Applied Econometrics JEM007, IES
Lecture 7

REGRESSION DISCONTINUITY DESIGN

Eva Hromádková, 18.11.2010

Regression Discontinuity Design
Introduction

- Based on a special type of natural experiments
- Occurs when the probability of participation in treatment discontinuously changes with the continuous variable z (forcing variable)
- Z itself may be associated with the potential outcomes, but this association is assumed to be smooth.
- Thus, any discontinuity of the outcome as a function of this covariate at the cutoff value is interpreted as evidence of a causal effect of the treatment.
- Summary in Imbens and Lemieux (2008)

Regression Discontinuity Design

Sharp design

- Probability of participation in treatment *d* changes from 0 to 1 at threshold z*

$$\lim_{z \to z^{*-}} \mathrm{P}(d = 1 | z) = \mathrm{P}(z^{*-}) = 0$$

$$\lim_{z \to z^{*+}} \mathrm{P}(d = 1 | z) = \mathrm{P}(z^{*+}) = 1$$

- treatment status is exogenously determined by the level of threshold – thus not affected by individual decision
- We estimate the effect of the treatment around the threshold – non-treated on one side are correct "counterfactuals" of treated on the other

RD-Sharp design

Graphical illustration

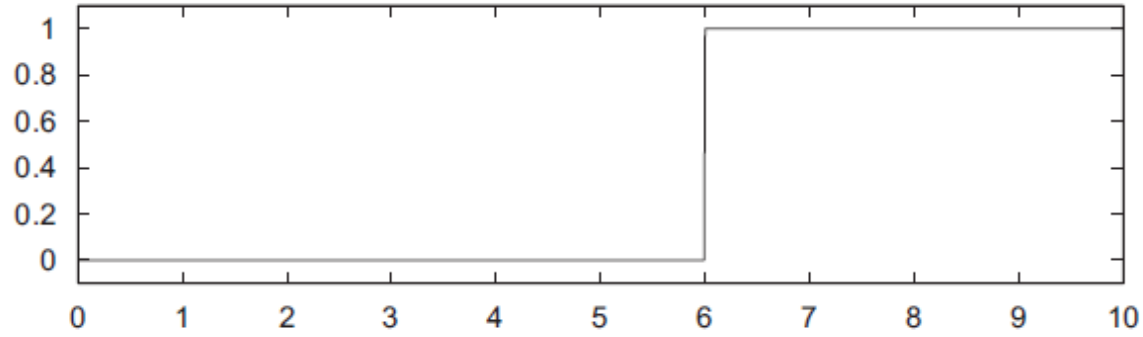


Fig. 1. Assignment probabilities (SRD).

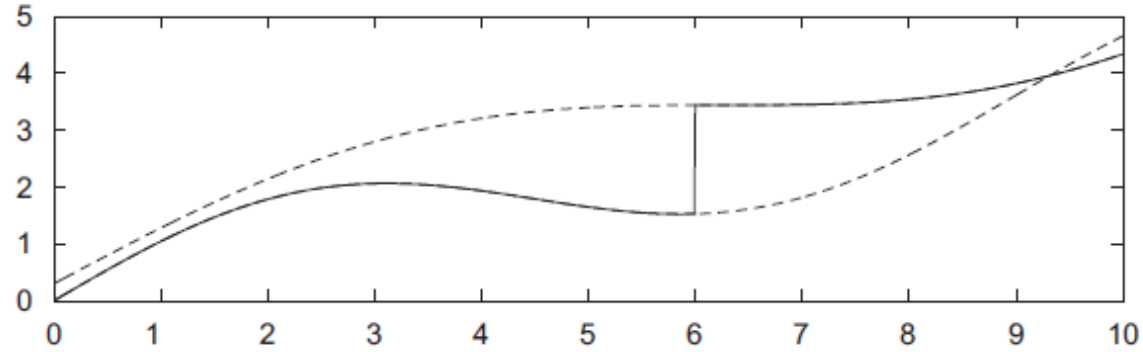


Fig. 2. Potential and observed outcome regression functions.

RD-Sharp design

Graphical illustration

Eligibility based on age

- Card and Shore-Sheppard (2004):
  - expansion of the Medicaid system to cover low income children of certain age range
  - eligibility rule was based on date of birth

Regression Discontinuity Design

Fuzzy design

- *z* does not fully determine participation in treatment
- Other unobserved factors also affect participation
- The discontinuity is not clear-cut and probability can change in smaller steps
- Incentives to participate in a program are not strong enough to move al from non-participation to participation over the particular treshold z*

RD-Fuzzy design

Graphical illustration

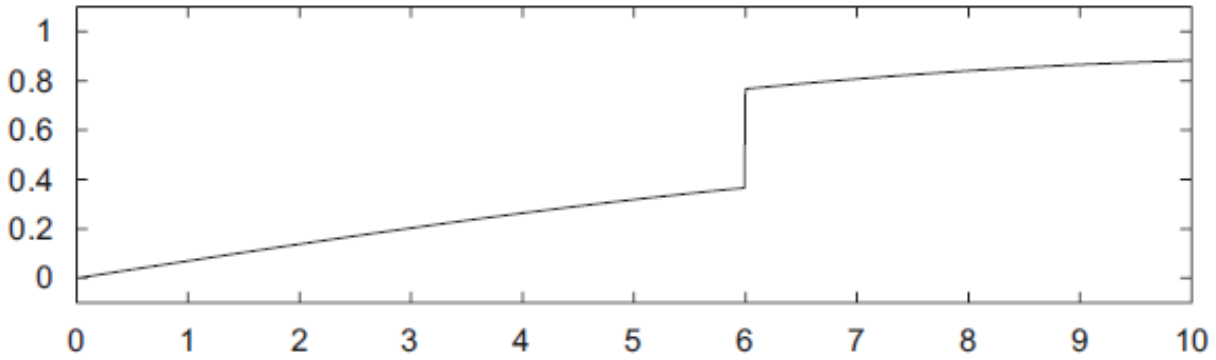


Fig. 3. Assignment probabilities (FRD).

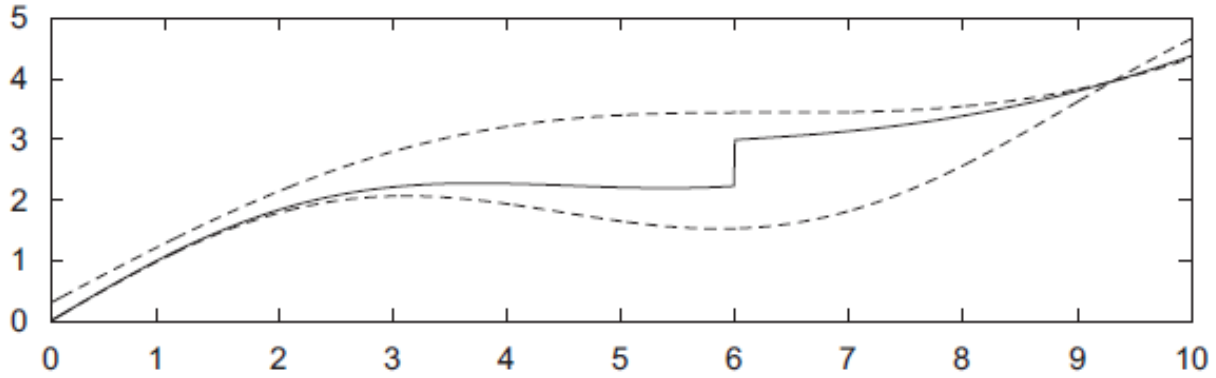


Fig. 4. Potential and observed outcome regression (FRD).

RD-Fuzzy design

Examples

- Effect of remedial teaching program (encouraged for students with score less than “c” (Matsudaira, 2007)
  - The effect is estimated only on compliers – those who are affected by the threshold “c” and decide to enroll
- Van der Klaauw (2002): effect of financial aid on college admission acceptance
  - SAT and other scores determine eligibility for financial aid – different groups
  - What is the effect of financial aid?
    - Causal effect: aid offer attracts more students to the college
    - Other effects: students with higher financial aid have usually better outside option from other schools
    - Other effects than discontinuity might matter too

IV vs. RD

- RD is equivalent to regressing outcome *Y* on treatment *d* using *z>c* (forcing variable) as the instrument, applied on the subsample of z from [c-h, c+h]
- RD is less demanding about exogeneity:
  - IV: z is exogenous (excluded)
  - RD: (i) Y is a continuous function of z at c; (ii) cut off level is exogenous

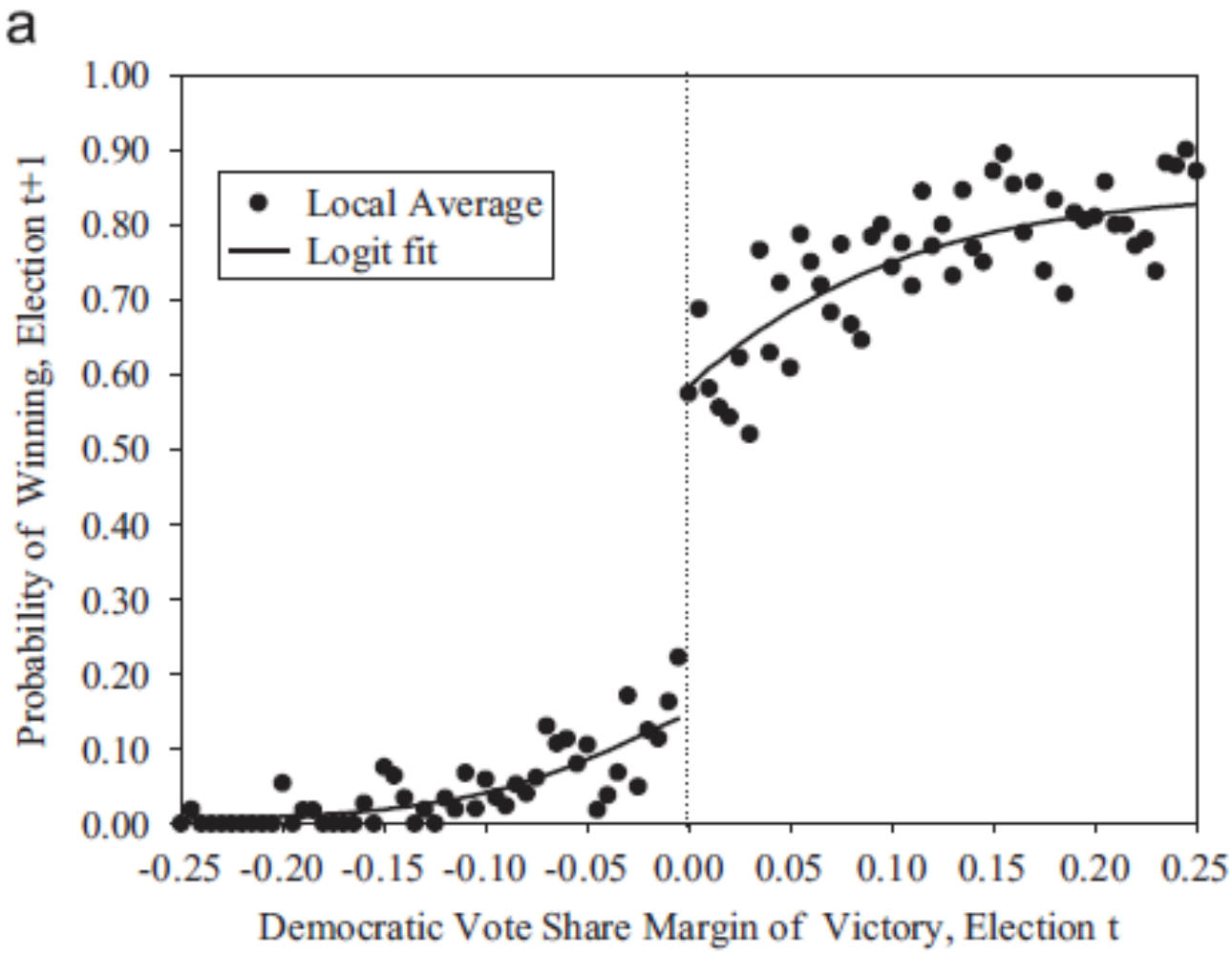
Implementation of RD (sharp)

Step 1: graphic analysis

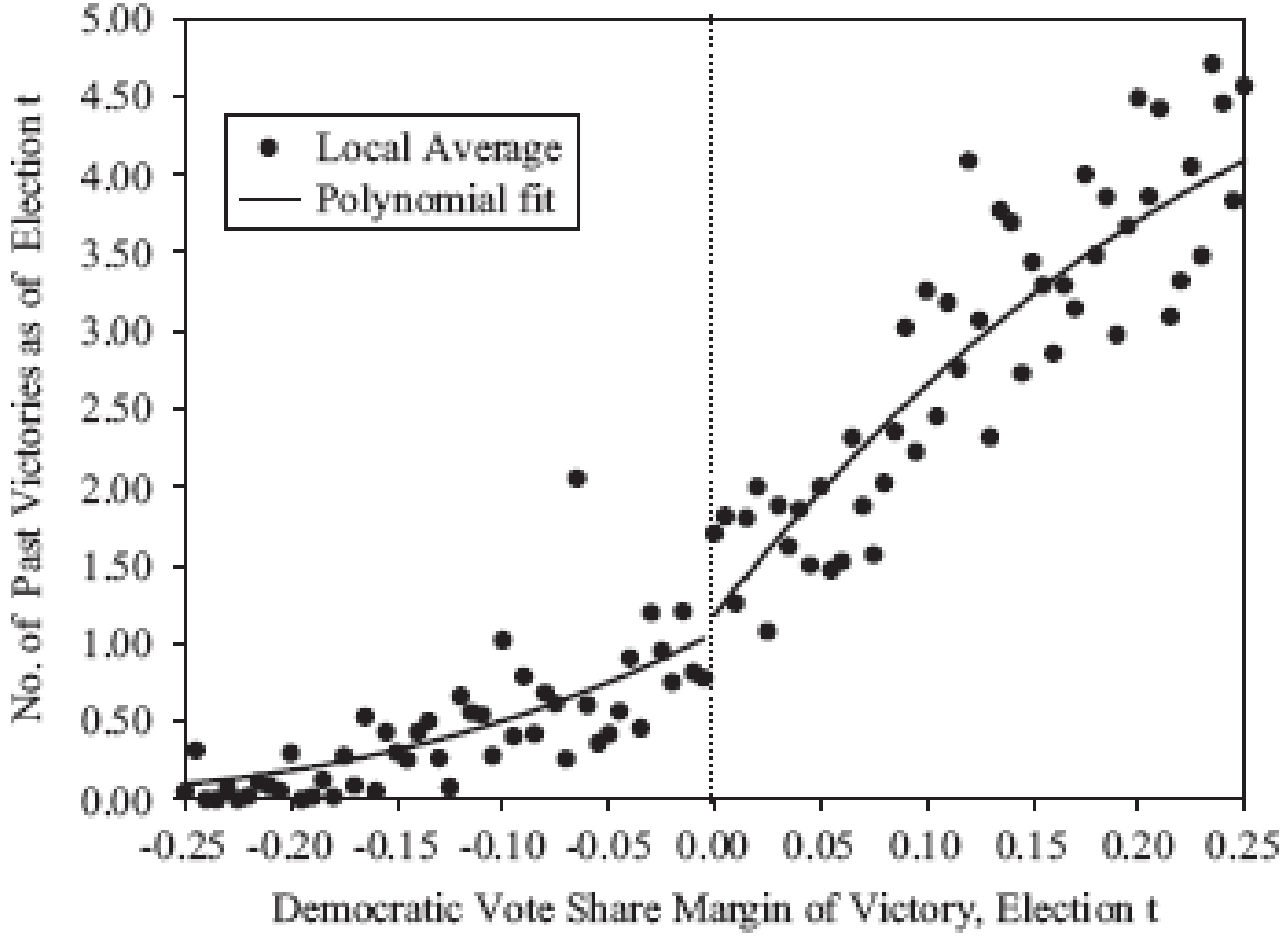
- Plot *d* (treatment) vs. *z* (forcing variable): Is there a jump at c (cutoff)?
- Plot *Y* (outcome) vs. *z:* Is there a jump at c?
- Plot other covariates vs. z: Is there NO jump at c?
- Plot the density of z: Look out for clustering of people just above c – indication of manipulation

Example: Lee (2007)

Looking at incumbent advantage in elections for Congress



b


Local Average
Polynomial fit
No. of Past Victories as of Election t
5.00
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
-0.25 -0.20 -0.15 -0.10 -0.05 0.00 0.05 0.10 0.15 0.20 0.25
Democratic Vote Share Margin of Victory, Election t

Implementation of RD (sharp)
Step 2: local estimation

Simplest:

- Choose a window width h
- Calculate $E(Y \mid c-h \leq z < c)$ and $E(Y \mid c \leq z < c+h)$
  - This is like fitting constant
- Problem 1: need lot of data in the neighborhood of c
- Problem 2: if Y is linear in z, the bias is linear in h

Local linear regression: (non-parametric methods)

- We fit linear regression functions to the observation within distance h of discontinuity point
- Calculate value at cut-off -> use to estimate effect

RD-summary of issues

- Identifies only local effect restricted to the discontinuity point
- Once the design is fuzzy
  - The discontinuity applies only to compliers
  - Unobserved factors can drive decision
  - If individuals can manipulate to which side of threshold belong based on their expectations, problems are even more severe
  - Similar problems with IV and LATE