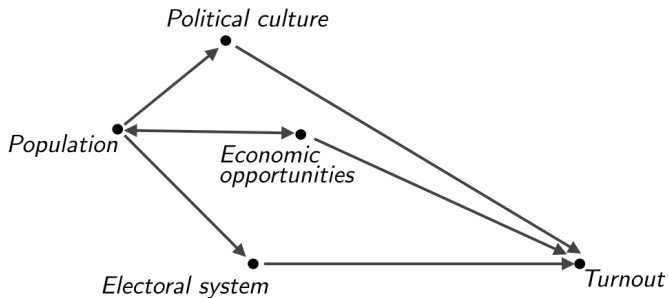
Wrap-up

Appendix: diagrams

The usual situation: e.g. Ladner & Milner



When electoral system depends on population threshold



When electoral system depends on population threshold, and you are at the threshold

