Econometrics for Causal Inference

URP Part 4 : Fuzzy Regression Discontinuity

Sungkyunkwan University
-Machine Learning and Econometrics-

Jan 14, 2025

Introduction

- Reminder for Sharp RD
- Intuition for Fuzzy RD (imperfect compliance, IV ...)
- Numerical Experession of Fuzzy RD

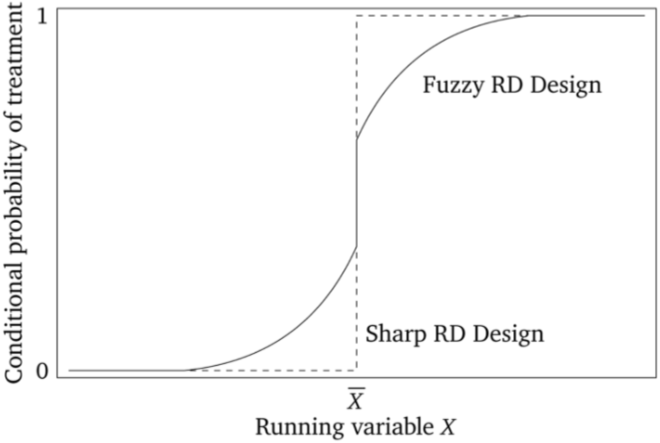
Reminder for Sharp RD

Regression Discontinuity (briefly)

- We can estimate **Local Average Treatment Effect (LATE)** via RD
- Score, Cutoff, Treatment was required to implement RD
- RD had two types based on "Change of Treatment Probability" (ignoring kink RD)
    1. Sharp RD : Probability increases from 0 to 1 in cutoff
    2. **Fuzzy RD** : Probability increases but difference is not 1

Intuition for Fuzzy RD

- Jump in treatment probability is smaller than 1



Intuition for Fuzzy RD

Fuzzy environment

- People can choose not to be treated although their score is higher than cutoff. (imperfect compliance)

  For example, in SKKU, students' whose GPA is highter than cutoff can major economics (treated), but not 100 %. In this case, probability of treatment increases in cutoff but jump is smaller than 1
- In order to get treatment, there may be additional needs other than X

  For example, in AirBnB, there exists 4 criteria to be AirBnB. Most important criterion is rating. Rating should be higher than 4.8 therefore X is rating. However, although host has rating higher than 4.8, if he does not satisfies other criteria he cannot be superhost. Therefore, probability to be superhost (treated) jumps in 4.8 but its size is smaller than 1.

Hullegie, P., & Klein, T. J. (2010).

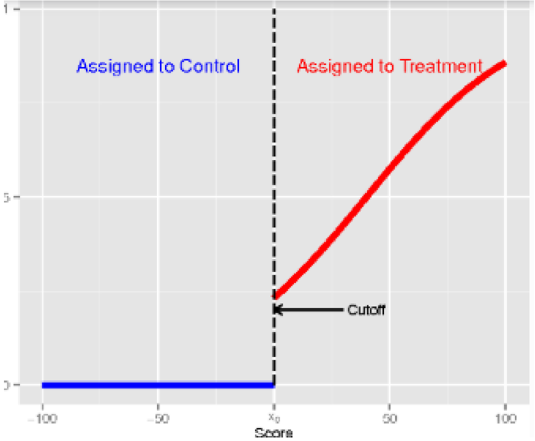
Effect of private insurance on individual's health

- Background
  In Germany, individual whose income is lower than cutoff cannot buy private insurance and they are mandated to take public insurance. For individual whose income is higher than cut off can opt out from public insurance and **may** buy private insurance
- Why fuzzy ? $\rightarrow$ *Imperfect Compliance*
  People can choose not to be treated although their score exceeded cutoff since they can stay in public insurance ($+$ measurement error)

Hullegie, P., & Klein, T. J. (2010).

Effect of private insurance on individual's health

X : income, Y : health status, T : private insurance



Fuzzy RD Designs

Unlike Sharp RD, there exists endogeneity problem directly regressing $Y = \beta_0 + \beta_1 T + \epsilon$ due to imperfect compliance

- There for we need to express Y and T in X

$$\tau_{FRD} = \frac{\lim_{x \downarrow c} \mathbb{E}[Y|X = x] - \lim_{x \uparrow c} \mathbb{E}[Y|X = x]}{\lim_{x \downarrow c} \mathbb{E}[T|X = x] - \lim_{x \uparrow c} \mathbb{E}[T|X = x]}$$

- Numerator: Discontinuity in outcome variable Y around the cutoff c.
- Denominator: Discontinuity in treatment variable T around the cutoff c.

Methodology

- Assume true treatment effect is ρ

$$E(Y_i) = \begin{cases} f_0(x_i) & \text{if } T_i = 0, \\ f_1(x_i) + \rho & \text{if } T_i = 1. \end{cases}$$

- Probability of treatment assignment:

$$P(T_i = 1 | x_i) = \begin{cases} g_0(x_i) & \text{if } x_i < 0, \\ g_1(x_i) & \text{if } x_i \geq 0. \end{cases}$$

Then,

$$E(Y_i|X_i) = f_0(x_i) + [f_1(x_i) - f_0(x_i) + \rho]T_i,$$
$$E(T_i|X_i) = g_0(x_i) + [g_1(x_i) - g_0(x_i)]Z_i, \quad Z_i = 1(x_i \geq 0).$$

Methodology

- Define Y_i in terms of the cutoff point $x_i = 0$:

$$Y_i = \begin{cases} f_0(x_i) + [f_1(x_i) - f_0(x_i) + \rho]g_0(x_i), & \text{if } x_i < 0, \\ f_0(x_i) + [f_1(x_i) - f_0(x_i) + \rho]g_1(x_i), & \text{if } x_i \geq 0. \end{cases}$$

$$\begin{aligned} \tau_{\text{FRD}} &= \frac{\lim_{x \downarrow 0} \mathbb{E}[Y_i | X_i = x] - \lim_{x \uparrow 0} \mathbb{E}[Y_i | X_i = x]}{\lim_{x \downarrow 0} \mathbb{E}[T_i | X_i = x] - \lim_{x \uparrow 0} \mathbb{E}[T_i | X_i = x]} \\ &= \frac{\rho[g_1(x_i) - g_0(x_i)]}{g_1(x_i) - g_0(x_i)} = \rho \end{aligned}$$

Fuzzy RD is IV

Instrument Variable (IV)

- IV is used to solve endogeneity problem ($E(X * \epsilon) \neq 0$)
  We wanna solve this using exogenous variable Z (Expressing X in reduced form)

$$Y = \beta_0 + \beta_1 X + \epsilon$$

$$X = \alpha_0 + \alpha_1 Z + \eta$$

- First stage : regress X on Z
- Second stage : regress Y on $\hat{X}$

Fuzzy RD is IV

Instrument Variable (IV)

- To Z be IV, it must satisfy two conditions

$$E(Z * \eta) = 0 \tag{1}$$

$$E(Z * \epsilon) = 0 \tag{2}$$

We can express equation (2) as E$(Z(Y - X\beta)) = 0$
Then,

$$\hat{\beta}_{IV} = \frac{E(ZY)}{E(ZX)} = \frac{\frac{1}{n}\Sigma Z_i Y_i}{\frac{1}{n}\Sigma Z_i X_i} = \frac{\hat{Cov}(Y_i, Z_i)}{\hat{Cov}(X_i, Z_i)}$$

Fuzzy RD is IV

ITT:

- Effects of IV Z on Y:

$$\text{ITT}_Y = E[Y \mid Z = 1] - E[Y \mid Z = 0]$$

$$= \lim_{X \to c^+} E[Y \mid X] - \lim_{X \to c^-} E[Y \mid X]$$

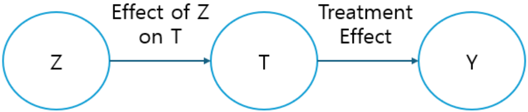
- Effects of IV Z on D:

$$\text{ITT}_T = E[T \mid Z = 1] - E[T \mid Z = 0]$$

$$\text{ITT}_T = \lim_{X \to c^+} E[T \mid X] - \lim_{X \to c^-} E[T \mid X]$$

where Z=1 if $X >= c$ and T is "real" treatment

Fuzzy RD is IV



(observed)Effect of Z on Y = (observed)Effect of Z on T * Treatment Effect

Treatment Effect = (observed)Effect of Z on Y / (observed)Effect of Z on T

Assume Z can affect Y only through T

$$
\begin{aligned}
\tau_{FRD} &= \frac{\text{Effect of Z on Y}}{\text{Effect of Z on T}} \\
&= \frac{Cov(Z_i, Y_i)/Var(Z_i)}{Cov(Z_i, T_i)/Var(Z_i)} = \frac{Cov(Z_i, Y_i)}{Cov(Z_i, T_i)} \\
&= \frac{E(ZY)}{E(ZT)} = \hat{\beta}_{IV}
\end{aligned}
$$

Further Readings

- Theoretical

  Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434), 444-455.

  Angrist, J., & Imbens, G. (1995). Identification and estimation of local average treatment effects.

Further Readings

- Empirical - Fuzzy RD

  Matsudaira, J. D. (2008). Mandatory summer school and student achievement. *Journal of Econometrics*, 142(2), 829-850.

  Bleemer, Z., & Mehta, A. (2022). Will studying economics make you rich? A regression discontinuity analysis of the returns to college major. *American Economic Journal: Applied Economics*, 14(2), 1-22.