# Section: Module 6

## Regression Discontinuity Design

Kentaro Nakamura

GOV 2002

October 24th, 2025

# Logistics

- Congratulations on finishing the midterm!

- Two homework
  - Revenge Midterm (due October 27th 10am)
  - Problem set 6 (due October 29th 10am)

- You need to submit 3 problem sets out of 5
  - Revenge midterm is counted as one

# Today's Agenda

- Overview of Observational Studies

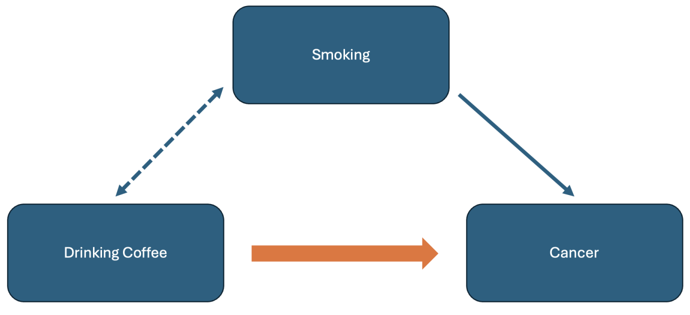
- Sharp RDD
  - Basic Setup / Intuition
  - Estimand / Assumption
  - Identification
  - Estimation

- Fuzzy RDD (PSet Question 2)

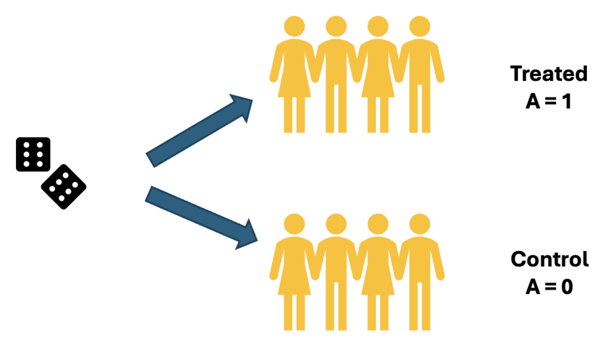
# Observational Studies: Overview

- From this week, we are in observational studies
  - Difference: Lack of randomization

- Example: Does drinking coffee cause cancer?



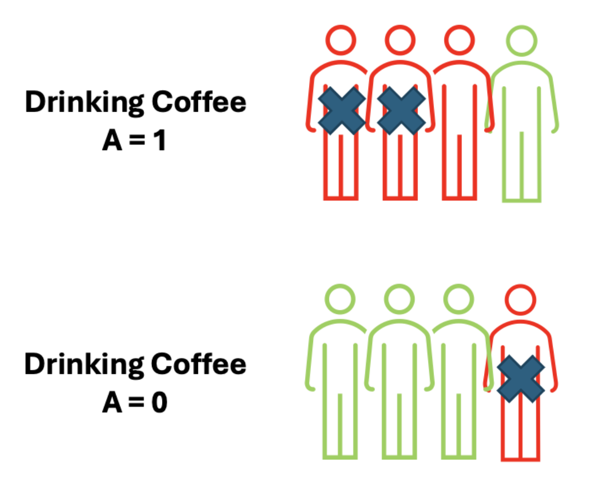
# Observational Studies: Overview




- Randomization makes two group comparable
  - Thus, difference-in-means works!

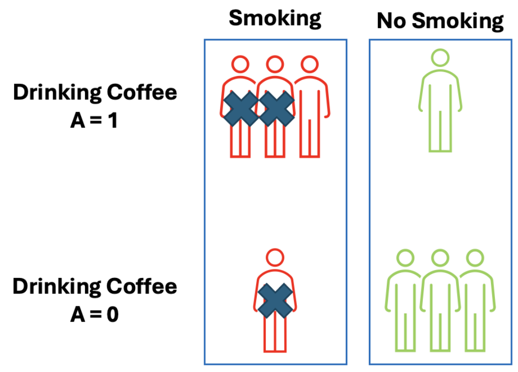
# Observational Studies: Overview

- It is not the case for observational studies



# Observational Studies: Overview

- Basic Approach: Control all the confounders
  - **Confounders**: Variables affecting both treatment and outcome




- Requirement: Observe all the confounders
  - Also, within each strata, there must exist counterparts of the comparison

# Observational Studies: Overview

- For confounders $C_i$, we formally need the followings:

$$\text{Ignorability / Exchangeability}: \quad Y_i(t) \perp\!\!\!\perp T_i \mid C_i$$
$$\text{Positivity / Common Support}: \quad 0 < \mathbb{P}(T_i = 1 \mid C_i) < 1$$

- We will cover these strategies in future modules
  - Module 8: Controlling confounders
    - Different ways of modeling (e.g., matching, weighting, regression)
  - Module 7: Sensitivity analysis + Partial identification
    - Checking credibility of assumptions

- BUT this is really hard! We rarely observe all the confounders

- **Quasi-experimental design**: Since we have no randomization, we will use several designs to estimate the causal effect
  - Module 5: Instrumental Variable
  - Module 6: Regression Discontinuity Design (this week!)
  - Module 10: Difference-in-Difference

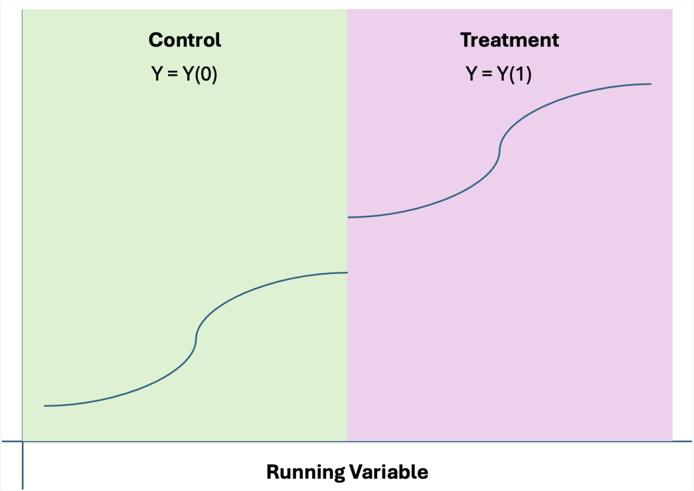
# Sharp RDD: Basic Setup and Intuition

- $T_i \in \{0, 1\}$: Treatment
- $X_i$: **Running variable** that perfectly determines the value of $T_i$ with the cutpoint $c$

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

- $X_i$ may be correlated with $Y_i(0)$ and $Y_i(1)$ (i.e., no selection of observable)
- Simply adjusting running variable $X_i$ does not work because of lack of overlap assumption
- **Intuition**: At the cutpoint $X_i = c$, assignment to treatment may be as if random
  - Only thing that differs is treatment assignment
- But no local randomization is necessary
  - See de la Cuesta and Imai (2016, ARPS)

# Sharp RDD: Basic setup



# Sharp RDD: Estimand and Assumption

- **Estimand**: Average treatment effect **on the threshold**

$$\tau = \mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = c]$$

  - Problem: External validity. **Local** ATE, applicable only to people at the threshold

- **Assumption**: $\mathbb{E}[Y_i(t) \mid X_i = x]$ is continuous in $x$ at $X_i = c$ for $t = 0, 1$
  - Continuity $\rightarrow$ Does not change abruptly
  - Formally, $\lim_{x \rightarrow c} \mathbb{E}[Y_i(t) \mid X_i = x] = \lim_{x \leftarrow c} \mathbb{E}[Y_i(t) \mid X_i = x]$
  - Example of violation (sorting): students strategically retaking an exam to just exceed a scholarship cutoff
  - Barely below and above the cutoff is no longer as-if random
    - Not the local randomization (randomization in the range)

# Sharp RDD: Identification

- Now, the estimand is $\tau = \mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = c]$
- Then, for $T_i = 1$

$$\begin{aligned}\mathbb{E}[Y_i(1) \mid X_i = c] &= \lim_{x \leftarrow c} \mathbb{E}[Y_i(1) \mid X_i = x] \quad (\because \text{continuity}) \\ &= \lim_{x \leftarrow c} \mathbb{E}[Y_i \mid X_i = x] \quad (\because \text{consistency})\end{aligned}$$

- Similariy, for $T_i = 0$

$$\mathbb{E}[Y_i(0) \mid X_i = c] = \lim_{x \rightarrow c} \mathbb{E}[Y_i(0) \mid X_i = x] = \lim_{x \rightarrow c} \mathbb{E}[Y_i \mid X_i = x]$$

- Therefore,

$$\tau = \underbrace{\lim_{x \downarrow c} \mathbb{E}[Y_i \mid X_i = x]}_{=\mathbb{E}[Y_i(1) \mid X_i = c]} - \underbrace{\lim_{x \uparrow c} \mathbb{E}[Y_i \mid X_i = x]}_{=\mathbb{E}[Y_i(0) \mid X_i = c]}$$

# Sharp RDD: Estimation

- Identification formula is

$$\tau = \underbrace{\lim_{x \downarrow c} \mathbb{E}[Y_i \mid X_i = x]}_{\text{Obtained from above threshold}} - \underbrace{\lim_{x \uparrow c} \mathbb{E}[Y_i \mid X_i = x]}_{\text{Obtained from below Threshold}}$$

- You need to estimate **the cutoff**
- Fit two **local linear regression**: for treated,

$$\arg\min_{\alpha,\beta} \sum_{i=1}^{n} \mathbf{1}\{X_i \geq c\}\{Y_i - \alpha - (X_i - c)\beta\}^2 K\left(\frac{X_i - c}{h}\right)$$

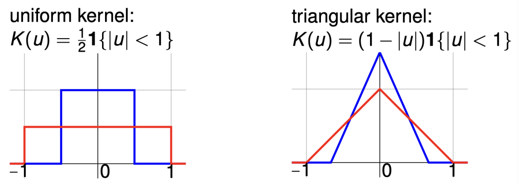
where

- $\mathbf{1}\{X_i \geq c\}$: indicator for treated unit (for control, $\mathbf{1}\{X_i \leq c\}$)
- $K\left(\frac{X_i - c}{h}\right)$ is the weight (kernel)
- Regressor is centered by $X_i - c$ so that $\alpha$ represnts the intercept at $X_i = c$ (i.e., cutoff)

# Local Linear Regression

$$\arg\min_{\alpha,\beta} \sum_{i=1}^{n} \mathbf{1}\{X_i \geq c\} \underbrace{\{Y_i - \alpha - (X_i - c)\beta\}^2}_{\text{Square of Errors}} \underbrace{K\left(\frac{X_i - c}{h}\right)}_{\text{Kernel}}$$

- **Kernel**: Gives more weight around cutoff
  - Recall that we want to model the local behavior around cutoff
  - The regression above is special version of weighted least squares




- **Bandwidth**: determines how local the regression is
  - Look at the value of running variable $X_i \pm h$
  - **Optimal bandwidth**: select bandwidth $h$ so that it minimizes MSE (Imbens and Kalyanaraman 2012)

# Review of Weighted Least Square

- **Weighted Least Square**: Weight the observation and solve OLS

- Formally, we minimize

$$\min_{\beta}(Y - X\beta)^{\top}W(Y - X\beta)$$

- Thus the first order condition is

$$\frac{\partial}{\partial\beta}(Y - X\beta)^{\top}W(Y - X\beta) = -2X^{\top}W(Y - X\beta) = 0$$
$$\rightarrow \hat{\beta} = (X^{\top}WX)^{-1}X^{\top}WY$$

# Local Linear Regression as a Weighted Regression

- Recall that local linear regression solves the minimization problem

$$\arg\min_{\alpha,\beta} \sum_{i=1}^{n} \mathbf{1}\{X_i \geq c\} \underbrace{\{Y_i - \alpha - \tilde{X}_i\beta\}^2}_{\text{Square of Errors}} \underbrace{K\left(\frac{\tilde{X}_i}{h}\right)}_{=W_i}$$

  where $\tilde{X}_i = X_i - c$

- Thus, the intercept $\hat{\alpha}$ is obtained by

$$\hat{\alpha}_+ = \boldsymbol{e}_1^\top (\boldsymbol{Z}^\top \boldsymbol{W} \boldsymbol{Z})^{-1} \boldsymbol{Z}^\top \boldsymbol{W} \boldsymbol{Y}$$

  where $\boldsymbol{e}_1 = \begin{bmatrix} 1 & 0 \end{bmatrix}^\top$, $\boldsymbol{Z}_i = \begin{bmatrix} 1 & \tilde{X}_i \end{bmatrix}^\top$, and $\boldsymbol{W} = \mathrm{diag}(W_i)$
  - You can see that $\hat{\alpha}_+$ is a linear combination of outcome

- Indeed, we can explicitly write $\hat{\alpha}_+$ with a weight $\tilde{W}_i$; i.e.,

$$\hat{\alpha}_+ = \sum_{i:\tilde{X}_i \geq 0} \underbrace{\frac{\tilde{W}_i}{\sum_{i:\tilde{X}_i \geq 0} \tilde{W}_i}}_{=\omega_i} Y_i \quad \text{where} \quad \tilde{W}_i = W_i \left(1 - \frac{\sum_{i:\tilde{X}_j \geq 0} W_j \tilde{X}_j}{\sum_{i:\tilde{X}_j \geq 0} W_j \tilde{X}_j^2} \tilde{X}_i\right)$$

# Why Local Linear Regression is good? (1)

- **Question**: Why local linear regression?
- **Answer**: Local linear regression behaves nicely at the boundary
  - When you estimate a regression near a boundary, bias can appear since there are fewer observations
  - Local linear regression corrects this bias up to the first order
- Let's see why it is the case.
- Recall that we derive the estimator $\hat{\alpha}_+$ is written as

$$\hat{\alpha}_+ = \sum_{i:\tilde{X}_i \geq 0} \underbrace{\frac{\tilde{W}_i}{\sum_{i:\tilde{X}_i \geq 0} \tilde{W}_i}}_{=\omega_i} Y_i$$

and thus

$$\mathbb{E}[\hat{\alpha}_+ \mid \tilde{\boldsymbol{X}}] = \sum_{i:\tilde{X}_i \geq 0} \omega_i \underbrace{\mathbb{E}[Y_i \mid \tilde{X}_i]}_{=\mu_1(\tilde{X}_i)}$$

# Why Local Linear Regression is good? (2)

- **Goal**: Show $\mathbb{E}[\hat{\alpha}_+ \mid \tilde{\boldsymbol{X}}] = \mathbb{E}[Y_i \mid \tilde{X}_i = 0] + \text{Small Bias}$
- Let's consider Taylor expansion of conditional expectation around $\tilde{X}_i = 0$

$$\mu_1(\tilde{X}_i) = \mu_1(0) + \mu_1'(0)\tilde{X}_i + \frac{1}{2}\mu_1''(0)\tilde{X}_i^2 + \cdots$$

  - The higher-order terms tends to be small

- Therefore,

$$\mathbb{E}[\hat{\alpha}_+ \mid \tilde{\boldsymbol{X}}] = \sum_{i:\tilde{X}_i \geq 0} \omega_i \underbrace{\mathbb{E}[Y_i \mid \tilde{X}_i]}_{=\mu_1(\tilde{X}_i)}$$

$$= \sum_{i:\tilde{X}_i \geq 0} \omega_i \left( \mu_1(0) + \mu_1'(0)\tilde{X}_i + \frac{1}{2}\mu_1''(0)\tilde{X}_i^2 + \cdots \right)$$

$$= \mu_1(0) \underbrace{\sum_{i:\tilde{X}_i \geq 0} \omega_i}_{=1} + \mu_1'(0) \sum_{i:\tilde{X}_i \geq 0} \omega_i \tilde{X}_i + \frac{1}{2}\mu_1''(0) \sum_{i:\tilde{X}_i \geq 0} \tilde{X}_i^2 + \cdots$$

# Why Local Linear Regression is good? (3)

- We can show that $\sum_{i:\tilde{X}_i \geq 0} \omega_i \tilde{X}_i$ is actually 0 (as local linear weight is 0)
  - Question 2 of PS6 (STAT)

- **Takeaway**: Bias at the boundary is zero up to first order

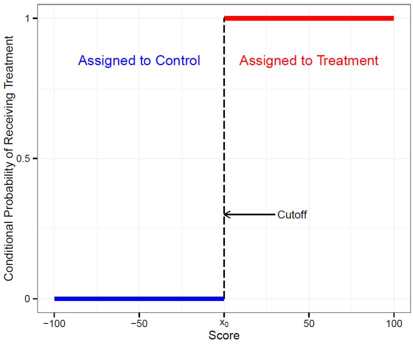
$$\mathbb{E}[\hat{\alpha}_+ \mid \tilde{\boldsymbol{X}}] = \mu_1(0) + \underbrace{\frac{1}{2}\mu_1^{''}(0) \sum_{i:\tilde{X}_i \geq 0} \tilde{X}_i^2 + \cdots}_{\text{Bias}}$$

  - If you use $k$th order local polynomial, bias is zero up to $k$-th order
  - However, as $k$ increases, the variance also increase

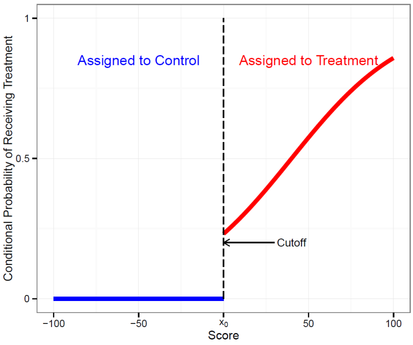
# Fuzzy RDD: Intuition

- **Fuzzy RDD**: Instrumental variable version of RDD
  - You have a running variable that does not perfectly determine the treatment status
  - e.g., eligibility of enrollment

Figure 1: Conditional Probability of Receiving Treatment in Sharp vs. Fuzzy RD Designs




(a) Sharp RD




(b) Fuzzy RD (One-Sided)

# Fuzzy RDD: Setup and Estimand

- Setup
  - $Z_i = \mathbf{1}\{X_i \geq c\}$: instrument
    - Instrument is determined by running variable
  - $T_i$: treatment
  - $Y_i$: outcome
  - Potential outcome: $Y_i(Z_i = z, T_i = t)$
- **Estimand**: LATE among complier at threshold

$$\mathbb{E}[Y_i(Z_i = 1, T_i = T_i(1)) - Y_i(Z_i = 0, T_i = T_i(0)) \mid X_i = c, \text{Complier}]$$

# Fuzzy RDD: Assumption and Identification

- **Assumptions**: IV + Sharp RDD
  - Monotonicity: $T_i(1) \geq T_i(0)$
  - Exclusion restriction: $Y_i(0, t) = Y_i(1, t)$
  - Continuity: $\mathbb{E}[T_i(z) \mid X_i = x]$ and $\mathbb{E}[Y_i(z, T_i(z)) \mid X_i = x]$ are continuous in $x$

- Recall that the identification formula of IV is given by

$$\frac{\text{Instrument Effect on Outcome (Z's effect on Y)}}{\text{Instrument Effect on Treatment (Z's effect on T)}}$$

- As Instrument's local effect is estimated by RDD, we use

$$\frac{\lim_{x \downarrow c} \mathbb{E}[Y_i \mid X_i = x] - \lim_{x \uparrow c} \mathbb{E}[Y_i \mid X_i = x]}{\lim_{x \downarrow c} \mathbb{E}[T_i \mid X_i = x] - \lim_{x \uparrow c} \mathbb{E}[T_i \mid X_i = x]}$$