Regression Discontinuity Design

Kosuke Imai

Harvard University

Spring 2021

Observational Studies

- In many cases, we cannot randomize the treatment assignment
  - ethical constraints
  - logistical constraints
- But, some important questions demand empirical evidence even though we cannot conduct randomized experiments!
- Designing observational studies $\rightsquigarrow$ find a setting where credible causal inference is possible
- Key $=$ Knowledge of treatment assignment mechanism
- Regression discontinuiety design (RD Design):
  1. Sharp RD Design: treatment assignment is based on a *deterministic* rule
  2. Fuzzy RD Design: encouragement to receive treatment is based on a deterministic rule
- Originates from a study of the effect of scholarships on students' career plans (Thistlethwaite and Campbell. 1960. *J. of Educ. Psychol*)

Regression Discontinuity Design

- Idea: Find an arbitrary cutpoint c which determines the treatment assignment such that $T_i = \mathbf{1}\{X_i \geq c\}$
- Close elections as RD design (Lee et al. 2004. *Q. J. Econ*):

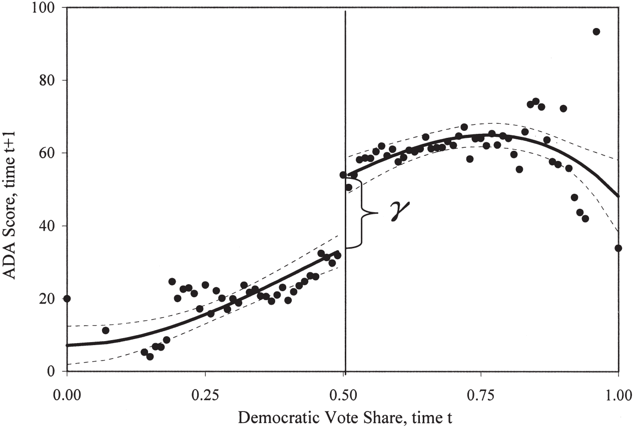
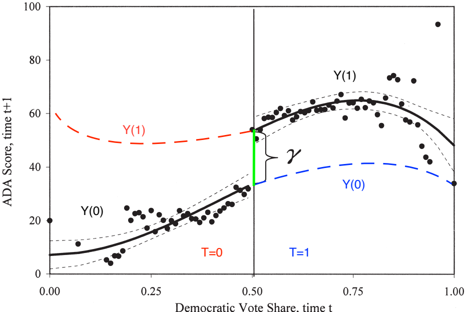


FIGURE I
Total Effect of Initial Win on Future ADA Scores: γ

Identification

- Estimand:
$$\mathbb{E}(Y_i(1) - Y_i(0) \mid X_i = c)$$
- Assumption: $\mathbb{E}(Y_i(t) \mid X_i = x)$ is continuous in x for $t = 0, 1$
  - nothing else that matters to potential outcomes is going on at $X_i = c$
  - deterministic rather than stochastic treatment assignment
  - violation of the overlap assumption: $0 < \Pr(T_i \mid X_i = x) < 1$ for all x
  - RD design is based on extrapolation
- Advantage: internal validity
- Disadvantage: external validity



Regression Modeling

- Two regressions:
  1. Treated group:

$$\mathbb{E}(Y_i(1) \mid X_i = c) = \lim_{x \downarrow c} \mathbb{E}(Y_i(1) \mid X_i = x) = \lim_{x \downarrow c} \mathbb{E}(Y_i \mid X_i = x)$$

  2. Control group:

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

- Simple linear regression within a window around the threshold c:

$$Y_i = \alpha + \beta(X_i - c) + \epsilon_i$$

  - Two separate regressions or a single regression with full interaction
  - How should we choose a window in a principled manner?
  - How should we relax the functional form assumption?

- Higher-order polynomial regression using the entire data $\rightsquigarrow$ sensitive to outliers and degree of polynomials (Imbens and Gelman. 2019. *J Bus Econ Stat*)

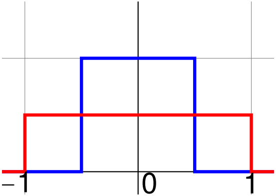
Local Linear Regression

- Better behavior at the boundary than other nonparametric regressions
- Two weighted regression with a kernel function and bandwidth h:

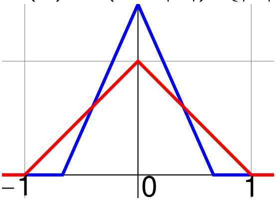
$$(\hat{\alpha}_+, \hat{\beta}_+) = \underset{\alpha,\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \mathbf{1}\{X_i > c\}\{Y_i - \alpha - (X_i - c)\beta\}^2 \cdot K\left(\frac{X_i - c}{h}\right)$$

$$(\hat{\alpha}_-, \hat{\beta}_-) = \underset{\alpha,\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \mathbf{1}\{X_i < c\}\{Y_i - \alpha - (X_i - c)\beta\}^2 \cdot K\left(\frac{X_i - c}{h}\right)$$

uniform kernel:
$K(u) = \frac{1}{2}\mathbf{1}\{|u| < 1\}$



triangular kernel:
$K(u) = (1 - |u|)\mathbf{1}\{|u| < 1\}$



Optimal Bandwidth (Imbens and Kalyanaraman. 2012. *Rev. Econ. Stud.*)

- Choose the bandwidth by minimizing the MSE:

$$
\begin{aligned}
\text{MSE} &= \mathbb{E}[\{\underbrace{(\hat{\alpha}_+ - \hat{\alpha}_-)}_{\text{estimated effect}} - \underbrace{(\alpha_+ - \alpha_-)}_{\text{true effect}}\}^2 \mid \mathbf{X}] \\
&= \underbrace{\mathbb{E}\{(\hat{\alpha}_+ - \alpha_+)^2 \mid \mathbf{X}\}}_{\text{MSE of } \hat{\alpha}_+} + \underbrace{\mathbb{E}\{(\hat{\alpha}_- - \alpha_-)^2 \mid \mathbf{X}\}}_{\text{MSE of } \hat{\alpha}_-} \\
&\quad -2 \cdot \underbrace{\mathbb{E}(\hat{\alpha}_+ - \alpha_+ \mid \mathbf{X})}_{\text{Bias}_+ = \text{Bias of } \hat{\alpha}_+} \cdot \underbrace{\mathbb{E}(\hat{\alpha}_- - \alpha_- \mid \mathbf{X})}_{\text{Bias}_- = \text{Bias of } \hat{\alpha}_-} \\
&= (\text{Bias}_+ - \text{Bias}_-)^2 + \text{Variance}_+ + \text{Variance}_-
\end{aligned}
$$

- Use the asymptotic approximation to bias and variance of local linear regression estimator at the boundary
- Refinements, e.g., bias correction (Calonico et al. 2014. *Econometrica*)