STATISTICAL HORIZONS

Regression Discontinuity Designs

Rocío Titiunik, Ph.D.

Upcoming Seminar:
December 6-7, 2019, Philadelphia, Pennsylvania

Overview

The RD Design: Definition and Taxonomy

- Basic setup
- Local Nature of Effects
- Graphical illustration of RD models

Causal Inference

- Main goal: learn about treatment effect of policy or intervention
- If treatment randomization available $\rightarrow$ easy to estimate effects
- If treatment randomization not available $\rightarrow$ observational studies
  - Selection on observables
  - Instrumental variables, etc.
- **Regression discontinuity (RD) design**
  - Simple assignment, based on known external factors
  - Objective basis to evaluate assumptions
  - *Careful*: very local!

Regression Discontinuity Design

Defined by the triplet: score, cutoff, treatment.

- Units receive a **score**.
- A treatment is assigned based on the score and a *known* **cutoff**.
- The **treatment** is:
  - given to units whose score is greater than the cutoff.
  - withheld from units whose score is less than the cutoff.
- Under assumptions, the abrupt change in the probability of treatment assignment allows us to learn about the effect of the treatment.

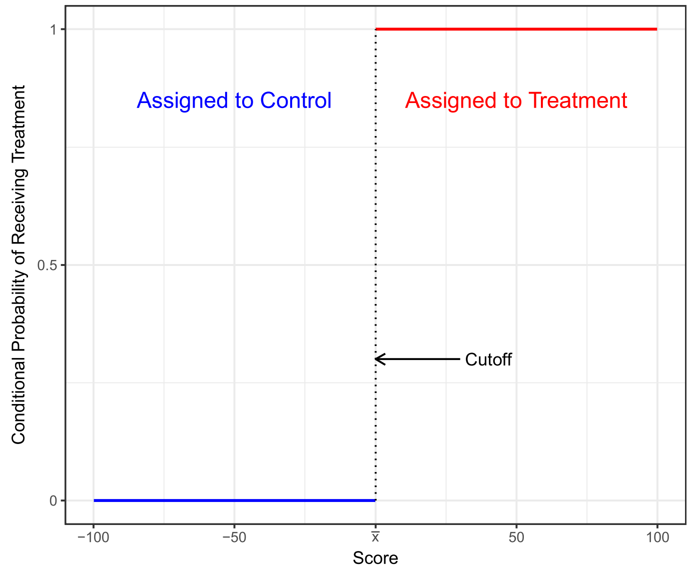
Regression Discontinuity Design

Defined by the triplet: score, cutoff, treatment.

- Units receive a **score**.
- A treatment is assigned based on the score and a *known* **cutoff**.
- The **treatment** is:
  - given to units whose score is greater than the cutoff.
  - withheld from units whose score is less than the cutoff.
- Under assumptions, the abrupt change in the probability of treatment assignment allows us to learn about the effect of the treatment.
- **Some examples**:

	X_i	Y_i
Education:	entry test score	test score, enrollment, performance, etc
Development:	pov index	educ, labor, health, etc
Health:	age / birthdate	insurance coverage, mortality, etc.

Treatment Assignment in (Sharp) RD Design



Sharp Regression Discontinuity Design

- n units, indexed by $i = 1, 2, \ldots, n$
- Unit's score is X_i, treatment is $T_i = \mathbf{1}(X_i \geq \bar{x})$
- Each unit has two potential outcomes:

  $Y_i(1)$: outcome that would be observed if i received treatment

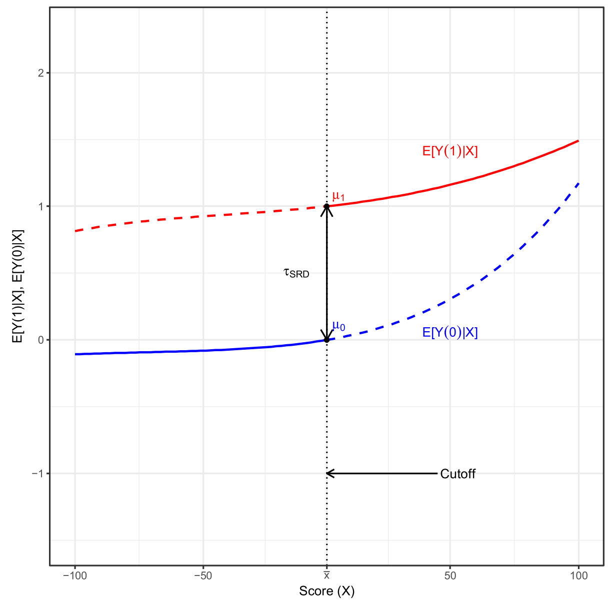
  $Y_i(0)$: outcome that would be observed if i received control

- The *observed* outcome is

$$Y_i = \begin{cases} Y_i(0) & \text{if } X_i < \bar{x}, \\ Y_i(1) & \text{if } X_i \geq \bar{x}. \end{cases}$$

- Fundamental problem of causal inference: only observe $Y_i(0)$ for units below cutoff and only observe $Y_i(1)$ for units above cutoff

RD Treatment Effect in Sharp RD Design



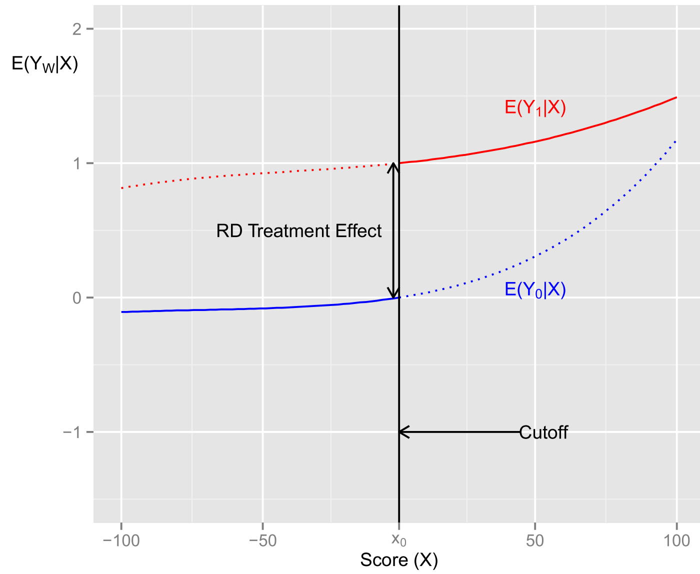
Fundamental Missing Data Problem

- A special situation occurs at the cutoff $X = \bar{x}$, the only point at which we may "almost" observe both curves
- Imagine two groups of units:

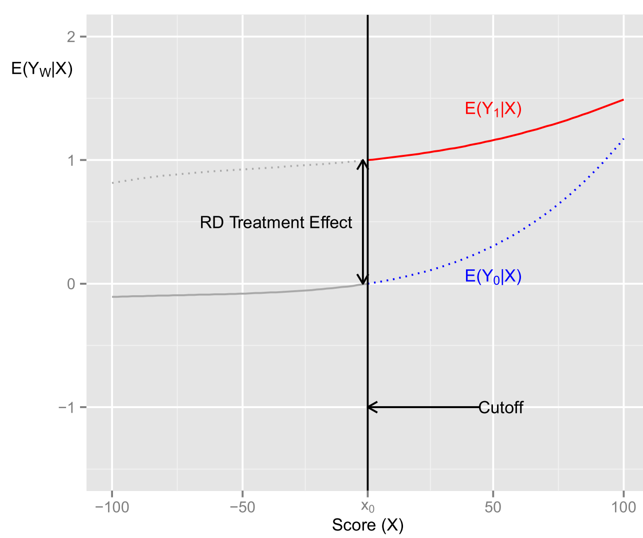
  with score equal to $\bar{x}$, $X_i = \bar{x} \rightarrow$ treated
  with with score barely below $\bar{x}$, $X = \bar{x} - \varepsilon \rightarrow$ control

- Yet *if values of the average potential outcomes at* $\bar{x}$ *are not abruptly different from their values at points near* $\bar{x}$, these two sets of units would be identical except for their treatment status
- Vertical distance at $\bar{x}$: the average treatment effect at this point
- This is the feature on which all RD designs are based

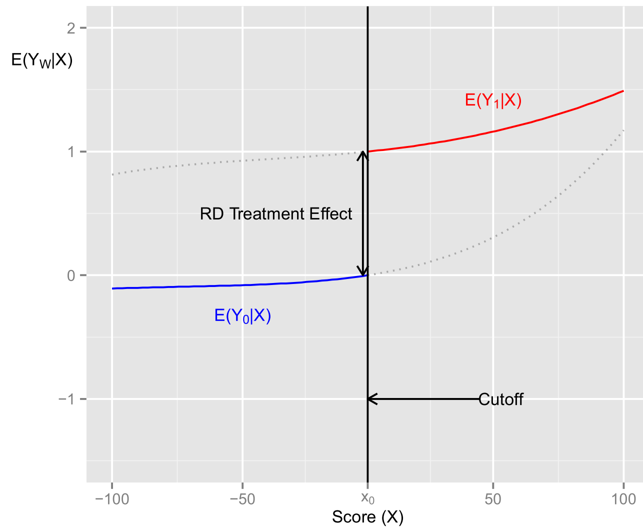
$$\tau_{\text{SRD}} = \mathbb{E}[Y_i(1) - Y_i(0)|X_i = \bar{x}]$$



$$\tau_{\text{SRD}} = \underbrace{\mathbb{E}[Y_i(1) - Y_i(0)|X_i = \bar{x}]}_{\text{Unobservable}} = \lim_{x \downarrow \bar{x}} \mathbb{E}[Y_i|X_i = x] - \lim_{x \uparrow \bar{x}} \mathbb{E}[Y_i|X_i = x]$$



$$\tau_{\texttt{SRD}} = \underbrace{\mathbb{E}[Y_i(1) - Y_i(0)|X_i = \bar{x}]}_{\text{Unobservable}} = \underbrace{\lim_{x \downarrow \bar{x}} \mathbb{E}[Y_i|X_i = x]}_{\text{Observable}} - \underbrace{\lim_{x \uparrow \bar{x}} \mathbb{E}[Y_i|X_i = x]}_{\text{Observable}}$$



Sharp RD design: Summary

- Canonical Parameter:

$$\tau_{\texttt{SRD}} = \mathbb{E}[Y_i(1) - Y_i(0)|X_i = \bar{x}] = \lim_{x \downarrow \bar{x}} \mathbb{E}[Y_i|X_i = x] - \lim_{x \uparrow \bar{x}} \mathbb{E}[Y_i|X_i = x]$$

- Perfect compliance:
  - every unit with score above $\bar{x}$ receives treatment.
  - every unit with score below $\bar{x}$ receives control.

- Not a “causal parameter” in the “proper” sense.

- Lee (2008) interpretation:

$$\tau_{\texttt{SRD}} = \int (y_1^+(w) - y_0^+(w)) \frac{f_{X|W}(\bar{x}|w)}{f_W(w)} dw$$

- Different interpretation under “local randomization”.

Example: Incumbency Advantage in U.S. Senate

- **Problem**: incumbency advantage in the U.S. Senate.
- Single-member district elections + two party system.
- Democratic party
  - runs for election t in state i and gets vote share X_i.
  - wins the election if vote share is 50% or more, $X_i \geq 50$.
  - loses the election if vote share is less than 50%.
- Outcome of interest: vote share in following election $t+1$, Y_i.
- Fundamental problem of causal inference: only observe Democratic's vote share at $t+1$ when the Democratic Party is incumbent in those districts where Democrats won election t.
- Cattaneo, Frandsen & Titiunik (2015, JCI).

Example: Incumbency Advantage in U.S. Senate

- **Problem**: incumbency advantage (U.S. senate).
- **Data**:

  $Y_i =$ election outcome at $t+1$.

  $T_i =$ whether party wins election at t .

  $X_i =$ margin of victory at $t \quad (\bar{x} = 0)$.

  $Z_i =$ covariates (*demvoteshlag1*, *demvoteshlag2*, *dopen*, etc.).
- **Potential outcomes**:

  $Y_i(0) =$ election outcome at $t+1$ if **had not been** incumbent.

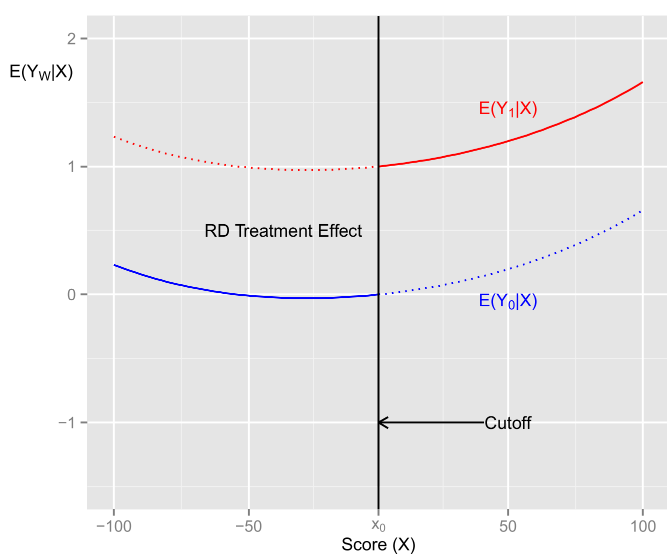
  $Y_i(1) =$ election outcome at $t+1$ if **had been** incumbent.
- **Causal Inference**:

  $$Y_i(0) \neq Y_i | T_i = 0 \qquad \text{and} \qquad Y_i(1) \neq Y_i | T_i = 1$$

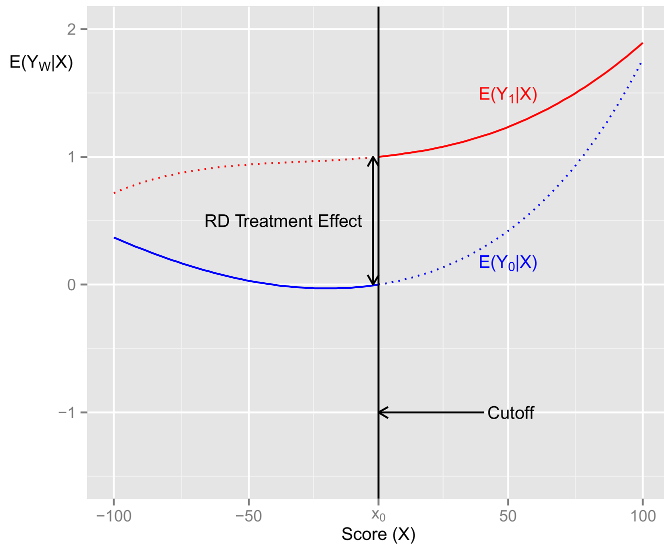
Local Nature of RD Effects

- RD parameters can be interpreted as causal in the sense that they are based on a comparison of potential outcomes ($Y_i(1)$ and $Y_i(0)$).
- But, in contrast to other parameters, average treatment effect is calculated at a single point on support of continuous random variable (X_i).
- This results in RD treatment effects having limited external validity:
  - $\tau_{\texttt{SRD}}$, the average treatment effect at $\bar{x}$, may not be informative about treatment effect at values of $x \neq \bar{x}$.
- Absent specific assumptions about global shape of regression functions, RD effects are average treatment effects *local to the cutoff*.
- How much can be learned from such local treatment effects will depend on each particular application.

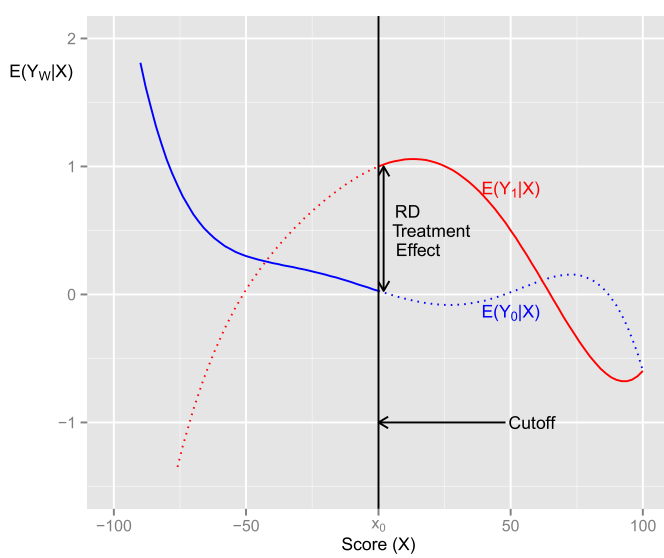
The RD Parameter: No Heterogeneity



The RD Parameter: Mild Heterogeneity



The RD Parameter: Wild Heterogeneity



RD Packages

`https://sites.google.com/site/rdpackages`

- **`rdrobust`**: estimation, inference and graphical presentation using local polynomials, partitioning, and spacings estimators; bandwidth selection.
- **`rdlocrand`** package: covariate balance, binomial tests, randomization inference methods (window selection & inference).
- **`rddensity`**: discontinuity in density test at cutoff (a.k.a. manipulation testing) using novel local polynomial density estimator.
- **`rdmulti`**: RD plots, estimation, inference, and extrapolation with multiple cutoffs and multiple scores.
- **`rdpower`** : power calculation and sample selection for local polynomial methods.

Graphical illustration of RD models

- Appealing feature of RDD: it can be illustrated graphically.
- Combined with formal approaches to estimation and inference, adds transparency to the analysis.
- Scatter plot: limited effectiveness for visualizing RD design.
- Usually useful to aggregate or smooth the data before plotting.

Graphical illustration of RD models

- Typical RD plot:
  - global polynomial fit
  - local sample means

(i) Global fit: smooth approximation to the unknown regression functions
  - 4th or 5th order polynomials, separately above and below the cutoff.

(ii) Local sample means:
  - disjoint intervals (bins) of the score, calculating the mean of the outcome within each bin.

- Combination of (i) and (ii) allows for:
  - visualize the overall shape of the regression functions for T and C
  - retain information about local behavior of the data

Graphical illustration of RD models

- Two types of bins:
  - Evenly-spaced
  - Quantile-spaced
- How to choose the number of bins optimally:
  - Tracing out the regression function: IMSE (balances bias and variance)
  - Mimicking Variance

Empirical Illustration:
Head Start (Ludwig and Miller, 2007,QJE)

- **Problem**: impact of Head Start on Infant Mortality
- **Data**:

  Y_i = child mortality 5 to 9 years old

  T_i = whether county received Head Start assistance

  X_i = 1960 poverty index $\quad (\bar{x} = 59.1984)$

  Z_i = see database.

- **Potential outcomes**:

  $Y_i(0)$ = child mortality if **had not received** Head Start

  $Y_i(1)$ = child mortality if **had received** Head Start

- **Causal Inference**:

  $Y_i(0) \neq Y_i | T_i = 0 \qquad \text{and} \qquad Y_i(1) \neq Y_i | T_i = 1$

- See Cattaneo, Titiunik and Vazquez-Bare (2017, JPAM) for details.

Effect of Head Start Assistance on Child Mortality