Lecture 12: Difference in Differences and Regresssion Discontuity Design

Introduction ot Econometrics,Fall 2017

Zhaopeng Qu

Nanjing University

12/25/2017

Review the last lecture

Panel Data

Difference in Differences(Cha.13)

Introduction

- A typical RCT design requires a causal studies to do as follow
  1. Randomly assignment of treatment to divide the population into a "treatment" group and a "control" group.
  2. Collecting the data at the time of post-treatment then comparing them.
- It works because *treatment* and *control* are randomized.
- what if we have the treatment group and the control group, but they are not fully randomized?
- If we have observations across two times at least(one before treatment, the other after treatment), then an easy way to make causal inference is **Difference in Differences(DID)** method.

DID with Regression

- Formally, an simple DID regression is

$$Y_{it} = \beta_0 + \beta_1 T_i + \beta_2 D_t + \beta_3 (T_i \times D_t) + u_{it}$$

- T_i: a time dummy denotes pre($T_i = 0$) or post treatment($T_i = 1$),thus
- D_t: a treatment dummy denotes in treatment($D_t = 0$) or control group($D_t = 1$)
- $T_i \times D_t$: an interaction term

DID with Regression

- $$Y_{it} = \beta_0 + \beta_1 T_i + \beta_2 D_t + \beta_3 (T_i \times D_t) + u_{it}$$

	Treatment group	Control group	Difference
Before	$\beta_0 + \beta_2$	β_0	β_2
After	$\beta_0 + \beta_1 + \beta_2 + \beta_3$	$\beta_0 + \beta_1$	$\beta_2 + \beta_3$
Differece	$\beta_1 + \beta_3$	β_1	β_3

- then DID estimator

$$\hat{\beta}_{DID} = (\bar{Y}_{treat,after} - \bar{Y}_{treat,before}) - (\bar{Y}_{control,after} - \bar{Y}_{control,before})$$

DID estimator

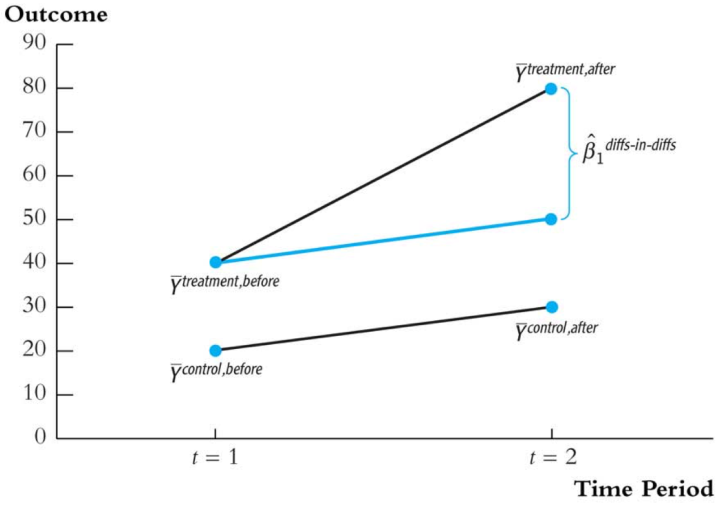


Figure 1:

Key Assumption For DID

- A key identifying assumption for DID is: **Common Trend**
  - Treatment would be the same "trend" in both groups in the absence of treatment.
- There are some unobservable factors affected on outcomes of both group. But as long as the effects have the same trends on both groups, then DID will eliminate the factors.
- Advantage: Omitted variables and other bias are likely reduced by the use of the untreated comparison group
- Other changes are not likely to always influence all groups in the same way.

Card and Krueger(1994): minimum wage on employment

- Theoretically,in competitive labor market, increasing binding minimum wage decreases employment.But what about the reality?
- Ideal experiment: randomly assign labor markets to a control group (minimum wage kept constant) and treatment group (minimum wage increased), compare outcomes.
- Policy changes affecting some areas and not others create natural experiments.
    - Unlike ideal experiment, control and treatment groups not randomly assigned.

Card and Krueger(1994): Backgroud

- Policy Change: in April 1992
  - Minimum wage in New Jersey from $4.25 to $5.05
  - Minimum wage in Pennsylvania constant at $4.25
- Research Design:
  - Collecting the data on employment at fast food restaurants in NJ(treatment group) in Feb.1992 (before treatment)and again November 1992(after treatment).
  - Also collecting the data from the same type of restaurants in eastern Pennsylvania(PA) as control group where the minimum wage stayed at $4.25 throughout this period.

Card & Krueger(1994): Geographic background

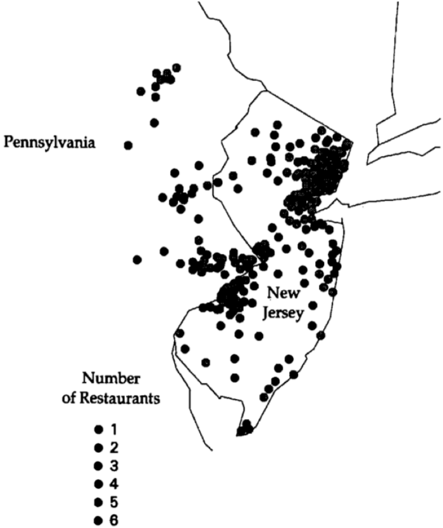


Figure 2:

Card & Krueger(1994):

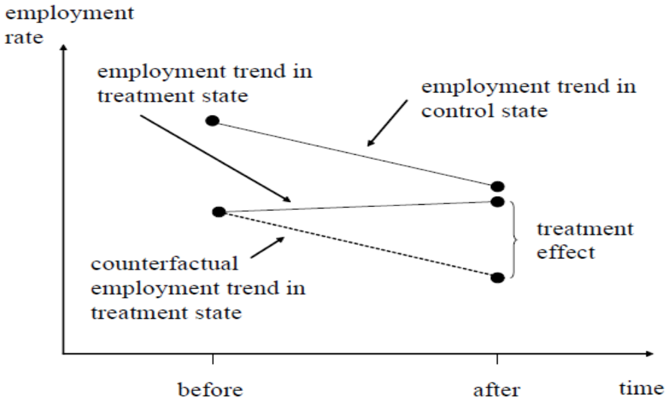


Figure 5.2.1: Causal effects in the differences-in-differences model

Figure 2:

Card & Krueger(1994):Result

Table 5.2.1: Average employment per store before and after the New Jersey minimum wage increase

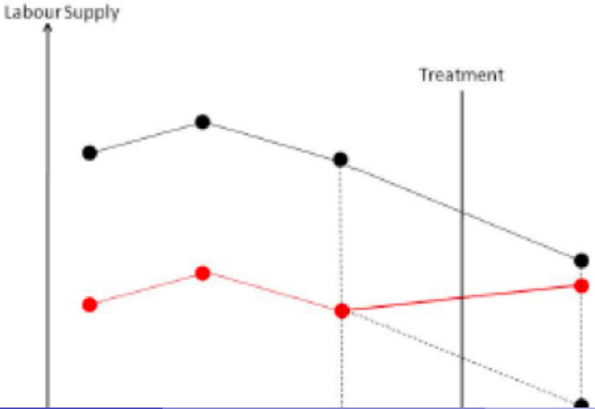
Variable	PA (i)	NJ (ii)	Difference, NJ-PA (iii)
1. FTE employment before,	23.33	20.44	-2.89
all available observations	(1.35)	(0.51)	(1.44)
2. FTE employment after,	21.17	21.03	-0.14
all available observations	(0.94)	(0.52)	(1.07)
3. Change in mean FTE	-2.16	0.59	2.76
employment	(1.25)	(0.54)	(1.36)

Notes: Adapted from Card and Krueger (1994), Table 3. The

Figure 4:

Assessing natural experiment

- Common Trend: Common trend assumption is difficult to verify but one often uses pre-treatment data to show that the trends are the same.



An Encouraging Example: Pischeke(2007)

- Topic: the length of school year on student performance
- Backgroud:
  - Until the 1960s, children in all German states except Bavaria started school in the Spring. In 1966-1967 school year, the Spring moved to Fall.
  - It make two shorter school years for affected cohort, 24 weeks long instead of 37.
- Reseach Design:
  - Dependent Variable: Retreating rate
  - Independent Variable: spending time on school
  - Treatment group: Students in the German **States except Bavaria**.
  - Control group: Students in **Bavaria**.

An Encouraging Example: Pischeke(2007)

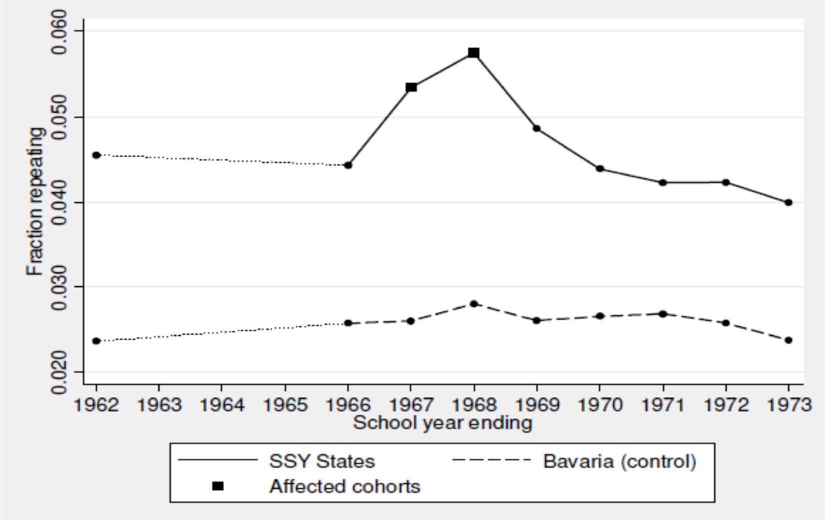


Figure 6:

Pischke(2007)

- This graph provides strong visual evidence of treatment and control states with a common underlying trend.
- A treatment effect that induces a sharp but transitory deviation from this trend.
- It seems to be clear that a short school years have increased repetition rates for affected cohorts.

Extensions of DID: DDD

- More convincing analysis sometime comes from higher-order contrasts: **DDD** or **Triple D** design.
    - Build the third dimension of contrast to eliminate the potential bias.
- e.g: Health Plan for elderly
    - Treatment group: Elderly aged above 65 in treatment place.
    - Control group 1: Elderly aged between 55-65 in treatment place.
    - Assumption 1 : the elderly on different age would have the same trends of health status if there were not the plan.
    - Control group 2: Elderly aged above 65 in control place.
    - Assumption 2 : the elderly in different places would have the same trends of health status if there were not the plan.
- It can loose the simple *common trend* assumption in simple DID.

Extensions of DID: Synthetic Controls Method

- In some cases, treatment and potential control groups do not follow parallel trends. Then standard DID method would lead to biased estimates.
- Synthetic Control (SC) is a method to evaluate the causal effect of treatment on aggregate outcomes of one (or very few) treated unit.
- The basic idea behind synthetic controls is that a combination of units often provides a better comparison for the unit exposed to the intervention than any single unit alone.
- It is a data-driven procedure to use a small number of non-treated units to build the suitable counterfactuals.

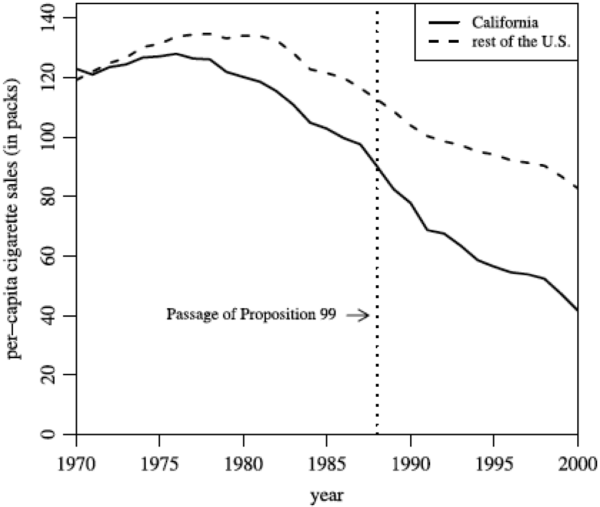
Extensions of DID: Synthetic Controls Method

- Use (long) longitudinal data to build the weighted average of non-treated units that best reproduces characteristics of the treated unit over time in pre-treatment period.
- The weighted average of non-treated units is the synthetic cohort.
- Causal effect of treatment can be quantified by a simple difference after treatment: treated vs synthetic cohort.

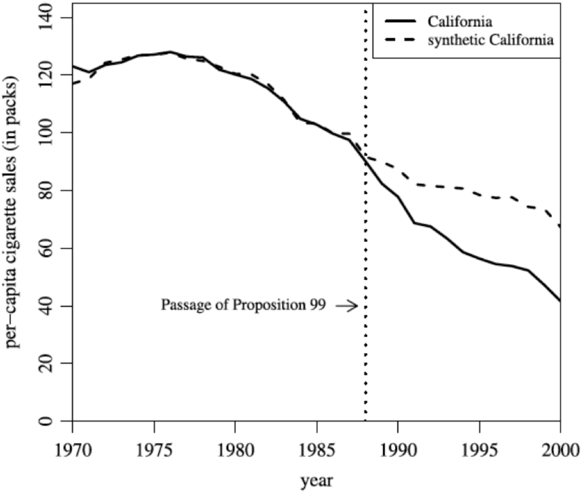
Abadie et. al (2010): tobacco tax on cigarette consumption

- In 1988, California passed comprehensive tobacco control legislation: Increased cigarette taxes by $0.25 per pack ordinances.
- estimate the effect of the policy on cigarette consumption in California.

Abadie et. al (2010): tobacco tax on cigarette consumption



Abadie et. al (2010): tobacco tax on cigarette consumption



Regression Discontinuity Design (RDD)

Main Idea of RDD

- Instead of controlling treatment assignment process, if researchers have detailed institutional knowledge of treatment assignment process
- Then we could use this information to create an "experiment"

Main Idea of RDD

- Regression Discontinuity Design (RDD) exploits the facts that:
  - Some rules are arbitrary and generate a discontinuity in treatment assignment.
  - The treatment assignment is determined based on whether a unit exceeds some threshold on a variable (**assignment variable**, **running variable** or **forcing variable**)
  - Assume other factors do not change abruptly at threshold.
  - Then any change in outcome of interest can be attributed to the assigned treatment.

A Motivating Example: Elite University

- A large number of studies have shown that graduates from more selective programs or schools earn more than others.
    - e.g Students graduated from NJU earn more than those graduated from other ordinary university.
- But it is difficult to know whether the positive earnings premium is due to
    - true "causal" impact of human capital acquired in the academic program
    - a spurious correlation linked to the fact that good students selected in these programs would have earned more no matter what(**Selection Bias**)

A Motivating Example: Elite University

- But say that the entry cutoff for a score of entrance exam is 400 at NJU.
- Those with scores 399 or even 395 are unlikely to attend NJU, instead attend NUFE(南京财经大学).
- Since the those get 399 and those get 400 are essentially identical,they get different scores due to some random events.
- **RD strategy**: I can do "as well" as in a randomized experiment by tracking down the long term outcomes for the 400 (admitted to NJU) and the 399 (admitted at NUFE)

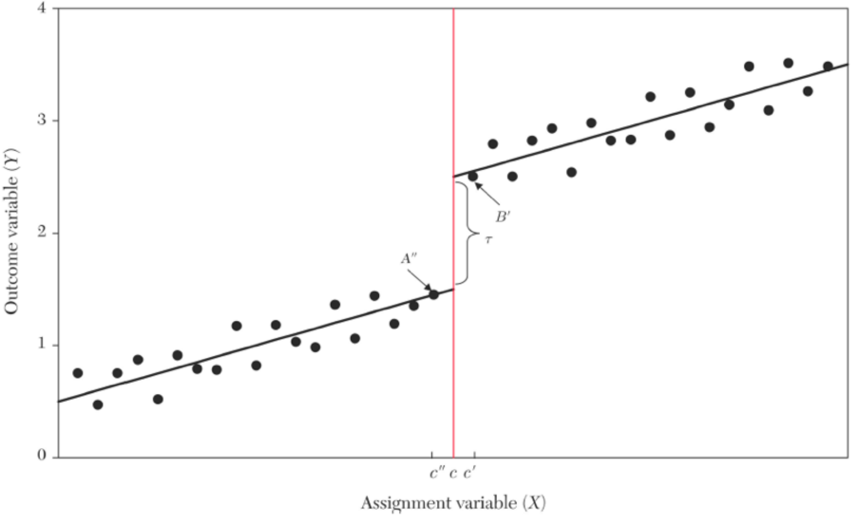
A Motivating Example: Elite University

- Mark Hoekstra (2009) "The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach" Review of Economics and Statistics
- This paper demonstrates the above RD idea by examining the economic return of attending the most selective public state university.
- In the United States, most schools used SAT (or ACT) scores in their admission process.
- For example, the flagship state university considered here uses a strict cutoff based on SAT score and high school GPA.

A Motivating Example: Elite University

- For the sake of simplicity, Hoekstra just focuses on the SAT score (adjusted depending on GPA).
- The author is then able to match (using social security numbers) students applying to the flagship university in 1986-89 to their administrative earnings data for 1998 to 2005.
- As in any good RD study, pictures tell it all, so let's just focus on those.

A Motivating Example: Test Score and Earnings



SAT Score and Enrollment

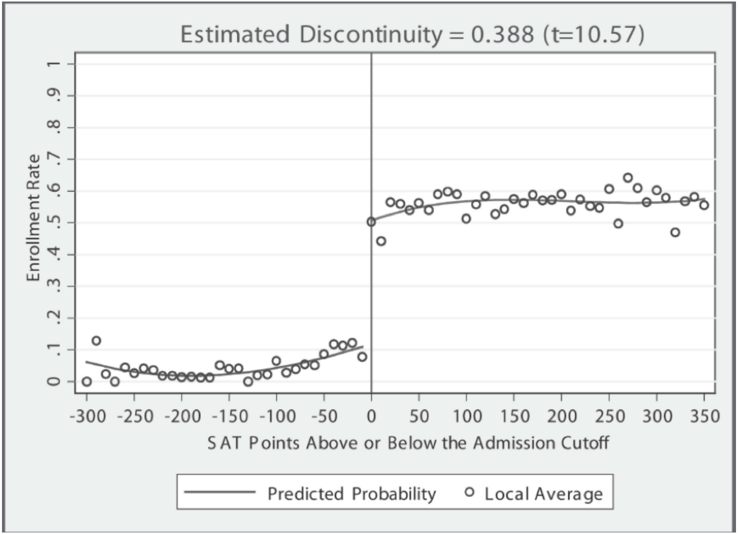


Figure 10:

SAT Score and Earnings

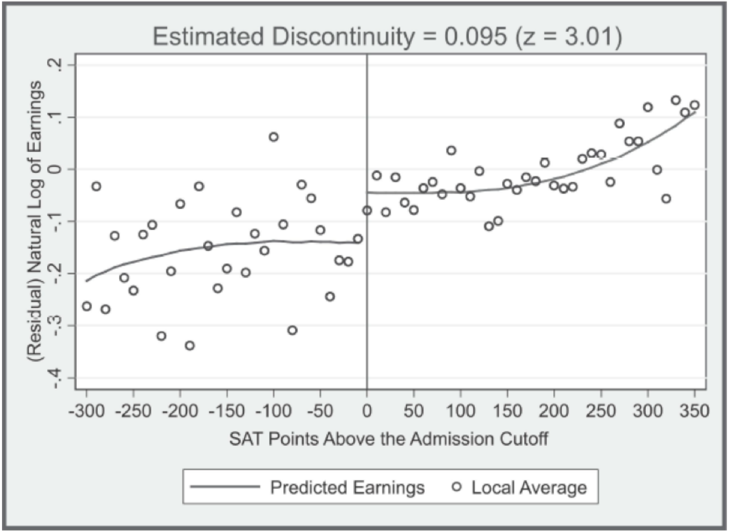


Figure 11:

The Effect of Health Intervention

- Prashant Bharadwaj, Katrine Vellesen Løken, and Christopher Neilson (2013) "Early Life Health Interventions and Academic Achievement".
- The effect of health intervention in early childhood on later life outcomes
- **Selection bias**: those who need health intervention in early childhood could be very sick and might have bad later life outcome (e.g. low educational attachment)
- **RDD solution**: infants with a birth weight below 1500 grams were eligible for additional healthcare while those with a birth weight just above the cutoff were not eligible
- Compares mortality rates and academic achievement between those infants *just below and above the cutoff of 1500 grams*

The Effect of Health Intervention

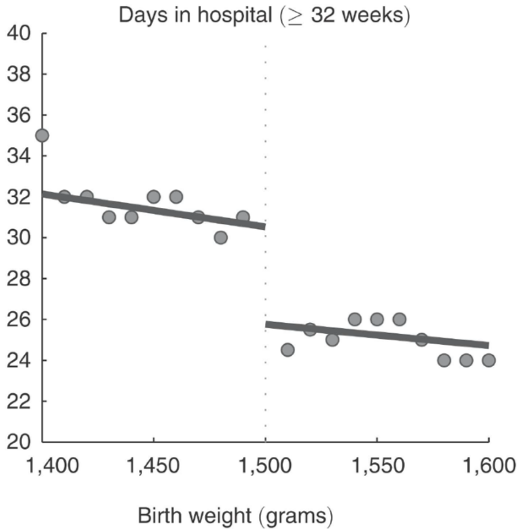
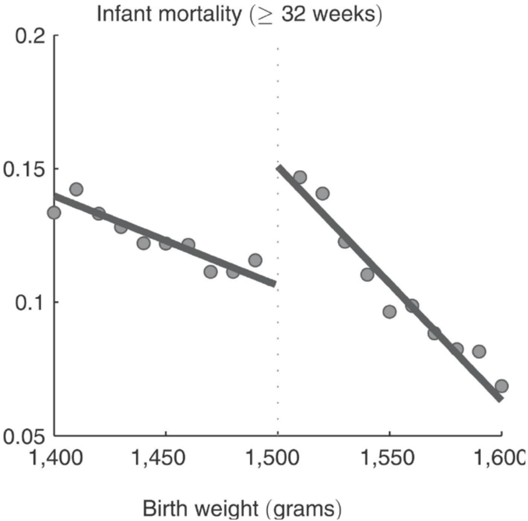


Figure 12:

The Effect of Health Intervention



The Effect of Health Intervention

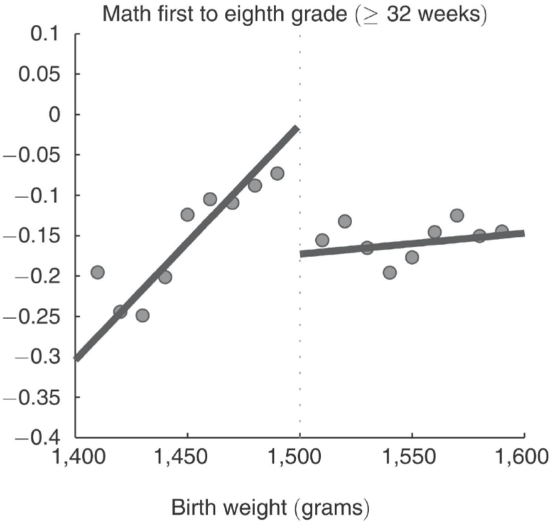


Figure 14:

RDD: theory and application

Sharp RDD and Fuzzy RDD

- In general, depending on enforcement of treatment assignment, RDD can be categorized into two types:
  1. **Sharp RDD**: nobody below the cutoff gets the "treatment", everybody above the cutoff gets it
     - Everyone follows treatment assignment rule (all are compliers).
     - Local randomized experiment with perfect compliance around cutoff.
  2. **Fuzzy RDD**: the probability of getting the treatment jumps discontinuously at the cutoff (NOT jump from 0 to 1)
     - Not everyone follows treatment assignment rule.
     - Local randomized experiment with partial compliance around cutoff.
     - Using initial assignment as an instrument for actual treatment.

Sharp RDD and Potential Outcomes

- Treatment
  - assignment variable (running variable):X_i
  - Threshold (cutoff) for treatment assignment:c
  - Treatment variable: D_i and treatment assignment rule is

$$D_i = 1\, if\; X_i \geq c\, and\, D_i = 0\, if\; X_i < c$$

- Potential Outcomes
  - Potential outcome for an individual i with treatment,Y_{1i}
  - Potential outcome for an individual i without treatment,Y_{0i}
- Observed Outcomes

$$Y_{1i}\, if\; D_i = 1(X_i \geq c)\, and\, Y_{0i}\, if\; D_i = 0(X_i < c)$$

Identification for Sharp RDD

- Intuitively, we are interested in the discontinuity in the outcome at the discontinuity in the treatment assignment.
- We can use sharp RDD to investigate the behavior of the outcome around the threshold

$$\alpha_{SRD} = lim_{\varepsilon \to 0} E[Y_i | X_i = c + \varepsilon] - lim_{\varepsilon \to 0} E[Y_i | X_i = c - \varepsilon]$$

- Under certain assumptions, this quantity identifies the ATE at the threshold

$$\alpha_{ATE} = E[Y_{1i} - Y_{0i} | X_i = c]$$

Identification for Sharp RDD

- **Deterministic Assumption**

$$D_i = 1(X_i \geq c)$$

- Treatment assignment is a deterministic function of the assignment variable X_i and the threshold c.
- **Continuity Assumption**
  - $E[Y_{1i}|X_i]$ and $E[Y_{0i}|X_i]$ are continuous at $X_i = c$
- Assume potential outcomes do not change at cutoff.
- This means that except treatment assignment, all other unobserved determinants of Y_i are continuous at cutoff c.
- This implies no other confounding factor affects outcomes at cutoff c.
- Any observed discontinuity in the outcome can be attributed to treatment assignment.

Graphical Interpretation

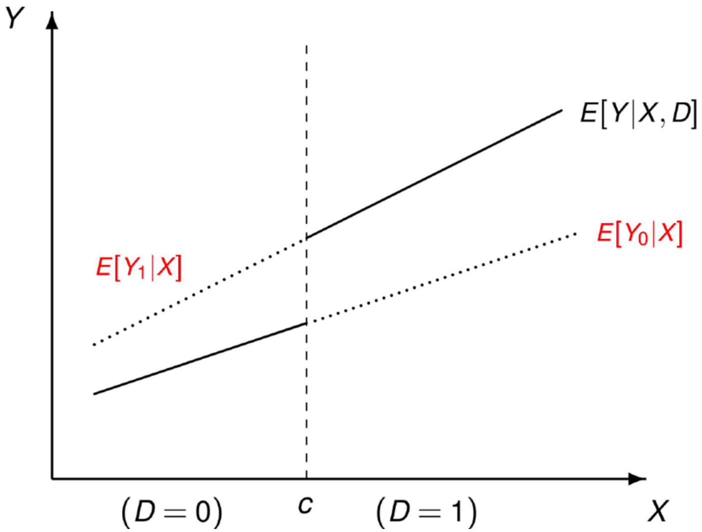


Figure 15:

Graphical Interpretation

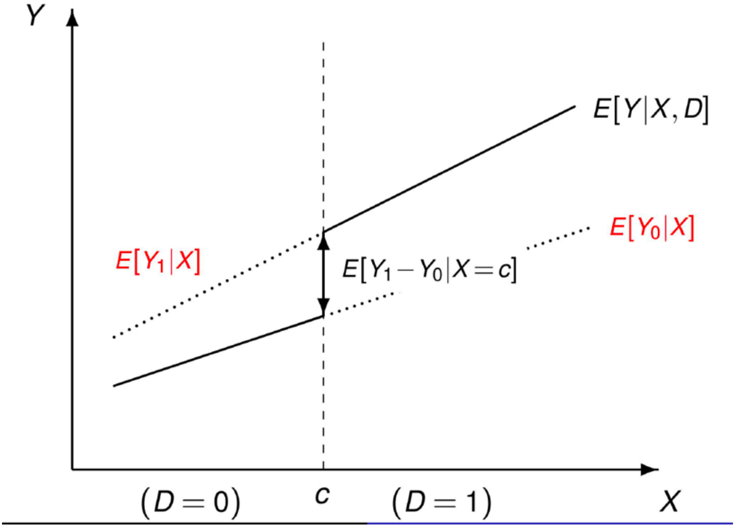


Figure 16:

Continuity Assumption

- Continuity is a natural assumption but could be violated if:
- There are differences between the individuals who are just below and above the cutoff that are NOT explained by the treatment.
- The same cutoff is used to assign some other treatment.
- Individuals can fully manipulate the running variable in order to gain access to the treatment or to avoid it.

Sharp RDD Estimation

- There are 2 types of strategies for correctly specifying the functional form in a RDD:
  1. Parametric/global method: Use all available observations and Estimate treatment effects based on a specific functional form for the outcome and assignment variable relationship.
  2. Nonparametric/local method:Use the observations around cutoff Compare the outcome of treated and untreated observations that lie within specific bandwidth.

Parametric/Global Approach

- Suppose that in addition to the assignment mechanism above, potential outcomes can be described by some reasonably smooth function $f(X_i)$

$$E[Y_{i0}|X_i] = \alpha + f(X_i)$$
$$Y_{i1} = Y_{i0} + \rho$$

- We can construct RD estimates by fitting

$$Y_i = \alpha + \rho D_i + f(X_i) + \eta_i$$

- There are essentially two ways of approximating the $f(x)$
  - Kernerl regression
  - Polynomial Function
- Several ways to determin the order of polynomials.

Fuzzy RD: Use the Discontinuity as Instrument

- Fuzzy RD exploits discontinuities in the probability of treatment conditional on a covariate.
- The discontinuity becomes an instrumental variable for treatment status.
- D_i is no longer deterministically related to crossing a threshold but there is a jump in the *probability* of treatment at X_o.

$$P[D_i = 1|X_i] = \begin{cases} g_1(X_i) \text{ if } x_i \geq x_o \\ g_0(X_i) \text{ if } x_i < x_o \end{cases}, \text{ where } g_1(X_i) \neq g_0(X_i)$$

- $g_1(X_i)$ and $g_0(X_i)$ can be anything as long as they differ at x_0.
- The relationship between the probability of treatment and X_i can be written as:

$$P[D_i = 1|X_i] = g_0(X_i) + [g_1(X_i) - g_0(X_i)]T_i$$

where $T_i = 1(X_i \geq X_0)$

Figure 17:

Fuzzy RD with Varying Treatment Effects - Second Stage

- We can write down a first stage relationship:

$$E[D_i|X_i] = \gamma_{oo} + \gamma_{o1}X_i + \gamma_{o2}X_i^2 + \ldots + \gamma_{op}X_i^p \\ + \pi T_i + \gamma_1^* X_i T_i + \gamma_2^* X_i^2 T_i + \ldots + + \gamma_p^* X_i^p T_i$$

- One can therefore use both T_i as well as the interaction terms as instruments for D_i.
- If one uses only T_i as IV one has a just identified model which usually has good finite sample properties. In that case the estimated first stage would be:

$$D_i = \gamma_0 + \gamma_1 X_i + \gamma_2 X_i^2 + \ldots + \gamma_p X_i^p + \pi T_i + \xi_{1i}$$

- The fuzzy RD reduced form is:

$$Y_i = \mu + \kappa_1 X_i + \kappa_2 X_i^2 + \ldots + \kappa_p X_i^p + \rho\pi T_i + \xi_{2i}$$

Figure 18:

Fuzzy RD with Varying Treatment Effects - Second Stage

- As in the sharp RD case one can allow the smooth function to be different on both sides of the discontinuity.
- The second stage model with interaction terms would be the same as before:

$$\begin{aligned} Y_i = \alpha + \beta_{01}\widetilde{x}_i + \beta_{02}\widetilde{x}_i^2 + \ldots + \beta_{0p}\widetilde{x}_i^p \\ +\rho D_i + \beta_1^* D_i \widetilde{x}_i + \beta_2^* D_i \widetilde{x}_i^2 + \ldots + \beta_p^* D_i \widetilde{x}_i^p + \eta_i \end{aligned} \quad (2)$$

- Where $\widetilde{x}$ are now not only normalized with respect to x_o but are also fitted values obtained from the first stage regression.

#

Fuzzy RD with Varying Treatment Effects - Second Stage

Practical Tips for Estimation

- It is probably advisable to report results for both estimation types:
    - Polynomials in X.
    - Local linear regression.
- In robustness checks you also want to show that including higher order polynomials does not substantially affect your findings.
- Your results are not affected if you vary the window around the cutoff.
- Standard errors may go up but hopefully the point estimate does not change.

Graphical Analysis in RD Designs

A graphical analysis should be an integral part of any RD study. You should show the following graphs:

1. **Outcome by forcing variable** (X_i) :
    - The standard graph showing the discontinuity in the outcome variable.
    - Construct bins and average the outcome within bins on both sides of the cutoff.
    - You should look at different bin sizes when constructing these graphs (see Lee and Lemieux (2010) for details).
    - Plot the forcing variable X_i on the horizontal axis and the average of Y_i for each bin on the vertical axis.
    - You may also want to plot a relatively flexible regression line on top of the bin means.
    - Inspect whether there is a discontinuity at x_0.
    - Inspect whether there are other unexpected discontinuities.

Figure 20:

Graphical Analysis in RD Designs

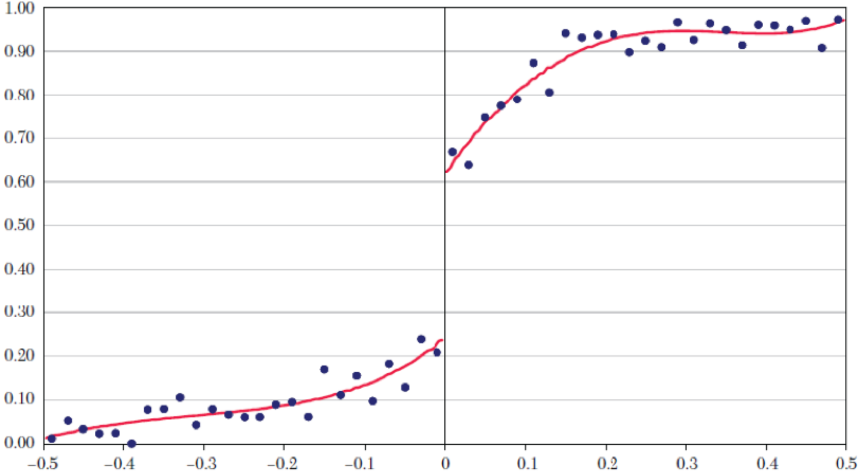


Figure 21:

Graphical Analysis in RD Designs

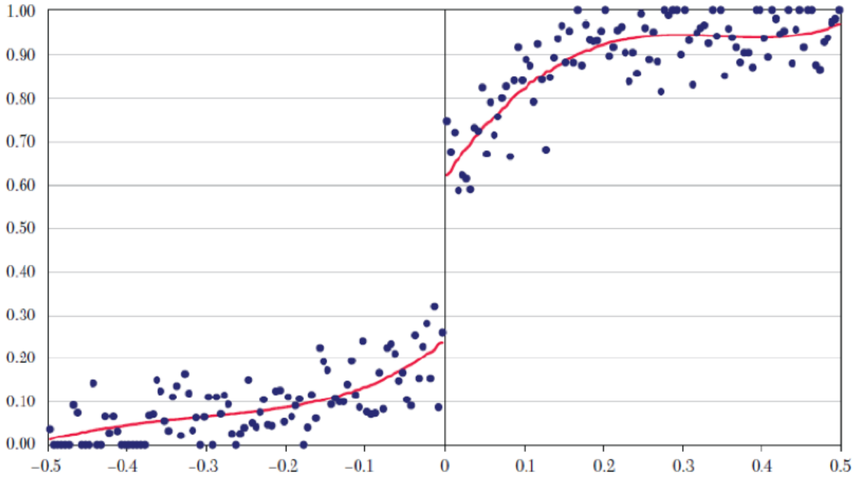


Figure 22:

Graphical Analysis in RD Designs

2. **Probability of treatment by forcing variable if fuzzy RD.**
   - In a fuzzy RD design you also want to see that the treatment variable jumps at x_0.
   - This tells you whether you have a first stage.
3. **Covariates by forcing variable.**
   - Construct a similar graph to the one before but using a covariate as the "outcome".
   - There should be no jump in other covariates.
   - If the covariates would jump at the discontinuity one would doubt the identifying assumption.

Figure 23:

Graphical:Example Covariates by Forcing Variable

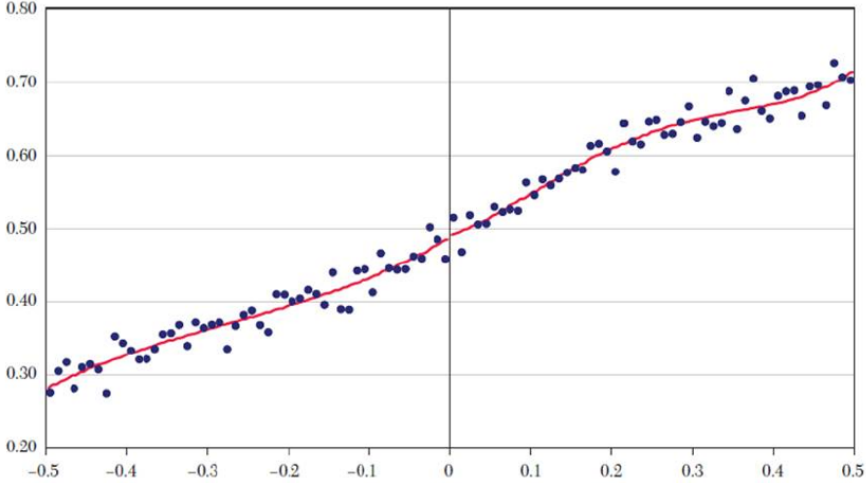


Figure 24:

Graphical:Density of the Forcing Variable

4. **The density of the forcing variable.**
    - One should plot the number of observations in each bin.
    - This plot allows to investigate whether there is a discontinuity in the distribution of the forcing variable at the threshold.
    - This would suggest that people can manipulate the forcing variable around the threshold.
    - This is an indirect test of the identifying assumption that each individual has imprecise control over the assignment variable.

Figure 25:

Graphical:Density of the Forcing Variable

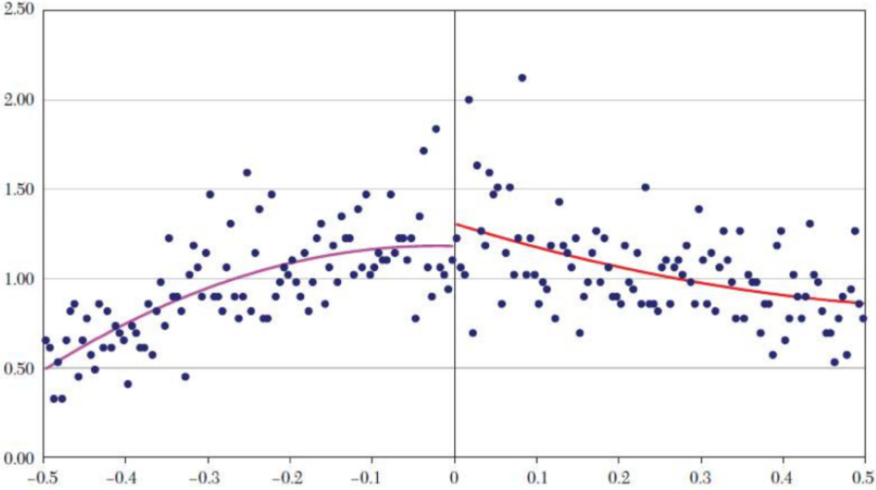


Figure 26:

Testing the Validity of the RD Design

- Testing the continuity of the density of X: McCrary(2008) test
- Test involving covariates:
  - Test whether other covariates exhibit a jump at the discontinuity. (Just re-estimate the RD model with the covariate as the dependent variable).This is a type of **placebo** test.
- Testing for jumps at non-discontinuity points

A Summary of Lectures

Causal Inference in Social Science

- **Causality and** is our main goal in the studies of empirical social science
- The existence of **Selection Bias** makes social science more difficult than science.
- Although experimental method is a powerful tool for economists, we can't carry on every project by it.
- It is the major reason *why modern econometrics exists and develops*

Selection on Observables and Unobservables

- **Conditional Independence Assumption(CIA)**

$$(Y_{0i}, Y_{1i}) \perp D_i \mid C_i$$

- CIA asserts that conditional on observable characteristics C_i, potential outcomes, Y_i are independent of assigned treatment, D_i.

– After controlling for value of covariates C_i, the assignment of units to treatment is *"as good as random"*.

– The probability of receiving treatment is same for the individuals with the same observable characteristics.

- Whether or not satisfies *CIA assumption* can divid methods into two categories
  - *selection on observables*
  - *selection on unobservables*

Causal Inference methods

- Furious Seven Weapons
  1. RCT （随机干预实验）
  2. OLS（普通最小二乘回归）
  3. Decomposition（分解）
  4. Matching and Propensity Score（匹配与倾向得分）
  5. Instrumental Variable（工具变量）
  6. Regression Discontinuity（断点回归）
  7. Difference in Differences （双差分）
- To make two groups comparable, we need to keep treatment and control group **"other thing equal"**in observed characteristics and unobserved characteristics.

An Simple Intuition: Mean Comparison

- Build a reasonable counterfactual state or find a proper control group is the core of econometrical methods.
- Common Idea: match similar units, and produce a mean comparison.
  - OLS gives conditional mean comparison.
  - By a counterfactual exercise,decomposition make a mean comparison between actual outcomes and predicted outcomes.
  - IV compares means of instrumented and non-instrumented(actually a weighted OLS)
  - DID compares difference in mean across groups which have a common trend.
  - RD compares means around the cutoff.
- Goal: give a believable and reliable mean comparison.

The path from cause to effect:Empricia

- **The path is never easy,but you are on the wheels!!!**