Lecture 10: Regression Discontinuity Designs (Part II)

Jaakko Meriläinen

5304 Econometrics @ Stockholm School of Economics

Regression Discontinuity Designs: A Quick Recap

- We introduced RDDs as a potential tool for uncovering causal effects using observational data
- Our focus was on binary treatments which were assigned, at least in part, based on an underlying observed variable:
  - Such that the probability of treatment "jumped" discontinuously at the threshold
  - With the assumption that nothing else jumped discontinuously at the threshold, allows us to uncover treatment effects
  - As if we had run a randomized trial even though we are working entirely with observational data
- Two cases:
  - **Sharp RDD:** Pr(treatment) goes from 0 to 1 at the threshold
  - **Fuzzy RDD:** Pr(treatment) increases discontinuously at the threshold but by less than 1
- Today, we will look at issues in implementing and interpreting RDDs more carefully

Plan for today

1. Review of RDD concepts
   - Sharp RDD
   - Fuzzy RDD
2. Functional form of CEFs
   - Investigating functional forms
   - Global polynomials
   - Local polynomial analyses of RD
3. Example: Exam schools in New York and Boston
4. Example: Incumbency advantage in Finland
5. Extensions, summary, and references

Regression Discontinuity Designs

- The basic idea is simple
- There is some underlying variable (X_i) that determines whether you get access to a program <u>based on a threshold</u>
- On crossing the threshold, the probability of treatment jumps discontinuously
- If potential outcomes vary continuously over the threshold, we can get consistent estimates
- At the cutoff, it is as-good-as-random whether you end up to the left-hand or the right-hand side of the cutoff—i.e., whether a unit is treated or not

Sharp RDD

- In sharp RDD, $W_i = 1X_1 \geq c$
- We can then look at the jump in the conditional expectation of the outcome (Y_i) given the covariate (X_i) at the threshold c

$$\lim_{x \downarrow c} E(Y_i|X_i = x) - \lim_{x \uparrow c} E(Y_i|X_i = x)$$

which can be interpreted as the average causal effect of treatment at the discontinuity point

$$\tau_{SRD} = E[Y_i(1) - Y_i(0)|X_i = c]$$

Sharp RDD (Imbens and Lemieux 2008)

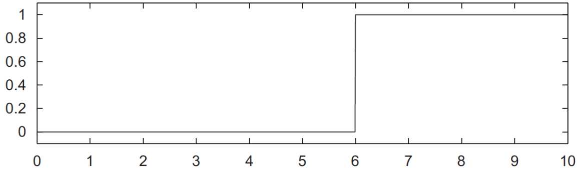


Fig. 1. Assignment probabilities (SRD).

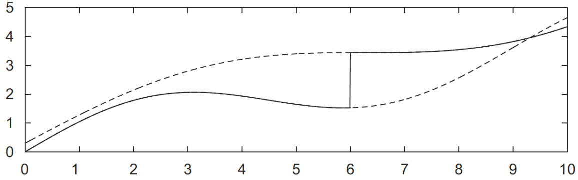


Fig. 2. Potential and observed outcome regression functions.

Fuzzy RDD

- To estimate a causal effect, we do not require the probability of treatment to go from zero to one at the discontinuity
- We just need it to be discontinuous ("jump")
- As long as that is the case, we can get consistent estimates from the Wald Estimator:

$$\tau_{FRD} = \frac{\lim\limits_{x \downarrow c} E(Y_i|X_i = x) - \lim\limits_{x \uparrow c} E(Y_i|X_i = x)}{\lim\limits_{x \downarrow c} E(W_i|X_i = x) - \lim\limits_{x \uparrow c} E(W_i|X_i = x)}$$

- This should look familiar by now
- It is the reduced-form (ITT) estimate divided by the first-stage— fuzzy RDD is IV!

Fuzzy RDD (Imbens and Lemieux 2008)

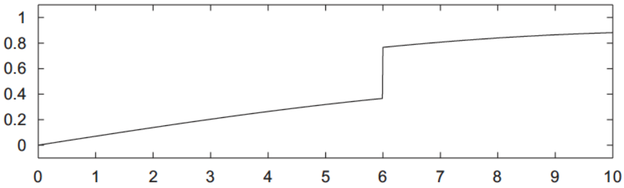


Fig. 3. Assignment probabilities (FRD).

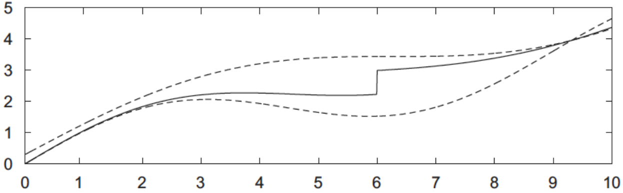


Fig. 4. Potential and observed outcome regression (FRD).

Functional Forms—Is the Discontinuity Really There?

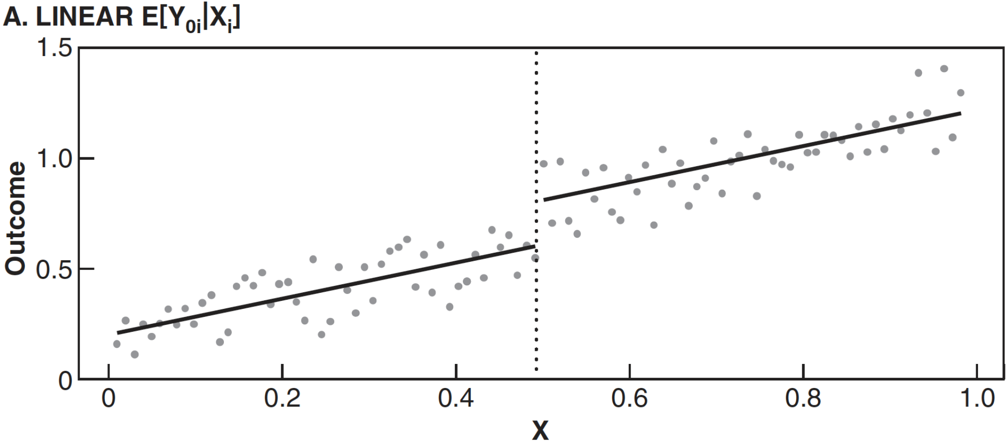
- The RDD relies for identification on the relationship between the assignment variable and the probability of treatment

- On either side of the threshold, the assignment variable and the probability of treatment could have a continuous relationship
  - For instance, the probability of treatment may be increasing with greater values of X on either side of $X = c$

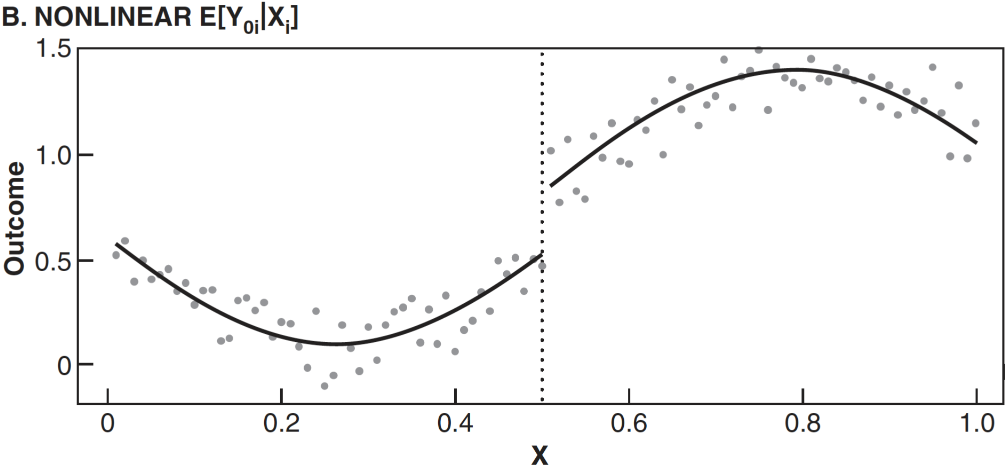
- The important thing is that, **at the threshold**, this probability should increase/decrease **discontinuously**
  - Based on this, the outcome must also jump discontinuously at the threshold

- There is a risk that we may mistake for a (smooth) non-linearity for a discontinuity
  - In which case our "discontinuity" does not really exist!
  - You **must** check for this graphically!

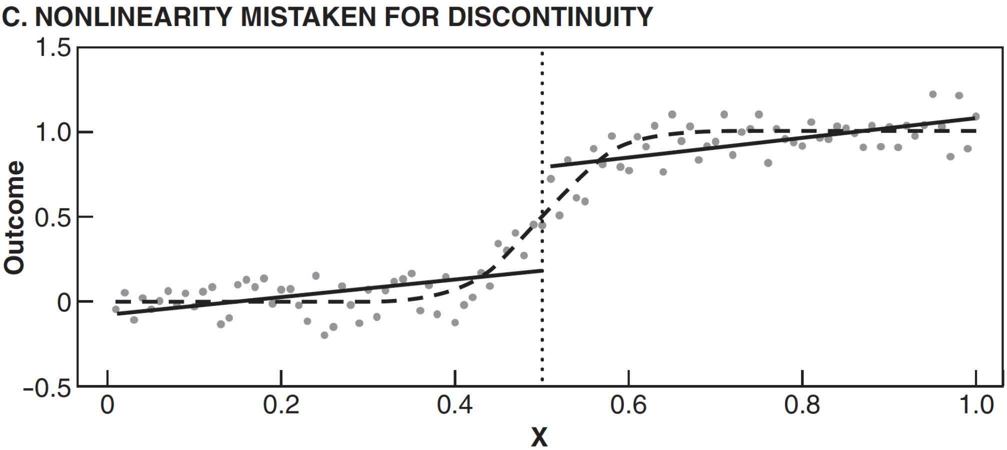
A Linear Relationship (Angrist and Pischke 2009)



A Non-Linear Relationship (Angrist and Pischke 2009)



Beware—the Discontinuity Could Be Spurious! (Angrist and Pischke 2009)



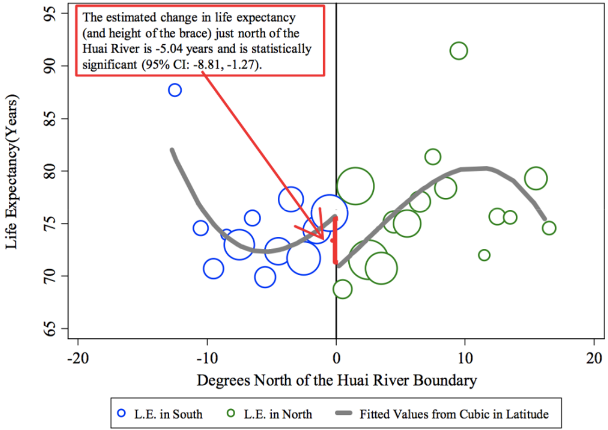
How to Tell What is Appropriate?

- The first step is always to visualize your data
  - See, for example, `binscatter` and `cmogram` in Stata
  - These both have RDD options—use them!
- What should you be looking for?
  - **Is there a discrete jump at the threshold?**
  - Is the relationship linear?
  - Is it similar on both sides of the threshold?
  - If non-linear, is it e.g. easily described by a quadratic control?
- A lot relies on getting this right
  - You must verify that the discontinuity is really there!
  - Your model specification should reflect the underlying data structure!

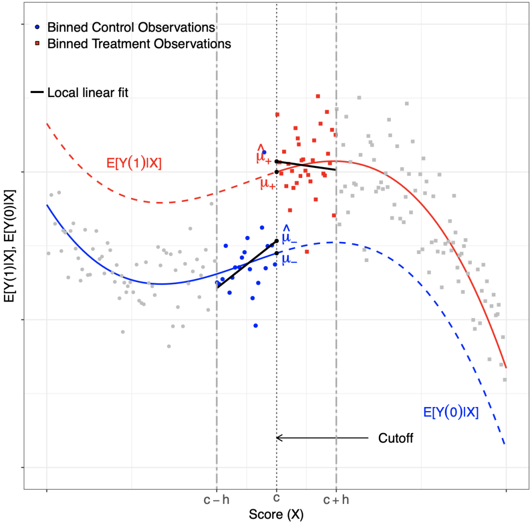
Control for Lots of Non-Linearities?

- One common way is to keep the full data but to include many non-linear terms ("global polynomials")
  - In practice, this means that you include not just x, but also squares, cubes etc. of x
  - Typically, we would also want to show the fit of the data to the estimated fit
  - We do this by showing "binned" means
- Turns out using global polynomials is not a reliable practice (Gelman and Imbens 2017)
  - It can lead to a major over-rejection of the null of no discontinuity
  - I.e. you see a discontinuity, even though there is no discontinuity in reality!
- There are other issues as well (see Gelman and Imbens 2017 for details)
- Results can be very sensitive to the degree of the polynomial

Is There a Discontinuity? (Chen et al. 2013)



Local Polynomial Regression with $p = 1$ (Cattaneo, Idrobo and Titiunik 2019)



Researchers Need to Make Many Choices!

1. Polynomial order?
2. Size of the bandwidth?
3. Weighting of the observations (choice of kernel)?

- (1) and (2) can make a big difference for the results!
- (3) is usually not a big deal
- We will next discuss each of these and best practices

How Far from the Threshold Should You Consider?

- In an RCT with perfect compliance, the full treatment group is comparable to the control group
- In an RDD, under maintained assumptions, observations "infinitesimally" away from the threshold are comparable
- How do you determine how close to the threshold you should consider?
- There is a stark bias/variance trade-off here
- The further away from the threshold you consider, the more observations you have (the precision of your estimates will be high)
- **This comes at a cost:** further from the threshold, your observations get less comparable (the more your identification relies on getting the functional form right; the higher the risk of bias!)

Optimal Bandwidths

- Let the data speak when using the bandwidth!
- Researchers typically rely on optimally selected bandwidths that minimize the MSE of the regression
- The math behind this is rather involved, so we will skip it—in practice, statistical software will do the estimation for you based on the data you have and the polynomial you want to use
- On Stata, `rdrobust` package is useful (for bandwidth estimation, see also `rdbwselect`)
- Something similar is also available on R

Choice of Polynomial

- In principle, one could have a high-order polynomial ($p = 1$ means linear, $p = 2$ means quadratic, $p = 3$ means cubic,....)
- In practice, higher-order polynomials lead to over-fitting and have poor properties at boundary points
- So, typically, researchers use $p = 1$ or $p = 2$
  - This means, running a linear or a quadratic regression within the selected bandwidth
  - A separate regression is run on each side of the threshold
- Placebo checks where the RDD is estimated at artificially chosen thresholds can reveal problems with the chosen model (Hyytinen et al. 2019)

Choice of Polynomial

- If there is strong curvature in the data close to the cutoff (NB. does not happen always), local linear specification can be problematic with an optimally chosen bandwidth
- Calonico, Cattaneo, and Titiunik (2014) propose a bias-corrected robust approach
- Their recommendation is to take the point estimate from the local linear specification with its optimal bandwidth and use a confidence interval that adjusts for the curvature issue
- In practice: Estimate a model with the optimal bandwidth for the local linear polynomial but use $p = 2$ and take the CI from this
- Theory behind this: Too involved for this course

Choosing a Kernel (Cattaneo, Idrobo, and Titiunik 2019)

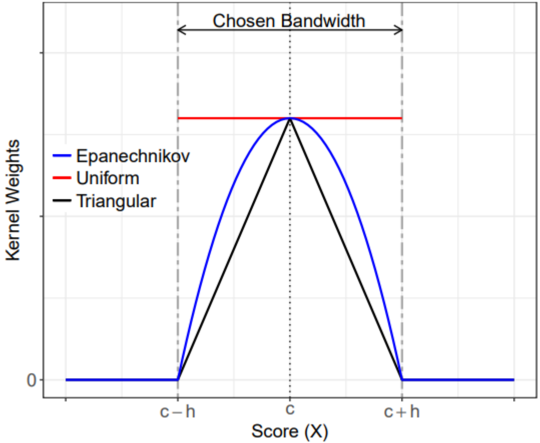


Figure 13: Different Kernel Weights for RD Estimation

Bias in Local Approximations (Cattaneo, Idrobo and Titiunik 2019)

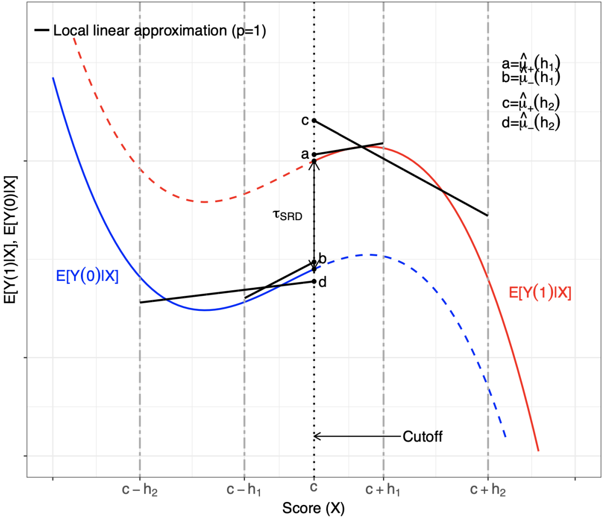


Figure 14: Bias in Local Approximations

Bias in Local Approximations (Cattaneo, Idrobo, and Titiunik 2019)

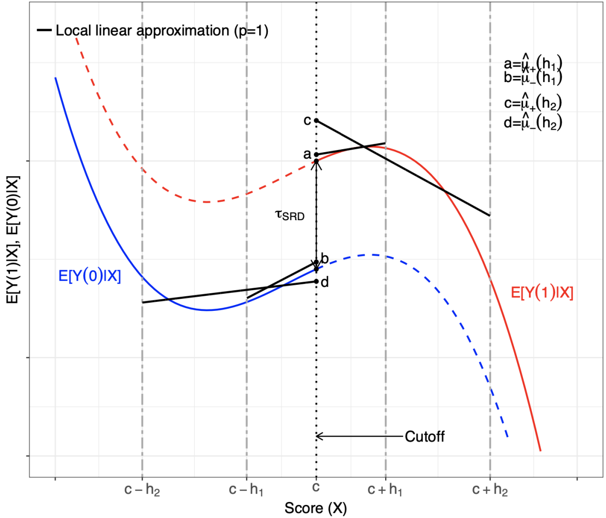


Figure 14: Bias in Local Approximations

Recommendations: What Should You Do?

- Plot the data (including the binned means)
  - What is the functional form approximately like?
  - Is a linear or quadratic approximation justifiable?

- If justifiable, run linear/quadratic regression in a narrow window of $\pm h$ observations from the threshold
  - I.e. throw out observations more than $\pm h$ away from the threshold and run your analysis as normal
  - Use optimal bandwidths but also show results with different values of h
  - NB. This can be quite data-intensive (bias-precision trade-off also kicks in)

- Assess identifying assumption: check the distribution of the running variable (manipulation and heaping), McCracy test, and covariate smoothness

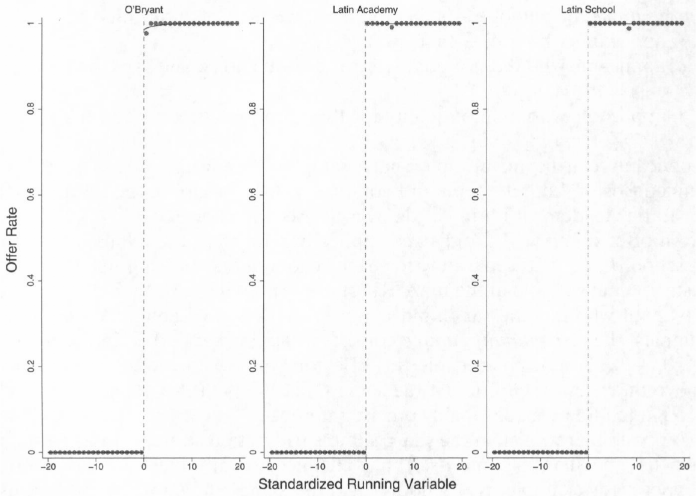
- Check that there are no discontinuities in placebo regressions

- You can also explore robustness to different kernels, but this typically matters less

Exam Schools in Boston

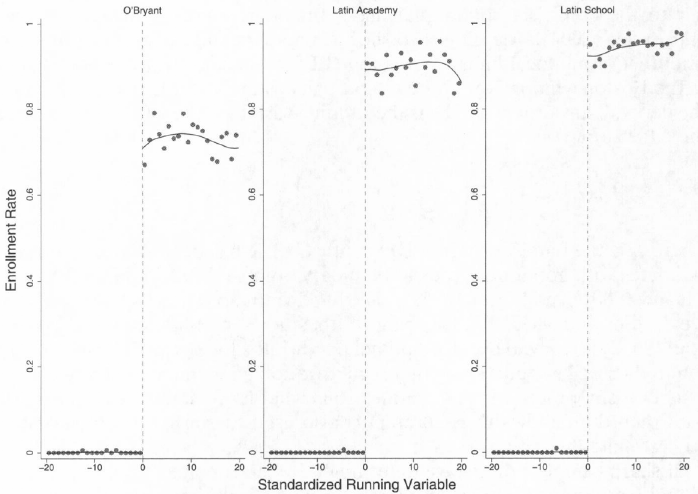
- Abdulkadiroglu, Angrist and Pathak (2014) look at very selective "Exam Schools" in Boston and New York
- These are selective public schools with much higher scoring pupils than the typical public system
- In Boston, the students are selected based on a combination of mechanisms
  - Suffice it to say that there is a single summary variable
  - Those just above the cut-off for a school are made offers
- Take-up is imperfect—not everyone who is offered may take up

Exam Schools in Boston (Abdulkadiroglu et al. 2014): Discontinuity in Offers



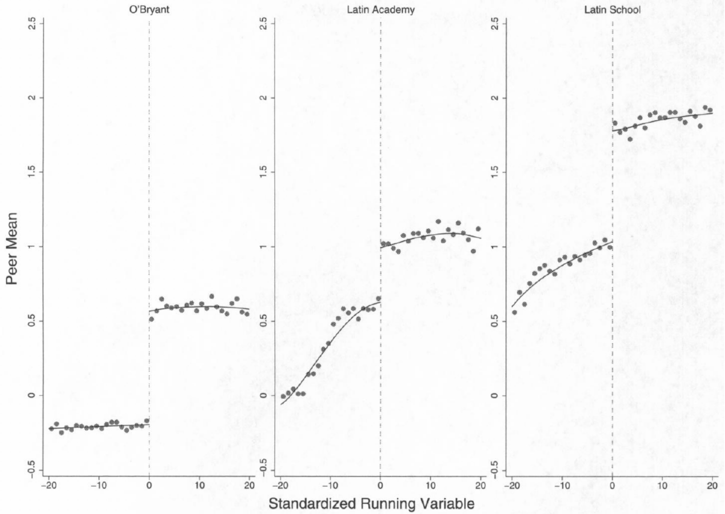
(a) Offers at each Boston exam school

Exam Schools in Boston: Discontinuity in School-Specific Enrollment



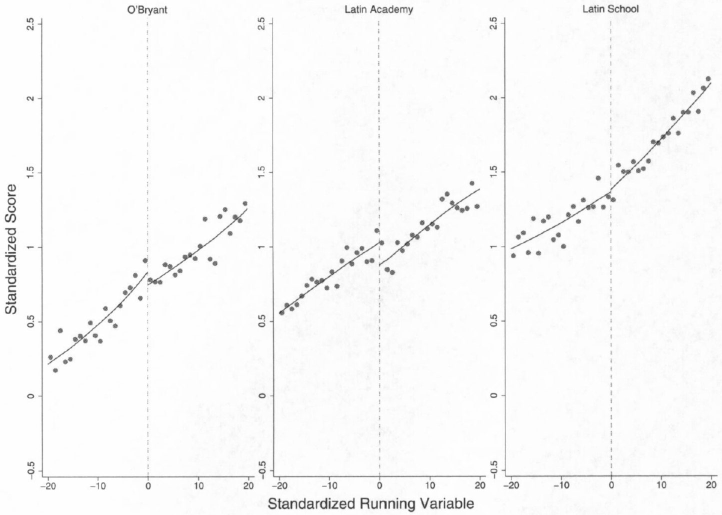
(b) Enrollment at each Boston exam school

Exam Schools in Boston: Discontinuity in Peer Group



(a) Baseline peer math score at Boston exam schools for 7th and 9th grade applicants

Exam Schools in Boston: Discontinuity in Outcomes (ITT)



(a) 7th and 8th grade math at Boston exam schools for 7th grade applicants

Randomized Experiment versus RDD

- RDD has become one of the most popular quasi-experimental tools—not just in economics but also in other social sciences
- The popularity of RDD is not surprising given that it is often heralded as one of the few observational study designs that is successful in approximating experimental benchmarks
- Theoretically, RDD mimicks a randomized experiment that occurs exactly at the cutoff
- Let us look into real-world data from such an experiment!
  - These data come from Hyytinen et al. (2018) who study election lotteries in Finnish local councils
  - Voters have to vote for an individual candidate (with some party affiliation), and lotteries occur when two or more candidates tie in votes for the last local council seat that their party obtains
  - Hyytinen et al. estimate the *personal incumbency advantage* using these lotteries and an RDD based on election margins

Incumbency Advantage in Election Lotteries

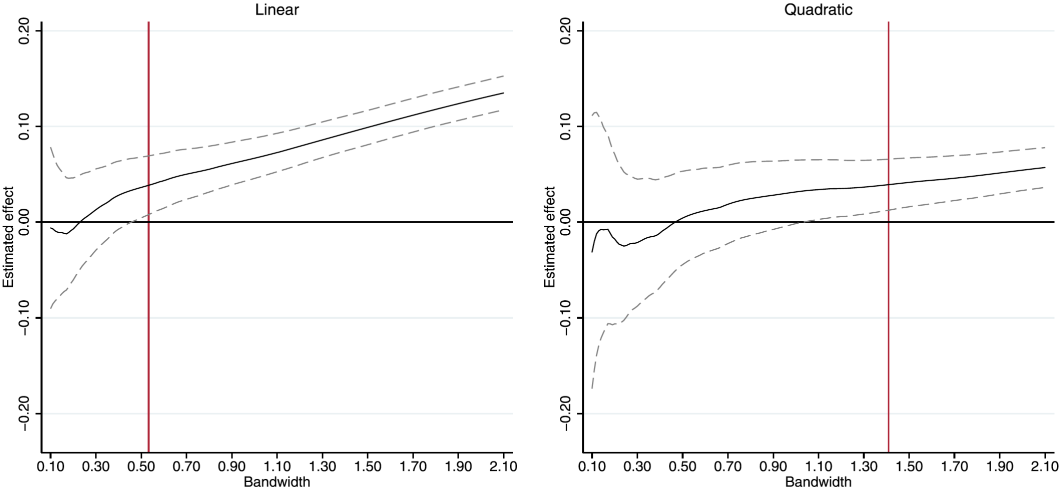
	(1)	(2)	(3)	(4)
Elected	0.004	0.001	−0.010	−0.010
95% confidence interval (robust)	[−0.046, 0.054]	[−0.049, 0.051]	[−0.064, 0.040]	[−0.060, 0.040]
95% confidence interval (clustered)	[−0.044, 0.053]	[−0.048, 0.050]	[−0.067, 0.047]	[−0.075, 0.055]
N	1351	1351	1351	1351
R^2	0.00	0.03	0.28	0.44
Controls	No	Yes	Yes	Yes
Municipality fixed effects	No	No	Yes	No
Municipality–year fixed effects	No	No	No	Yes

Note: Only actual lotteries are included in the regressions. Set of controls includes age, gender, party affiliation, socioeconomic status and incumbency status of a candidate, and total number of votes. Some specifications include also municipality or municipality–year fixed effects. Confidence intervals based on clustered standard errors account for clustering at the municipality level. The unit of observation is a candidate i at year t.

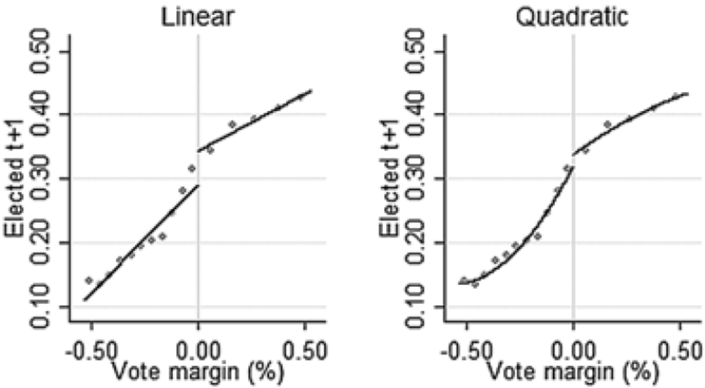
RDD Results

	(1)	(2)
	Linear	
Elected	0.039	0.052
95% confidence interval (robust)	[0.009, 0.070]	[0.027, 0.078]
95% confidence interval (clustered)	[0.011, 0.068]	[0.028, 0.077]
N	19,648	27,218
R^2	0.03	0.05
Bandwidth	0.54	0.74
Bandwidth selection method	IK	CCT

RDD Results Across Different Bandwidths



RDD Plots


Linear
Quadratic
Elected t+1
0.10 0.20 0.30 0.40 0.50
-0.50 0.00 0.50
Vote margin (%)

After Bias-Correction, Things Are Fine!

	(1)	(2)	(3)	(4)
	Linear		Quadratic	
Elected	0.039	0.052	0.008	0.022
95% confidence interval (robust)	[0.009, 0.070]	[0.027, 0.078]	[−0.038, 0.055]	[−0.018, 0.062]
95% confidence interval (clustered)	[0.011, 0.068]	[0.028, 0.077]	[−0.037, 0.053]	[−0.015, 0.059]
N	19,648	27,218	19,648	27,218
R^2	0.03	0.05	0.03	0.05
Bandwidth	0.54	0.74	0.54	0.74
Bandwidth selection method	IK	CCT	IK	CCT

Extensions

- There are several interesting extensions to what we have seen in these lectures—time permitting, we may discuss (some of) them briefly
  - RDDs with multiple cutoffs
  - Multi-dimensional RDDs
  - Regression kink design (RKD)
  - Difference-in-discontinuities
  - Non-parametric estimation
  - Extrapolation away from the cutoff
  - Bounds for treatment effects when treatment status is manipulated

Summary: Guide to Good Practices

- **Is there a discontinuity?**
  - Plot the data (including binned means and regression line)
  - Can the underlying function be reasonably approximated by a linear or quadratic polynomial?
- **Is the RD valid?**
  - Manipulation in the running variable?
  - Heaping in the running variable?
  - Discontinuities in covariates?
  - Discontinuities in placebo regressions?
- **How sensitive are results to bandwidth choice?**
  - How far from the threshold can you go?

Readings: Regression Discontinuity Design

- Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. Journal of Econometrics, 142(2), 615-635.

Optional

- Chapter 6 in Angrist and Pischke (2009)
- Lee, D. S., & Lemieux, T. (2010). Regression Discontinuity Designs in Economics. Journal of Economic Literature, 48, 281-355.
- Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2017). A practical introduction to regression discontinuity designs. Cambridge Elements: Quantitative and Computational Methods for Social Science-Cambridge University Press I
- Hyytinen, A., Meriläinen, J, Saarimaa, T., Toivanen, O., & Tukiainen, J. (2018). When does regression discontinuity design work? Evidecne from random election outcomes. Quantitative Economics 9(2), 1019-1051.

Other References Mentioned in This Lecture

- Abdulkadiroğlu, A., Angrist, J., & Pathak, P. (2014). The elite illusion: Achievement effects at Boston and New York exam schools. Econometrica, 82(1), 137-196.
- Gelman, A., & Imbens, G. (2017). Why high-order polynomials should not be used in regression discontinuity designs. Journal of Business & Economic Statistics.
- McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. Journal of Econometrics, 142(2), 698-714.