Regression Discontinuity Design

A. Colin Cameron
Univ. of California, Davis

These slides are part of the set of slides
A. Colin Cameron, Introduction to Causal Methods
https://cameron.econ.ucdavis.edu/causal/

March 2023

Introduction

- These slides give an introductory example of regression discontinuity design (RDD)
  - RDD is a method for causal inference
  - it can be applied when treatment occurs when a variable that determines in part the outcome crosses a threshold
  - it lends itself to graphical analysis.
- RDD relies on the assumption that there is no manipulation at the threshold
  - and estimation varies with how the outcome is modeled either side of the threshold.

- Separately the Stata file `rdd.do` implements these methods
  - using dataset AED_INCUMBENCY.DTA
- The data are from chapter 13.7 of A. Colin Cameron (2022) *Analysis of Economics Data: An Introduction to Econometrics* https://cameron.econ.ucdavis.edu/.
  - and in more detail in A. Colin Cameron and Pravin K. Trivedi (2022) *Microeconometrics using Stata: Volume 2*, chapter 25.7.
- Data are originally from Sebastian Calonico, Matias Cattaneo, Max Farrell, and Rocío Titiunik (2017), "Rdrobust: Software for Regression-discontinuity Designs," *The Stata Journal*.

Outline

1. Introduction
2. Regression Discontinuity Design
3. Example: Gains to Political Incumbency
4. Results
5. Further Details
6. References

Regression Discontinuity Design

- A threshold variable (denoted s) determines treatment status
  - e.g. admission into treatment is based on a score denoted s
  - with scores above 100, say, leading to treatment $(d = 1)$.
- A simple RDD estimate compares the average value of outcome variable y for individuals on either side of the threshold.
- Complication: usually y itself varies with s (called a running variable).
- Suppose that $y = \beta_1 + \beta_2 s + u$ without treatment
  - then a simple RDD estimate of ATET is $\widehat{\gamma}$ from OLS of
    - ★ $y_i = \beta_1 + \gamma d_i + \beta_2 s_i + u_i$.
- In practice more flexible models are used
  - e.g. different linear or quadratic trends on either side of the threshold
    - ★ $y_i = \beta_1 + \gamma d_i + \beta_2 d_i \times s_i + \beta_3 d_i \times s_i^2 + \beta_4 (1 - d_i) \times s_i + \beta_5 (1 - d_i) \times s_i^2 + u_i$.

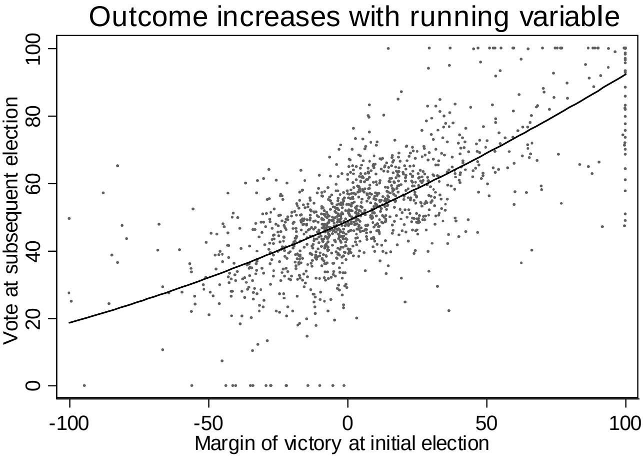
Gains to Political Incumbency

- Does being an incumbent on its own cause an increase the probability of winning the next election?
  - compare impact of just losing and just winning the previous election.
- Data on 1,390 U.S. Senate seat elections from 1914 to 2010.
- Running variable $s = \texttt{margin} =$ Democratic party's margin of victory in a Senate election
- Outcome variable $y = \texttt{vote} =$ Democratic party's vote share in the subsequent Senate election (usually six years later).
- Cutoff variable $d = 1$ if $\texttt{margin} > 0$ (so incumbent)

Variable	Obs	Mean	Std. dev.	Min	Max
margin	1,390	7.171159	34.32488	-100	100
vote	1,297	52.66627	18.12219	0	100
win	1,390	.5395683	.4986113	0	1

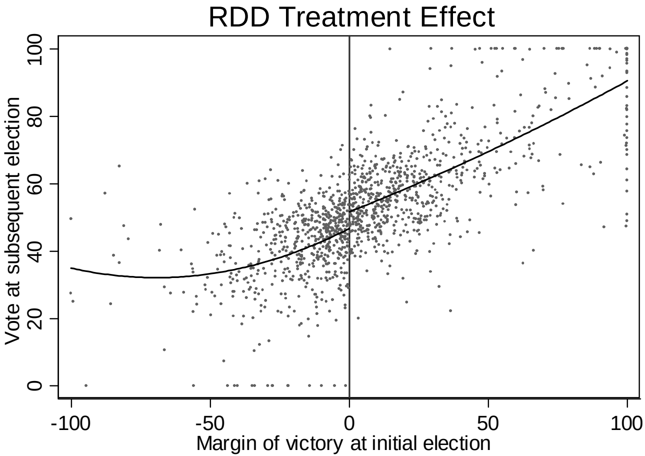
Vote (y) increases with margin (s)

- The running variable has an effect on the outcome.



Graphical analysis with quadratic on each side

- Fit separate quadratics either side of $d = 1$ (margin>0)
  - eyeballing suggests effect is around 5%.



Regression estimate with quadratic on each side

- Fit separate quadratics either side of $d = 1$ (margin> 0)
  - the estimated effect is a 4.9% boost if win previous election

```
. * Regression analysis - quadratic on each side
. generate winmarg = win*margin

. generate marginsq = margin^2

. generate winmargsq = win*marginsq

. regress vote win margin marginsq winmarg winmargsq, vce(robust) noheader
```

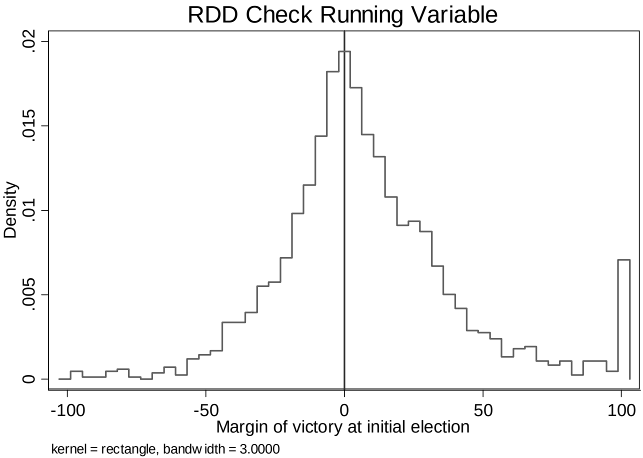
vote	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
win	4.934817	1.128685	4.37	0.000	2.72056	7.149074
margin	.4206294	.07891	5.33	0.000	.2658234	.5754353
marginsq	.0030298	.0012332	2.46	0.014	.0006105	.0054492
winmarg	-.094709	.0973555	-0.97	0.331	-.2857014	.0962833
winmargsq	-.0024027	.0013705	-1.75	0.080	-.0050913	.000286
_cons	46.73955	.842152	55.50	0.000	45.08742	48.39169

Further Details

- We could restrict analysis to close to the cutoff of zero
  - e.g. to `margin` between -25 and 25.
- Better graph replaces scatterplot of individual values with a scatterplot of bin averages of `vote` with, say, 20 bins either side.
- More sophisticated analysis uses local polynomials either side
  - use Stata user-written program `rdrobust` (used in MUS2 chapter 25.7).
- Fuzzy RD is a generalization that relaxes the assumption that all below the cutoff are untreated and all above the cutoff are treated.
- In some cases there can be manipulation of the running variable around the cutoff (in which case RDD estimates are invalid)
  - e.g. suppose students fail if course score is less than 50
    - ★ if teachers are reluctant to fail then expect a heaping just above 50
  - use a histogram or kernel density to visually check this
  - example is next.

Check no manipulation of running variable

- We do not expect manipulation in this example
  - and indeed see no heaping around `margin` $= 0$.



References on RDD

- D. S. Lee and T. Lemieux (2010), "Regression discontinuity designs in economics," *Journal of Economic Literature,* 48, pages 281–355.
- These books are given in approximate order of increasing difficulty.
- A. Colin Cameron (2022), Analysis of Economics Data: An Introduction to Econometrics, chapters 13.7 and 17.5.
- Joshua D. Angrist and Jörn-Steffen Pischke (2015), Mastering Metrics, Princeton University Press, chapter 4.
- Cunningham, Scott (2021), Causal Inference: The MixTape, Yale UP, chapter 9.
- A. Colin Cameron and Pravin K. Trivedi (2022), Microeconometrics using Stata: Volume 2, Second Edition, Stata Press, chapter 25.7.
- Joshua D. Angrist and Jörn-Steffen Pischke (2009), Mostly Harmless Econometrics: An Empiricist's Companion, Princeton University Press, chapter 6.
- A. Colin Cameron and Pravin K. Trivedi (2005), Microeconometrics: Methods and Applications, Cambridge University Press, chapter 25.6.
- Wooldridge, Jeffrey M. (2010), Econometric Analysis of Cross Section and Panel Data, Second Edition, MIT Press, chapter 21.5.

References on RDD by non-economists

- These books by non-economists are similar to *Mastering Metrics* in accessibility.
- Stephen L. Morgan and Christopher Winship (2015), Counterfactuals and Causal Inference: Methods and Principles for Social Research, Second edition, Cambridge University Press, chapter 11.2.
- Richard J. Murnane and John B. Willett (2010), Methods Matter: Improving Causal Inference in Educational and Social Science Research, Oxford University Press, chapter 9.
- Andrew Gelman, Jennifer Hill and Aki Vehtari (2022), Regression and Other Stories, Cambridge University Press, chapter 21.3.