# Public Policy Evaluation
# Lecture 4: Regression Discontinuity Designs

**Yao Thibaut Kpegli**

Licence 3 in Economics, ENS Paris-Saclay
2023–2024

March 2024

# Outline

# Context

- Unobserved heterogeneity: unobserved characteristics that explain both participation in the program and the outcome variable
- Threshold criteria: assignment to treatment is based on observable threshold/cutoff
- Assumption: individuals located on either side of the eligibility threshold for treatment, but close to it, are similar and can be compared

# Sharp design and Fuzzy design

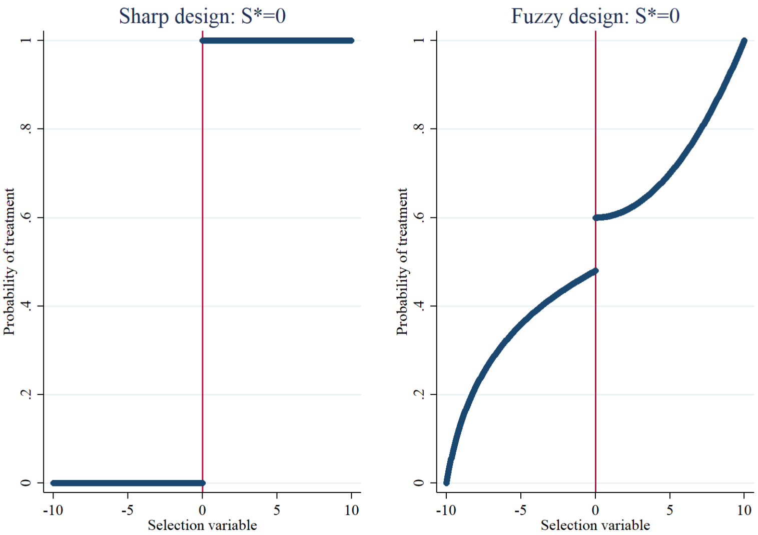
- Sharp design: strict application of the rule of cutoff so that the treatment depends deterministically on the value taken by a selection variable $S$

$$\begin{cases} T_i = 1 & \text{if} \qquad S_i \geq S^* \\ T_i = 0 & \text{if} \qquad S_i < S^* \end{cases}$$

- Fuzzy design: no strict application of the rule of cutoff but the probability of being treated should be higher when one is eligible than when one is not

$$P(T_i = 1 | S_i \geq S^*) > P(T_i = 1 | S_i < S^*)$$

# Sharp design and Fuzzy design



# Outline

# Notations

$$\begin{aligned}
\boldsymbol{Y}_i &= \underbrace{E[\boldsymbol{Y}_{0i}]}_{\alpha} + \underbrace{(\boldsymbol{Y}_{1i} - \boldsymbol{Y}_{0i})}_{\beta}\boldsymbol{T}_i + \underbrace{E[Y_{0i}] - \boldsymbol{Y}_{0i}}_{\epsilon_i} \\
&= \alpha + \beta\boldsymbol{T}_i + \epsilon_i
\end{aligned}$$

with $\beta$ the treatment effect

# Sharp design

- Independence assumption: individuals located on either side of the border, but very close to it, are strictly comparable:
  - in the absence of the treatment, there is no difference in the average value of $\boldsymbol{Y}_i$
  - other determinants of the outcome variable, modeled by the error term $\epsilon_i$, do not vary discontinuously around the cutoff $S^*$:

$$\lim_{S \longrightarrow S^*_-} E[\epsilon_i | S_i = S] = \lim_{S \longrightarrow S^*_+} E[\epsilon_i | S_i = S]$$

- Treatment effect: difference in mean between treated and untreated individuals around the cutoff

$$\beta = \lim_{S \longrightarrow S^*_+} E[Y_i | S_i = S] - \lim_{S \longrightarrow S^*_-} E[Y_i | S_i = S]$$

# Fuzzy design

- Independence assumption

$$\lim_{S \longrightarrow S^*_-} E[\epsilon_i|S_i = S] = \lim_{S \longrightarrow S^*_+} E[\epsilon_i|S_i = S]$$

- Difference in mean around the cutoff:

$$\begin{aligned} D &= \lim_{S \longrightarrow S^*_-} E[Y_i|S_i = S] - \lim_{S \longrightarrow S^*_+} E[Y_i|S_i = S] \\ &= \lim_{S \longrightarrow S^*_-} E[\beta T_i|S_i = S] - \lim_{S \longrightarrow S^*_+} E[\beta T_i|S_i = S] \\ &= \beta\Big(\lim_{S \longrightarrow S^*_-} E[T_i|S_i = S] - \lim_{S \longrightarrow S^*_+} E[T_i|S_i = S]\Big) \end{aligned}$$

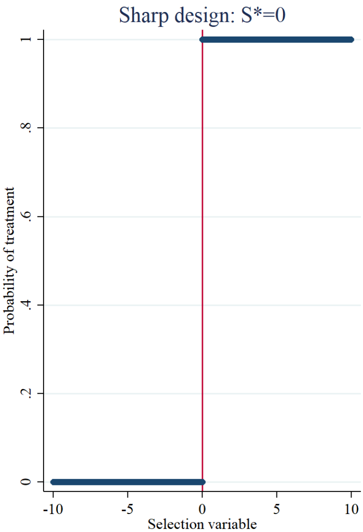
- Treatment effect:

$$\beta = \frac{\lim_{S \longrightarrow S^*_-} E[Y_i|S_i = S] - \lim_{S \longrightarrow S^*_+} E[Y_i|S_i = S]}{\lim_{S \longrightarrow S^*_-} E[T_i|S_i = S] - \lim_{S \longrightarrow S^*_+} E[T_i|S_i = S]}$$

# Outline

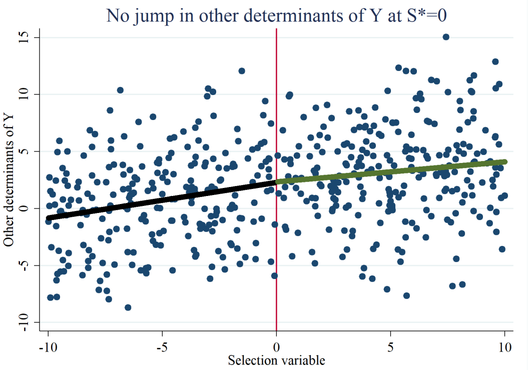
# RDD in regression

Verify that T exhibits a discontinuity at $S = S^*$



# RDD in pratice

- Verify that the outcome variable does not exhibit any discontinuity at $S = S^*$ in the absence of treatment (if you have past observations)
  - otherwise, just check that other determinants of outcome variable do not show any discontinuity at $S^*$



## Sharp design

- Estimate the two functions $f_-()$ and $f_+()$:

$$\boldsymbol{Y}_i = \begin{cases} f_-(\boldsymbol{S}_i) + \epsilon_i & \text{if} \quad \boldsymbol{S}_i < S^* \\ f_+(\boldsymbol{S}_i) + \epsilon_i & \text{if} \quad \boldsymbol{S}_i \geq S^* \end{cases}$$

- Estimate the effect:

$$\widehat{\beta} = \widehat{f}_+(S^*) - \widehat{f}_-(S^*)$$

- Parametric method (simplification)

$$\boldsymbol{Y}_i = f(\boldsymbol{S}_i) + \beta \boldsymbol{T}_i + \epsilon_i$$

with $f_-(\boldsymbol{S}_i) = f(\boldsymbol{S}_i)$, $f_+(\boldsymbol{S}_i) = f(\boldsymbol{S}_i) + \beta$, and $f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + ...$

# Fuzzy design

- Estimate the four functions $f_-()$, $f_+()$, $g_-()$ and $g_+()$:

$$\boldsymbol{Y}_i = \begin{cases} f_-(\boldsymbol{S}_i) + \epsilon_i & \text{if} \quad \boldsymbol{S}_i < S^* \\ f_+(\boldsymbol{S}_i) + \epsilon_i & \text{if} \quad \boldsymbol{S}_i \geq S^* \end{cases}$$

$$\boldsymbol{T}_i = \begin{cases} g_-(\boldsymbol{S}_i) + \epsilon_i & \text{if} \quad \boldsymbol{S}_i < S^* \\ g_+(\boldsymbol{S}_i) + \epsilon_i & \text{if} \quad \boldsymbol{S}_i \geq S^* \end{cases}$$

- Estimate the effect:

$$\widehat{\beta} = \frac{\widehat{f}_+(S^*) - \widehat{f}_-(S^*)}{\widehat{g}_+(S^*) - \widehat{g}_-(S^*)}$$

- Parametric method (simplification)

$$\boldsymbol{T}_i = f(\boldsymbol{S}_i) + \gamma \boldsymbol{S}_i + \epsilon_i$$

$$\boldsymbol{Y}_i = f(\boldsymbol{S}_i) + \beta \widehat{\boldsymbol{T}}_i + \epsilon_i$$

with $\widehat{\boldsymbol{T}}_i$ the predicted values from first regression

# Limits

Local Average Treatment Effect (LATE): estimate results are strictly valid only for individuals located at the cutoff

# Outline

# Application

Jens Ludwig and Douglas Miller "Does Head Start improve children's life chances? Evidence from a regression discontinuity design" (The Quarterly journal of economics, 2007).

Thank you for your attention !