Financial Econometrics

Regression Discontinuity Design

Tim C.C. Hung 洪志清

June 6^{th}, 2022

Main Idea of Regression Discontinuity Design

- In an RCT, researchers can eliminate selection bias by **controlling treatment assignment process**.
  - An RCT randomizes who receives a treatment (treatment group) and who does not (control group).
  - Since we randomly assign treatment, the probability of getting treatment is unrelated to other confounding factors.
- But conducting an RCT is very expensive and may have ethical issue.
- 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 Regression Discontinuity Design

- Regression Discontinuity Design (RDD) exploits the facts that:
  - Some rules can generate a discontinuity in treatment assignment.
  - The treatment assignment is determined based on whether a unit exceeds some threshold on a variable.
  - Such variable is called **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.

Example: Healthcare and Health Condition

- Do healthcare services make people healthier?
  $\rightarrow$ This is an important (causal) question health economists care about.

- However, if we simply compare the health status of those who receive more healthcare service to the health of those who get less.

- We might find that people who receive more healthcare are sicker. :(
  $\rightarrow$ a spurious correlation due to the fact that sick people tend to (actively) seek for more healthcare services. $\Rightarrow$ **selection bias**!

Example: Healthcare and Health Condition

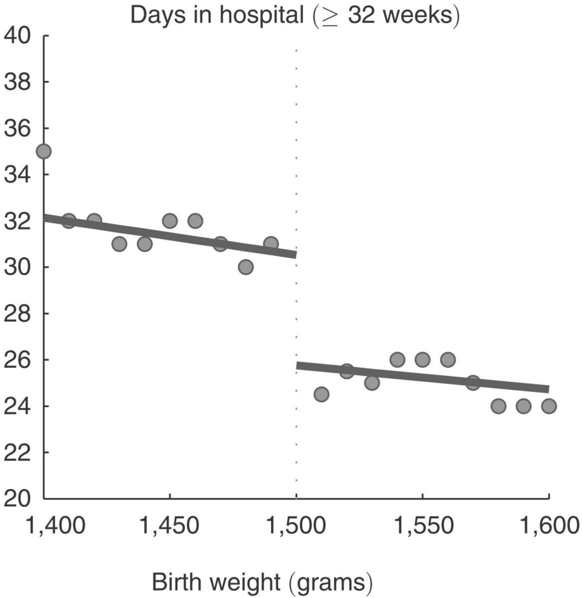
National Taiwan University

- A great way to answer that question would be to run an experiment:
  - Instead of letting people make healthcare decision, just flip a coin to decide whether they can receive healthcare.
  - Compare the health status or other outcomes of these two groups.
- Great idea, but nobody will let us run that experiment...

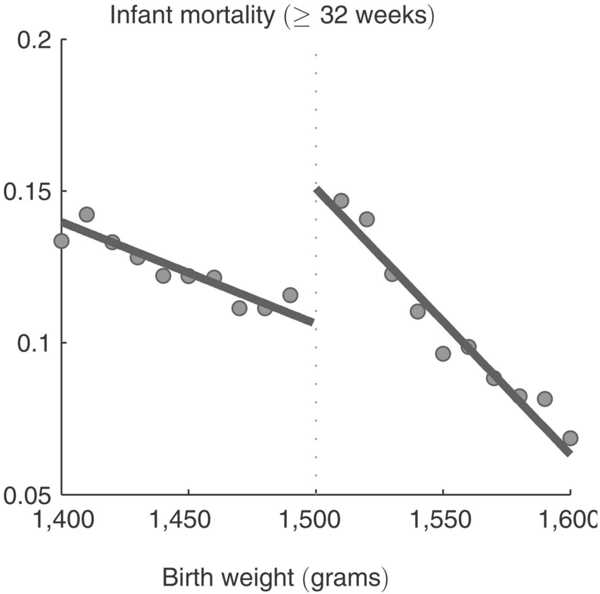
RDD is a local RCT

- Prashant Bharadwaj, Katrine Vellesen Løken, and Christopher Neilson (2013) "**Early Life Health Interventions and Academic Achievement**", *AER*
    - The effect of **health intervention** in early childhood on **later life outcomes**.
    - **Selection bias**: children 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.

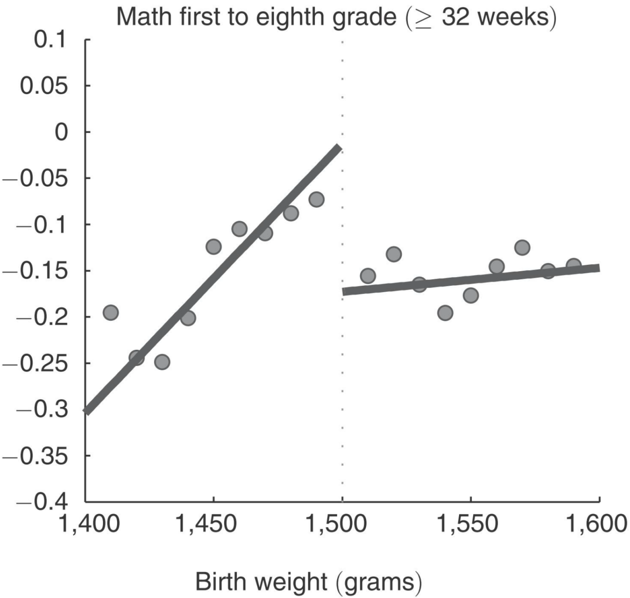
RDD is a local RCT



RDD is a local RCT



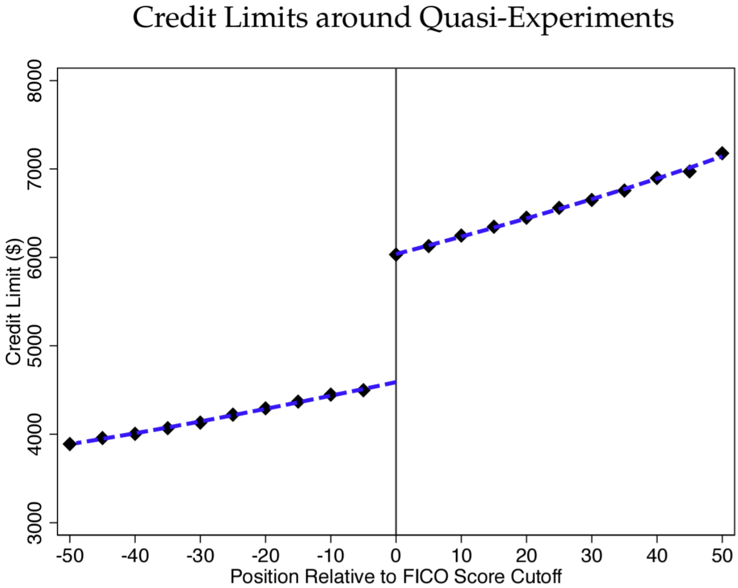
RDD is a local RCT



Finance Example

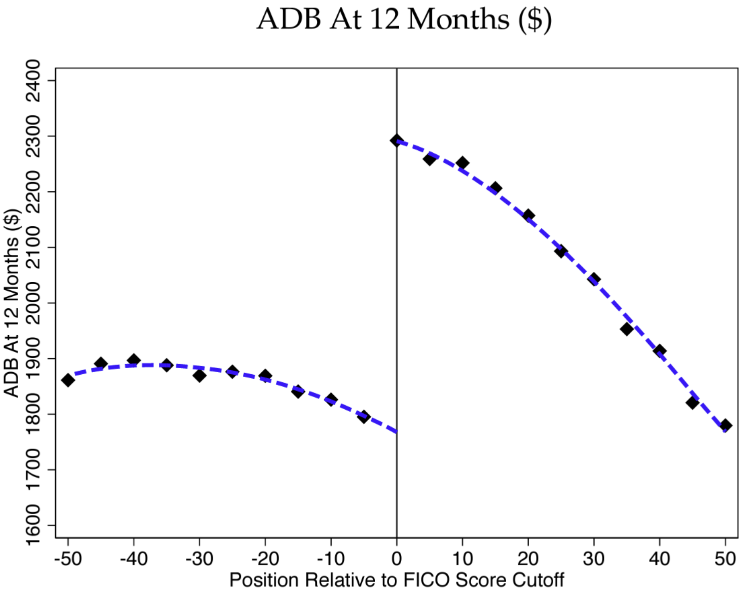
- Sumit Agarwal, Souphala Chomsisengphet, Neale Mahoney, Johannes Stroebel (2018) "**Do Banks Pass Through Credit Expansions to Consumers Who Want to Borrow?**", *QJE*
  - Does bank credit expansion stimulate more borrowing and consumption during the recession?
  - **Selection bias**: hard to disentangle demand (marginal propensity to borrow, MPB) and supple (marginal propensity to lend, MPL).
  - **RDD solution**: FICO scores!
    For example, a bank might grant a $2,000 credit limit to consumers with a FICO score below 720 and a $5,000 credit limit to consumers with a FICO score of 720 or above.

Finance Example

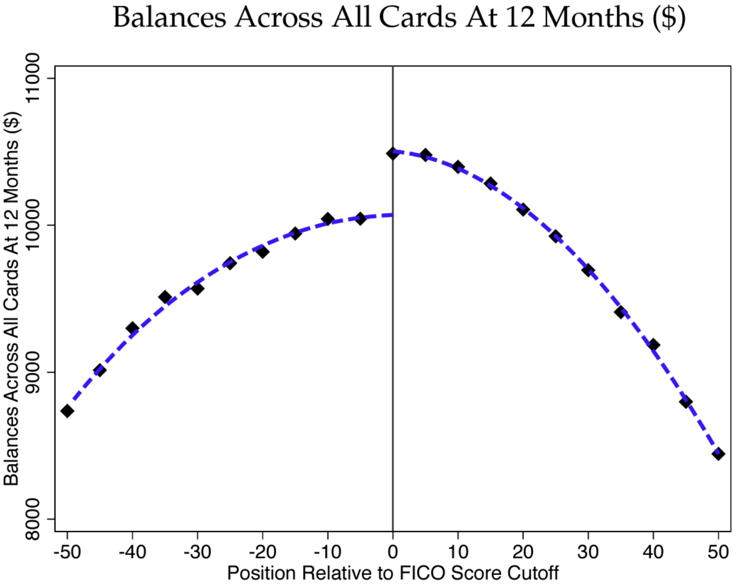


Finance Example

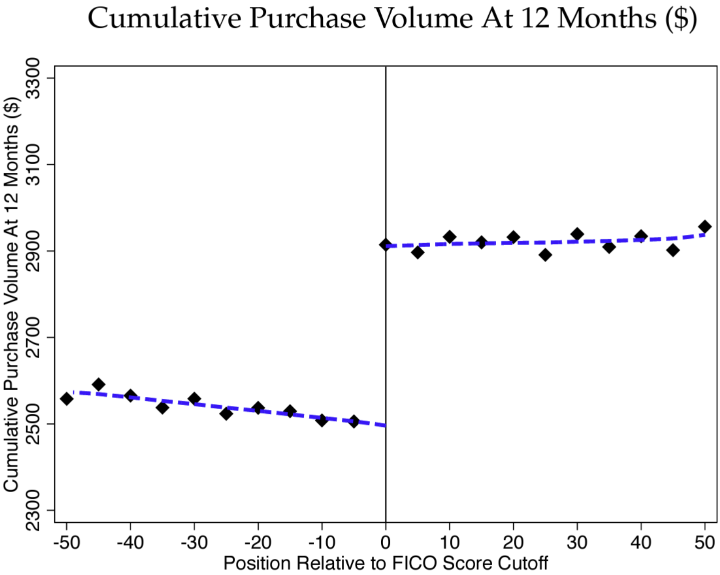
National Taiwan University



Finance Example



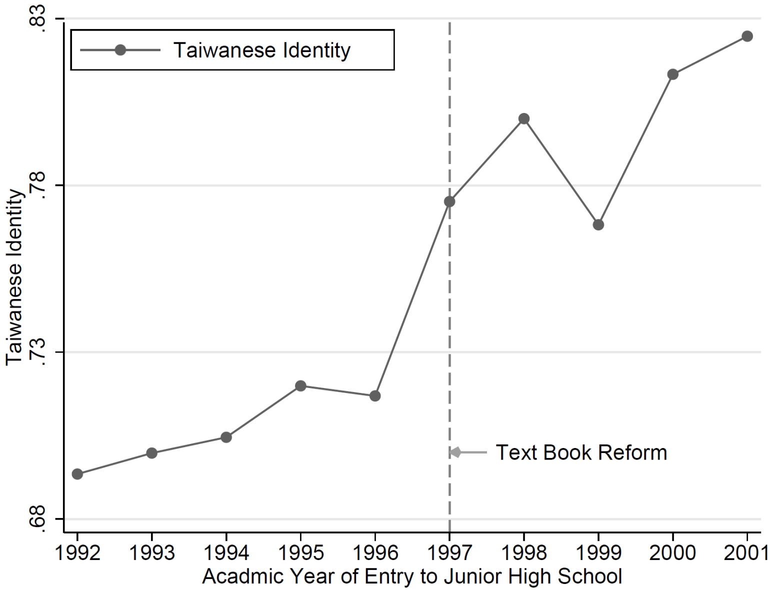
Finance Example



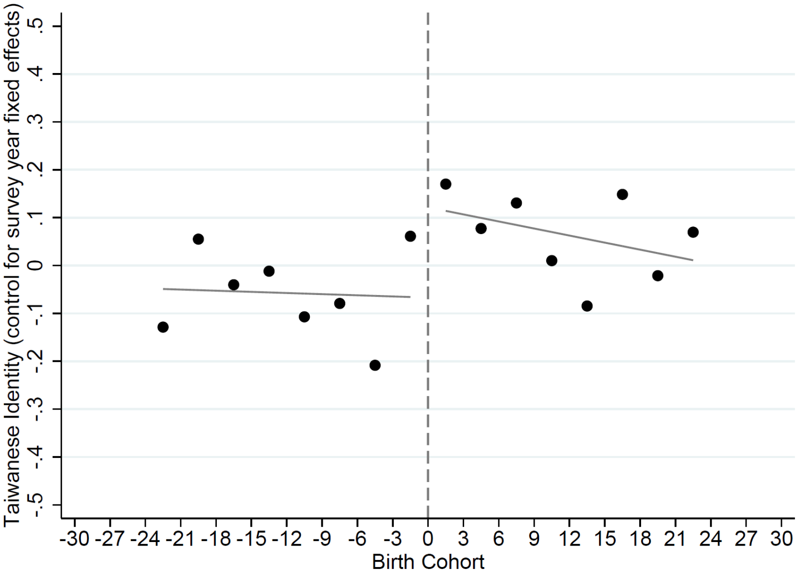
A Taiwanese Example

- Wei-Lin Chen (陳煒林), Ming-Jen Lin (林明仁) and Tzu-Ting Yang (楊子霆) (2022) "**Curriculum and National Identity: Evidence from the 1997 Textbook Reform in Taiwan**", R&R at *JDE*
  - The effect of curriculum for junior high school (中學教科書) on individuals' national identity (Taiwanese identity).
  - **Selection bias**: Government may change the content of textbook based on social trend.
  - **RDD solution**: In 1997, Taiwanese government implement a new curriculum (認識台灣) for the students who attend junior high school after September 1997.
  - Those born before September 1984 read old textbook, which exclusively focused on China. Those born after September 1984 read new textbook, which focused on Taiwanese history, geography, and society.

A Taiwanese Example



A Taiwanese Example



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 (IV) for actual treatment.

Sharp RDD

- Christopher Carpenter and Carlos Dobkin (2009) "**The Effect of Alcohol Consumption on Mortality: Regression Discontinuity Evidence from the Minimum Drinking Age**", *AEJ:Applied Economics*
  - This paper examine the effect of legal access to alcohol on mortality using a sharp RDD.
  - I will use this paper as an example to go through the key concept of sharp RDD.

Sharp RDD and Potential Outcomes

- Treatment:
  - Assignment variable (running variable): $X_i \in \mathbb{R}$
  - Threshold (cutoff) for treatment assignment: $c \in \mathbb{R}$
  - D_i: a dummy that indicate whether individual i receive treatment or not.
  - Treatment assignment:

$$D_i = \{X_i \geq c\}$$

$$D_i = \begin{cases} 1, & \text{if } X_i \geq c \\ 0, & \text{if } X_i < c \end{cases}$$

Sharp RDD and Potential Outcomes

- Potential Outcomes:
  - Y_i^1: Potential outcome for an individual i if he would receive treatment.
  - Y_i^0: Potential outcome for an individual i if he would NOT receive treatment.
- Observed Outcomes:
  - Observed outcomes Y_i are realized as:

$$Y_i = Y_i^1 D_i + Y_i^0 (1 - D_i)$$

$$Y_i = \begin{cases} Y_i^1, & \text{if } D_i = 1 \ (X_i \geq c) \\ Y_i^0, & \text{if } D_i = 0 \ (X_i < c) \end{cases}$$

Identification for Sharp RDD

- Ideally, for each individual i, if we could observe two potential outcomes at the same time, we can estimate average treatment effect (ATE):

$$\alpha_{ATE} = E[Y_i^1 - Y_i^0]$$

- But again, it is impossible to observe the two at the same time.

- Instead, we can use sharp RDD to investigate the behavior of the outcome around the threshold:

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

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

$$\alpha_{ATE\ at\ c} = E[Y_i^1 - Y_i^0 | X_i = c]$$

Identification for Sharp RDD

Assumption (Deterministic Assumption)

$$D_i = \mathbf{1}\{X_i \geq c\} \ \forall i$$

- Treatment assignment is a deterministic function of the assignment variable X_i and the threshold c.
  - An individual's age is above 21 $\rightarrow$ get legal access to alcohol.
  - An individual's age is below 21 $\rightarrow$ no legal access to alcohol.
- At the threshold c, we only see treated units and below the threshold $c - \varepsilon$, we only see non-treated units:

$$Pr(D_i = 1 | X_i = c) = 1$$
$$Pr(D_i = 1 | X_i = c - \varepsilon) = 0$$

Identification for Sharp RDD

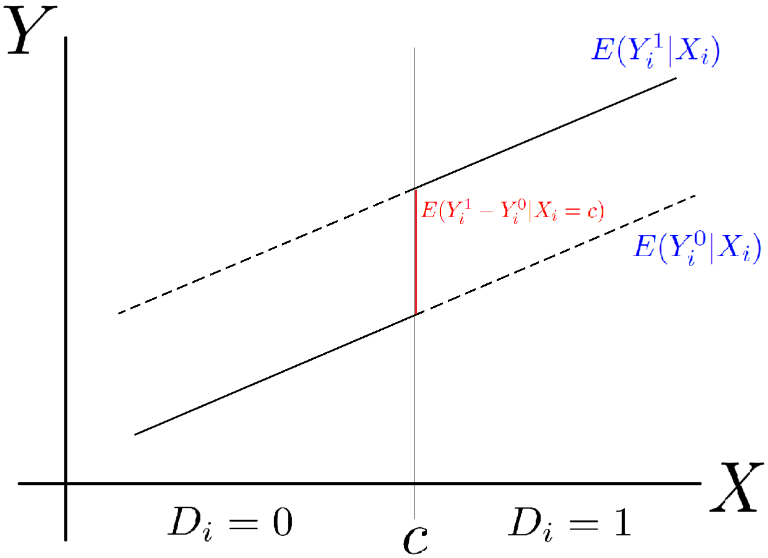
National Taiwan University

Assumption (Continuity Assumption)

$E[Y_i^1|X_i]$ and $E[Y_i^0|X_i]$ are continuous at $X_i = c$.

- Assume potential outcomes do NOT change at cutoff c.
  - 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.

Continuity Assumption



Continuity Assumption

- Remember we want to identify the ATE at the threshold:

$$\alpha_{ATE\ at\ c} = E[Y_i^1 - Y_i^0 | X_i = c]$$
$$= E[Y_i^1 | X_i = c] - E[Y_i^0 | X_i = c]$$

- But we do not observe $E[Y_i^0 | X_i = c]$ even due to the design, so we are going to extrapolate it from $E[Y_i^0 | X_i = c - \varepsilon]$

- In other words, we want to construct the counterfactual $E[Y_i^0 | X_i = c]$ using observed data:

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

Identification for Sharp RDD

- The continuity assumption and deterministic assumption imply the following:

$$\begin{aligned} E[Y_i^0|X_i = c] &= \lim_{\varepsilon \to 0} E[Y_i^0|X_i = c - \varepsilon] \\ &= \lim_{\varepsilon \to 0} E[Y_i^0|D_i = 0, X_i = c - \varepsilon] \\ &= \lim_{\varepsilon \to 0} E[Y_i|X_i = c - \varepsilon] \end{aligned}$$

- Note that this is the same for the treated group:

$$E[Y_i^1|X_i = c] = \lim_{\varepsilon \to 0} E[Y_i|X_i = c + \varepsilon]$$

- This allows us to use average outcomes of units just below the cutoff as a valid counterfactual for units right above the cutoff.

Identification for Sharp RDD

- The treatment effect is identified at the threshold as:

Theorem (Identification Results for Sharp RDD)

$$
\begin{aligned}
\alpha_{ATE\ at\ c} &= E[Y_i^1 - Y_i^0 | X_i = c] \\
&= E[Y_i^1 | X_i = c] - E[Y_i^0 | X_i = c] \\
&= \lim_{\varepsilon \to 0} E[Y_i | X_i = c + \varepsilon] - \lim_{\varepsilon \to 0} E[Y_i | X_i = c - \varepsilon] = \alpha_{SRD}
\end{aligned}
$$

- Under the above assumptions, we can identify **ATE at the threshold (unobserved)** using **sharp RD estimates (observed)**.

Sharp RDD Estimation

- To estimate the discontinuity at cutoff, we need to model the relationship between assignment variable X and outcome Y.

- Suppose that potential outcomes can be described by some reasonably smooth function $f(X_i)$:

$$E[Y_i^0|X_i] = \alpha + f(X_i)$$
$$Y_i^1 = Y_i^0 + \rho$$

- We can get RD estimates by fitting:

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

- In practice, people usually use a flexible polynomial (p^{th} order polynomial) regression to estimate $f(X_i)$:

$$Y_i = \alpha + \rho D_i + \beta_1 X_i + \beta_2 X_i^2 + \cdots + \beta_p X_i^p + \eta_i$$

Sharp RDD Estimation

- More general case:
  1. Allow X_i terms to differ on both sides of the threshold
     - ★ Include X_i both individually and interacting them with D_i.
     - ★ By doing this, we can estimate $f(X_i)$ on each side.
  2. Re-center X_i at c:
     - ★ $\tilde{X}_i = X_i - c$
     - ★ This step ensures that the treatment effect at $X_i = c$ is the coefficient on D_i in a regression model with interaction terms.
     - ★ So we do not have to add values of the D_i interacted with X_i to get the treatment effect at $X_i = c$.
  3. Selection polynomial orders:
     - ★ F-tests.
     - ★ AIC, BIC tests.

Sharp RDD Estimation

National Taiwan University

- Therefore, we usually estimate the following regression model:

$$
\begin{aligned}
Y_i = \alpha + \rho D_i + \beta_1 \tilde{X}_i + \cdots + \beta_p \tilde{X}_i^p \\
+ \beta_1^* D_i \tilde{X}_i + \cdots + \beta_p^* D_i \tilde{X}_i^p + \eta_i
\end{aligned}
$$

- Note that $\tilde{X}_i = X_i - c$.
- Equation on the previous pages is a special case where $\beta_1^* = \beta_2^* = \cdots = \beta_p^* = 0$
- The treatment effect (ATE) at $X = c$ is ρ.

Sharp RDD Estimation

- In the alcohol example, we estimate the following regression with cubic terms:

$$\begin{aligned} Y_i = \alpha + \rho D_i + \beta_1(X_i - 21) + \beta_2(X_i - 21)^2 + \beta_3(X_i - 21)^3 \\ + \beta_4 D_i(X_i - 21) + \beta_5 D_i(X_i - 21)^2 + \beta_6 D_i(X_i - 21)^3 + \eta_i \end{aligned}$$

- The effect of legal access to alcohol on mortality rate at age 21 is $\widehat{\rho}$

Sharp RDD Estimation

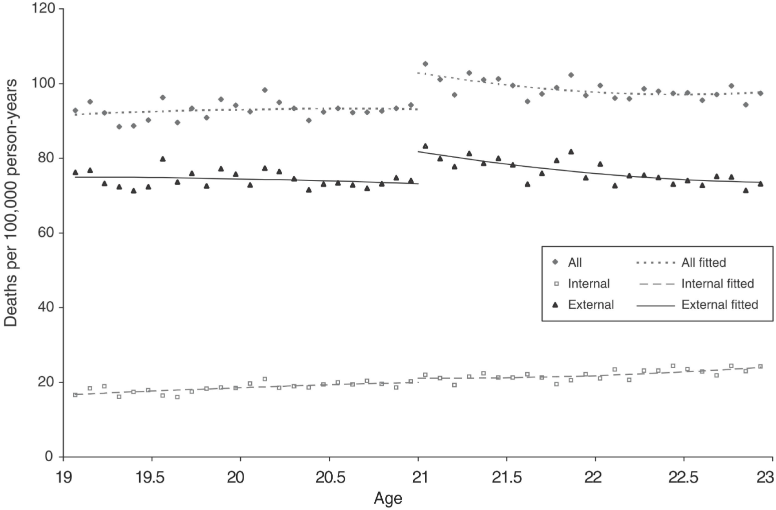
國立臺灣大學 National Taiwan University

TABLE 4—DISCONTINUITY IN LOG DEATHS AT AGE 21

	(1)	(2)	(3)	(4)
Deaths due to all causes				
Over 21	0.096	0.087	0.091	0.074
	(0.018)	(0.017)	(0.023)	(0.016)
Observations	1,460	1,460	1,460	1,458
R^2	0.04	0.05	0.05	
Prob > Chi-Squared		0.000	0.735	
Deaths due to external causes				
Over 21	0.110	0.100	0.096	0.082
	(0.022)	(0.021)	(0.028)	(0.021)
Observations	1,460	1,460	1,460	1,458
R^2	0.06	0.08	0.08	
Prob > Chi-Squared		0.000	0.788	
Deaths due to internal causes				
Over 21	0.063	0.054	0.094	0.066
	(0.040)	(0.040)	(0.053)	(0.031)
Observations	1,460	1,460	1,460	1,458
R^2	0.10	0.10	0.10	
Prob > Chi-Squared		0.000	0.525	
Covariates	N	Y	Y	N
Quadratic terms	Y	Y	Y	N
Cubic terms	N	N	Y	N
LLR	N	N	N	Y

Notes: See Notes from Table 1. The dependent variable is the log of the number of deaths that occurred x days from the person's twenty-first birthday. External deaths include all deaths with mention of an injury, alcohol use, or drug use. The Internal Death category includes all deaths Nt coded as external. Please see Web Appendix C for the ICD codes for each of the categories above. The first three columns give the estimates from polyNmial regressions on age interacted with a dummy for being over 21.

Sharp RDD Estimation



Sharp RDD Estimation: Local Method

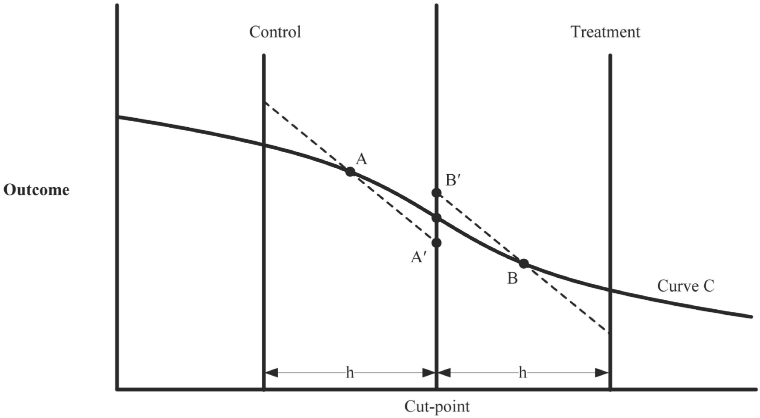
- Finding the right order of the polynomial may be troublesome.
- Instead, people use non-parametric/local approach which does NOT specify particular functional form of the outcome and the assignment variable.
- It uses only data within a small neighborhood (known as bandwidth) to estimate the discontinuity in outcomes **at the cutoff**:
    1. Compare means in the two bins adjacent to the cutoff (treatment v.s. control groups)
    2. Local linear regression

Sharp RDD Estimation: Local Method

國立臺灣大學 National Taiwan University

Figure 5

Boundary Bias from Comparison of Means vs. Local Linear Regression (Given Zero Treatment Effect)



Sharp RDD Estimation: Local Method

National Taiwan University

- Practically, we run local linear regression with a given window size h around the cutoff.

- Rather than $\overline{AB}$ (the average difference between the average of the two neighborhoods), the RD estimate is $\overline{B'A'}$

- The main challenge of non-parametric approach is to choose a bandwidth. There is essentially a trade-off between bias and precision.

- With larger bandwidth:

  pros more precise estimate because of larger number of the observations

  cons the estimate could be biased due to using data points far from cutoff

- People usually present RD estimates of different choices of bandwidth.

1. **Examine "Discontinuity" in Non-outcome Variables**
   - Construct a similar discontinuity graph for the covariates.
   - There should be **NO jump in other covariates**
   - If the covariates would jump at the discontinuity one would doubt the identifying assumption.

1. **Examine "Discontinuity" in Density of the assignment variable**
   - Individuals may invalidate the continuity assumption if they strategically manipulate assignment variable X to be just above or below the cutoff.
   - That is, people just above and below the cutoff are notcomparable.
   - This is related to a paper that will be presented: bunching estimator.

Course Surveys

National Taiwan University

Google Form: Survey Link