Short Course on Program Evaluation
Regression Discontinuity Designs

Matias D. Cattaneo
University of Michigan

May 28, 2017

RD Packages

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

- **rdrobust package**: estimation, inference and graphical presentation using local polynomials, partitioning, and spacings estimators.
- **rdlocrand package**: covariate balance, binomial tests, randomization inference methods (window selection & inference).
- **rddensity package**: discontinuity in density test at cutoff (a.k.a. manipulation testing) using local polynomial density estimator.
- **rdpower package**: power calculation and sample selection for local polynomial methods.
- New Practical Introduction Monograph and Replication files.

Outline

Recap: Causal Inference & Impact Evaluation

- Main goal: learn about treatment effect of policy or intervention.
- If treatment randomization available, easy to estimate treatment effects.
- If treatment randomization not available, turn to observational studies.
  - Instrumental variables.
  - Selection on observables.
- **Regression discontinuity (RD) designs**.
  - Simple and objective. Requires little information, if design available.
  - Might be viewed as a "local" randomized trial.
  - Easy to falsify, easy to interpret.
  - *Careful*: very local!

Randomized Control Trials

- **Notation**: $(Y_i(0), Y_i(1), X_i)$, $i = 1, 2, \dots, n$.
- **Treatment**: $T_i \in \{0, 1\}$, $\qquad T_i$ independent of $(Y_i(0), Y_i(1), X_i)$.
- **Data**: (Y_i, T_i, X_i), $i = 1, 2, \dots, n$, with

$$Y_i = \begin{cases} Y_i(0) & \text{if } T_i = 0 \\ Y_i(1) & \text{if } T_i = 1 \end{cases}$$

- **Average Treatment Effect**:

$$\tau_{\texttt{ATE}} = \mathbb{E}[Y_i(1) - Y_i(0)] = \mathbb{E}[Y_i | T = 1] - \mathbb{E}[Y_i | T = 0]$$

- **Experimental Design**.

Sharp RD design

- **Notation**: $(Y_i(0), Y_i(1), X_i)$, $i = 1, 2, \ldots, n$, $\qquad X_i$ continuous

- **Treatment**: $T_i \in \{0, 1\}$, $\qquad T_i = \mathbf{1}(X_i \geq \bar{x})$.

- **Data**: (Y_i, T_i, X_i), $i = 1, 2, \ldots, n$, with

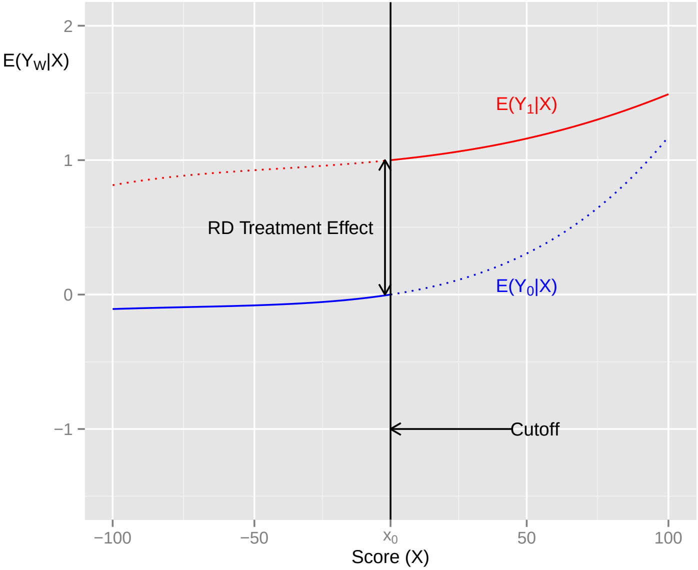
$$Y_i = \begin{cases} Y_i(0) & \text{if } T_i = 0 \\ Y_i(1) & \text{if } T_i = 1 \end{cases}$$

- **Average Treatment Effect at the cutoff**:

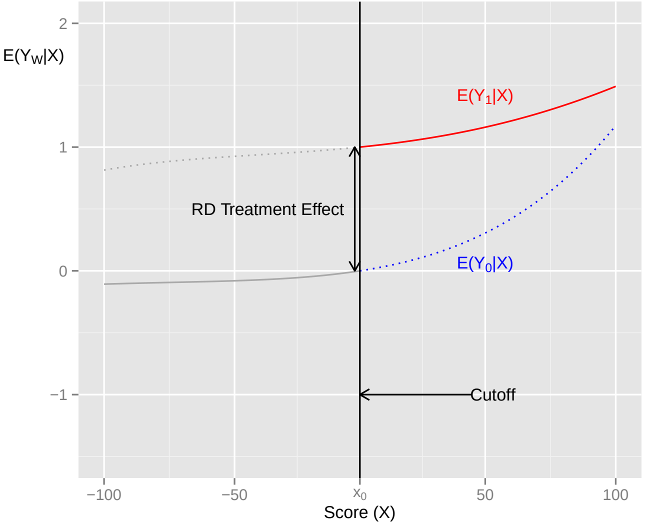
$$\tau_{\texttt{SRD}} = \mathbb{E}[Y_i(1) - Y_i(0) | X_i = \bar{x}] = \lim_{x \downarrow \bar{x}} \mathbb{E}[Y_i | X_i = x] - \lim_{x \uparrow \bar{x}} \mathbb{E}[Y_i | X_i = x]$$

- **Quasi-Experimental Design**: “local randomization” (more later)

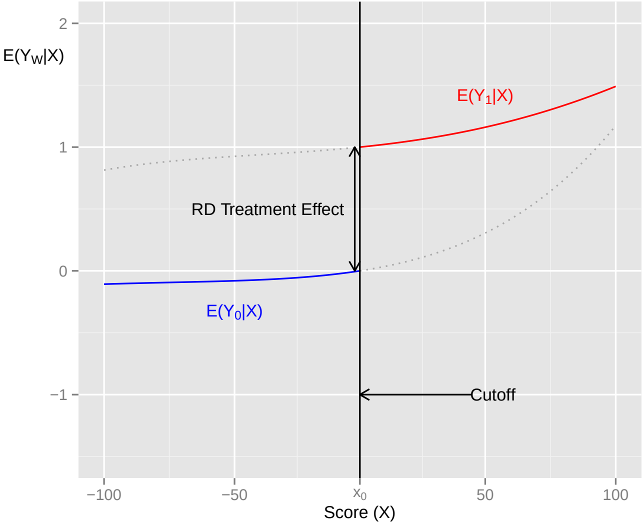
$$\tau_{\mathtt{SRD}} = \mathbb{E}[Y_i(1) - Y_i(0)|X_i = \bar{x}]$$



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



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



Empirical Illustration: 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).

Empirical Illustration: Incumbency Advantage (CFT, 2015, JCI)

- **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**:

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

Highlights and Main Ideas

- RD designs exploit "variation" near the cutoff.
- Graphical analysis is very useful: validation & falsification.
- Need to work with data near cutoff $\Longrightarrow$ bandwidth or window selection.
- Covariates and density of running variable should be similar near cutoff.
- Zero "overlap" so extrapolation is unavoidable (local or global).
- Causal effect is different (in general) than RCT.

Estimands and Identification

- Parameters of interests:
  - Sharp RD (SRD) and Fuzzy RD (FRD)
  - Kink RD (KRD) and Kink Fuzzy RD (KFRD)
  - Multiple scores RD and Geographic RD
  - Pooled RD v.s. Multiple Cutoff RD
- Inference methods roughly the same (some modifications required)
- Falsification methods more different in each case.

Sharp RD Design

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

- Canonical Parameter:

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

- 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".

Multi-cutoff RD Design

- Suppose cutoff $C_i \in \mathcal{C}$ with $\mathcal{C} = \{c_1, c_2, \cdots, c_J\}$ or $\mathcal{C} = [\underline{c}, \overline{c}]$.
  - General pooled RD designs (e.g., elections).
  - Geographic RD designs [Keele and Titiunik (2014)]
  - Multiple running variables RD designs.

- *Example*: Sharp RD design.
  - Pooled approach: $\tilde{X}_i = X_i - C_i$
  - Parameter:

$$\tau_{\texttt{SRD}} = \mathbb{E}[Y_i(1) - Y_i(0)|\tilde{X}_i = 0] = \lim_{x \downarrow 0} \mathbb{E}[Y_i|\tilde{X}_i = x] - \lim_{x \uparrow 0} \mathbb{E}[Y_i|\tilde{X}_i = x]$$
$$= \sum_{c \in \mathcal{C}} \mathbb{E}[Y_i(1,c) - Y_i(0,c)|X_i = c, C_i = c]\omega(c)$$

  where

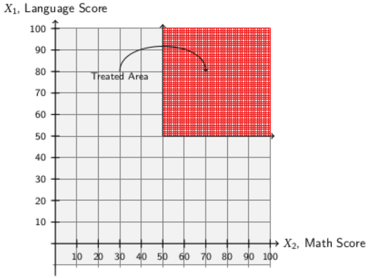
$$\omega(c) = \frac{f_{X|C}(c|c)\mathbb{P}[C = c]}{\sum_{c \in \mathcal{C}} f_{X|C}(c|c)\mathbb{P}[C = c]}$$

- Cattaneo, Keele, Titiunik and Vazquez-Bare (2016, JOP).

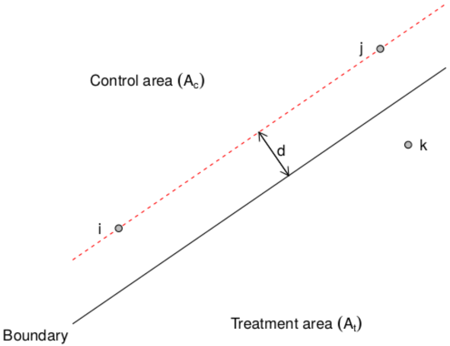
- Different interpretation under "local randomization"

Geographic and Multiple Running Variables RD Design

- Two scores: $X_i = (X_{1i}, X_{2,i})$ with "cutoff" $c = (c_1, c_2)$.
  - General case: two scores (e.g., math and language)
  - Specific case: geographic RD (Keele and Titiunik, 2014).
- Can be mapped back to RD with multiple-cutoff.
- Estimation and inference straightforward if $g(X_i) \in \mathbb{R}$, but this matters!
- Graphical and falsification methods differ a bit.
- Different interpretation under "local randomization"

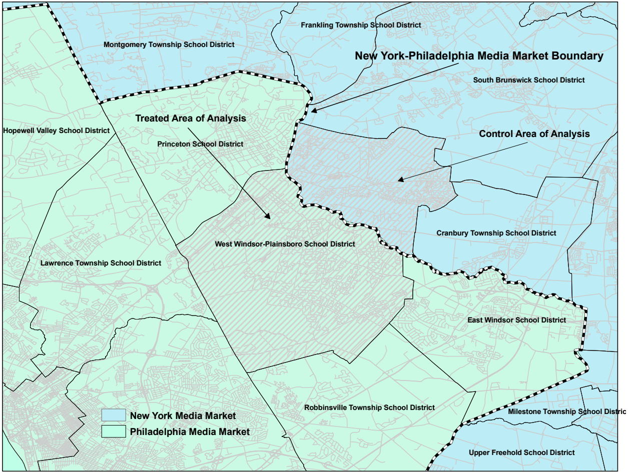


X_1, Language Score
100
90
80
70
60
50
40
30
20
10
Treated Area
10
20
30
40
50
60
70
80
90
100
X_2, Math Score



j
Control area (A_c)
d
k
i
Boundary
Treatment area (A_t)

Geo RD example: Effect of Advertisement of Voter Turnout



Local Randomization Approach

- **Idea**: near cutoff treatment assignment as-if randomly assigned.
- Employ RCT methods for "small" window around cutoff:
  - There exists window $W = [-h_n, h_n]$, with $-h_n < \bar{x} < h_n$, such that:

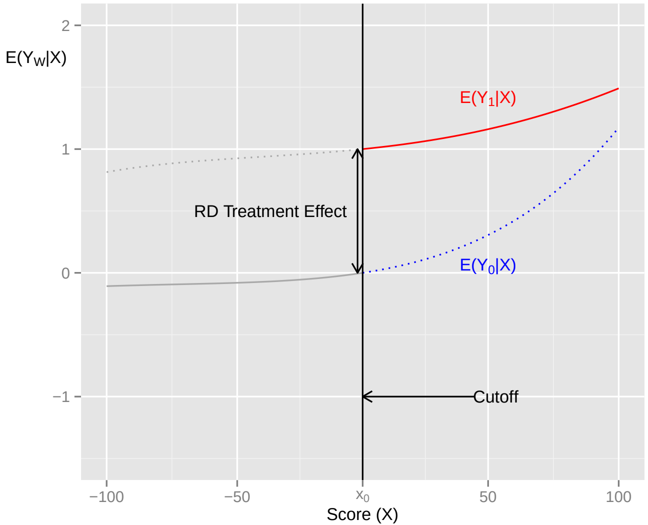
$$\text{for all } X_i \in W \Longrightarrow T_i \text{ independent of } (Y_i(0), Y_i(1))$$

  - Inference may use methods for randomized trials.
  - Example of test statistics: diff-in-means, regressions, etc.
  - As-if randomly assigned assumption can be relaxed somewhat, but it is strong.
- **Catch**: as-if random assumption may be good approximation *only very near cutoff*!
- Parameter is different, of course.
- Cattaneo, Frandsen, Titiunik (2015) & Cattaneo, Titiunik, Vazquez-Bare (2017, JPAM)

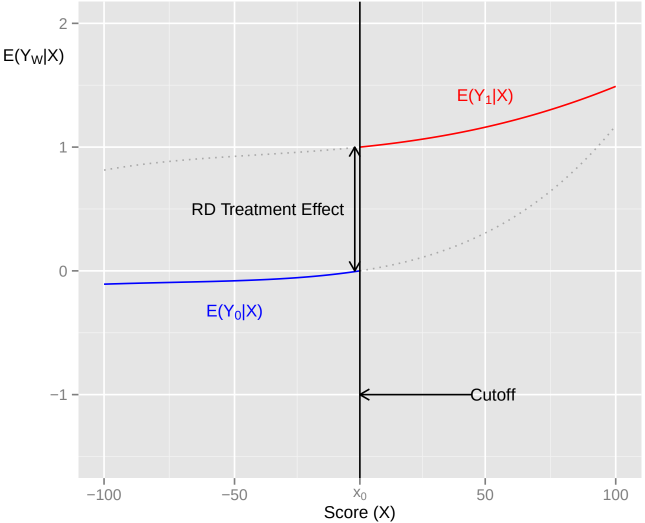
Estimation and Inference Methods

- Global polynomial approach (*not recommended*).
- Local polynomial approach.
  - Standard methods (useful for point estimation).
  - Robust bias-corrected methods (useful for inference).
- Local randomization approach and finite-sample inference.
  - Useful for both point estimation and inference.
  - Can be used for power calculations.
- Other methods exist (empirical likelihood, Bayesian, etc.).

$$\tau_{\mathtt{SRD}} = \underbrace{\mathbb{E}[Y_{1i} - Y_{0i}|X_i = x_0]}_{\text{Unobservable}} = \lim_{x \downarrow x_0} \mathbb{E}[Y_i|X_i = x] - \lim_{x \uparrow x_0} \mathbb{E}[Y_i|X_i = x]$$



$$\tau_{\texttt{SRD}} = \underbrace{\mathbb{E}[Y_{1i} - Y_{0i}|X_i = x_0]}_{\text{Unobservable}} = \underbrace{\lim_{x \downarrow x_0} \mathbb{E}[Y_i|X_i = x]}_{\text{Observable}} - \underbrace{\lim_{x \uparrow x_0} \mathbb{E}[Y_i|X_i = x]}_{\text{Observable}}$$



Estimation and Inference Methods

- Local polynomial methods.
  - Standard Approaches.
  - Bandwidth selection.
  - Robust Bias-correction.
  - Confidence intervals.
- Local randomization methods.
  - Interpreting RD as a local randomization in a window around the cutoff.
  - Conceptual differences with local polynomial estimation.
  - Window selection.
  - Estimation and inference using randomization-based methods.

Global Polynomial Approach (not recommended!)

- **Idea**: approximate regression functions for control and treatment units *globally*.

- **Naive approximation**: partially linear model,

$$Y_i = \alpha + \tau_{\mathtt{SRD}} \cdot T_i + (X_i - \bar{x}) \cdot \beta_1 + (X_i - \bar{x})^2 \cdot \beta_2 + \cdots + (X_i - \bar{x})^p \cdot \beta_p + \varepsilon_i$$

  - Bad idea: assumes constant treatment effect and equal slopes.

- **Flexible approximation**: full interaction model,

$$\begin{aligned} Y_i = \alpha + \tau_{\mathtt{SRD}} \cdot T_i &+ (X_i - \bar{x}) \cdot \beta_1 + (X_i - \bar{x})^2 \cdot \beta_2 + \cdots + (X_i - \bar{x})^p \cdot \beta_p \\ &+ T_i \cdot (X_i - \bar{x}) \cdot \gamma_1 + T_i \cdot (X_i - \bar{x})^2 \cdot \gamma_2 + \cdots + T_i \cdot (X_i - \bar{x})^p \cdot \gamma_p + \varepsilon_i \end{aligned}$$

  - Still bad idea: uses global approximation wit weights on observations are weird!

- Inference is standard; just a linear model.

Figure: Danger in Global Approximations

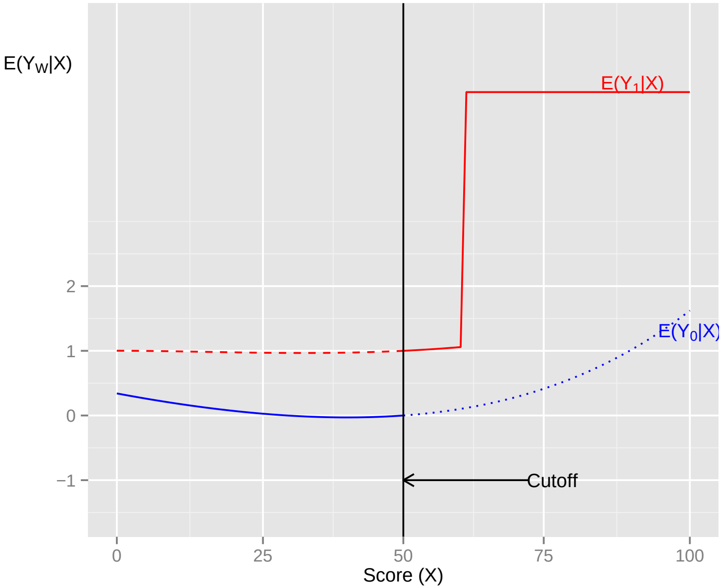


Figure: Danger in Global Approximations

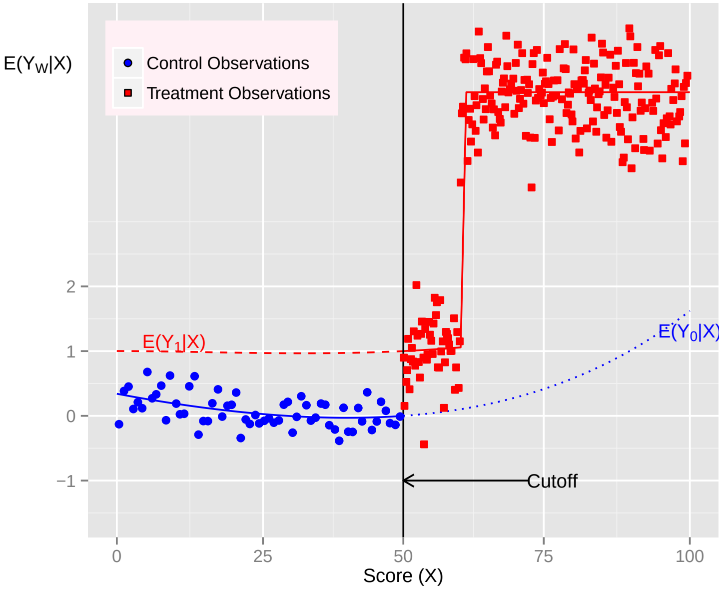


Figure: Danger in Global Approximations

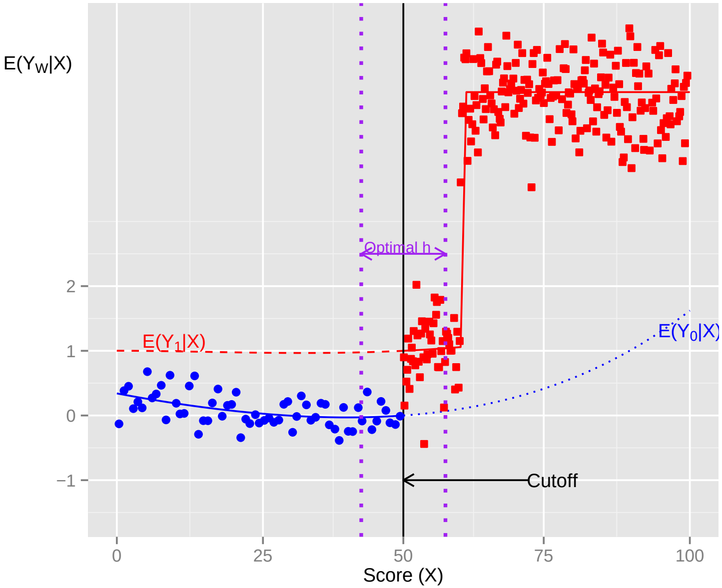


Figure: Danger in Global Approximations

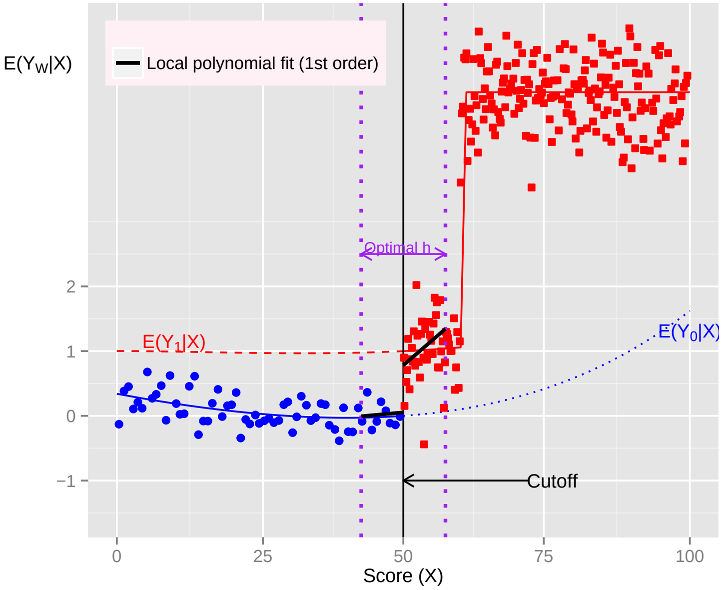
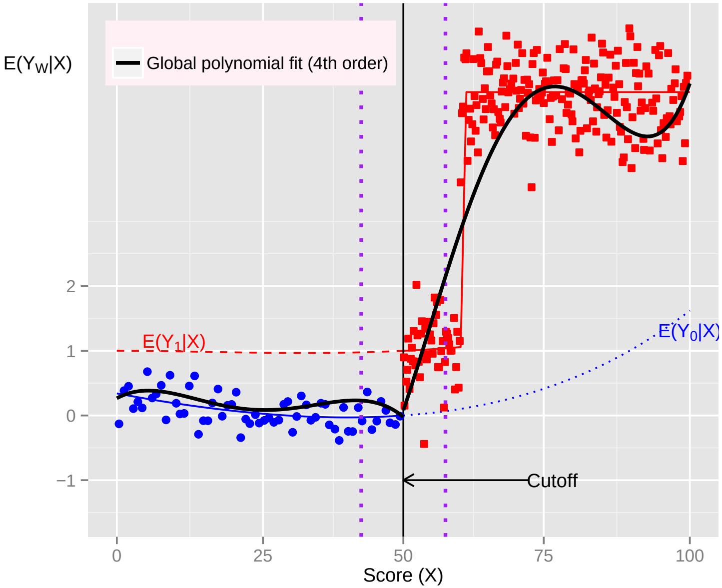


Figure: Danger in Global Approximations



Conventional Local Polynomial Approach

- **Idea**: approximate regression functions for control and treatment units *locally*.

- "Local-linear" estimator (w/ weights $K(\cdot)$):

$-h_n \leq X_i < \bar{x}$:	$\bar{x} \leq X_i \leq h_n$:
$Y_i = \alpha_- + (X_i - \bar{x}) \cdot \beta_- + \varepsilon_{-,i}$	$Y_i = \alpha_+ + (X_i - \bar{x}) \cdot \beta_+ + \varepsilon_{+,i}$

  - Treatment effect (at the cutoff): $\hat{\tau}_{\mathtt{SRD}}(h_n) = \hat{\alpha}_+ - \hat{\alpha}_-$

- Can be estimated using linear models (w/ weights $K(\cdot)$):

$$Y_i = \alpha + \tau_{\mathtt{SRD}} \cdot T_i + (X_i - \bar{x}) \cdot \beta_1 + T_i \cdot (X_i - \bar{x}) \cdot \gamma_1 + \varepsilon_i, \qquad -h_n \leq X_i \leq h_n$$

- Once h_n chosen, inference is "standard": weighted linear models.

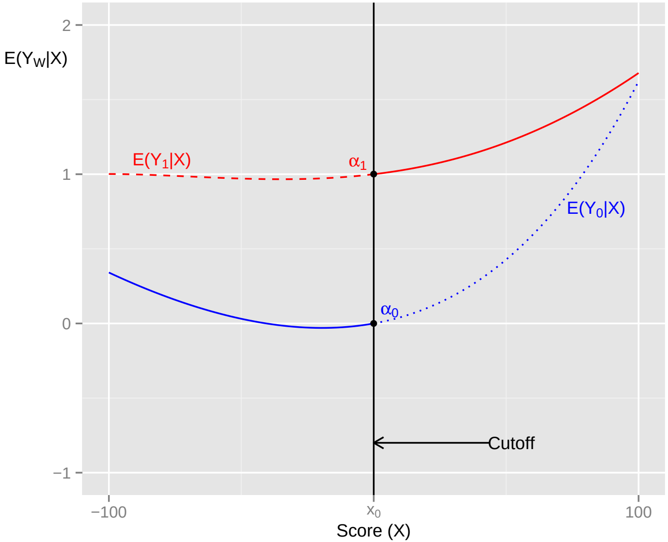
Conventional Local Polynomial Approach

- $\mathbb{E}[Y_i|X_i = x]$ approximated in neighborhood of $\bar{x}$ by polynomial

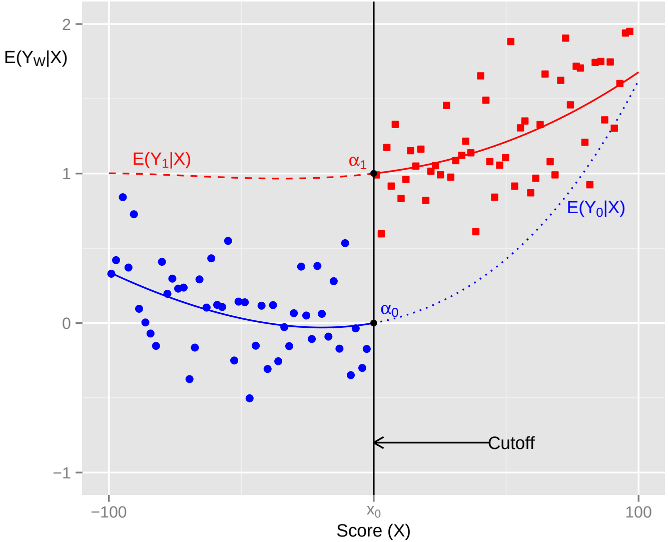
- Local polynomial estimation:

  1. Choose bandwidth h to keep observations in $[\bar{x} - h, \bar{x} + h]$
  2. Choose kernel function to weigh observations, $w_i = K(\frac{x_i - \bar{x}}{h})$
  3. Choose order of polynomial $(p = 1)$

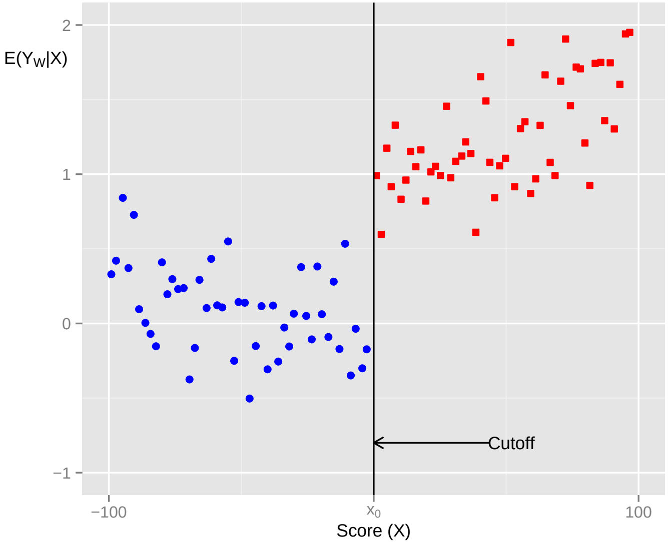
Local Polynomial Estimation



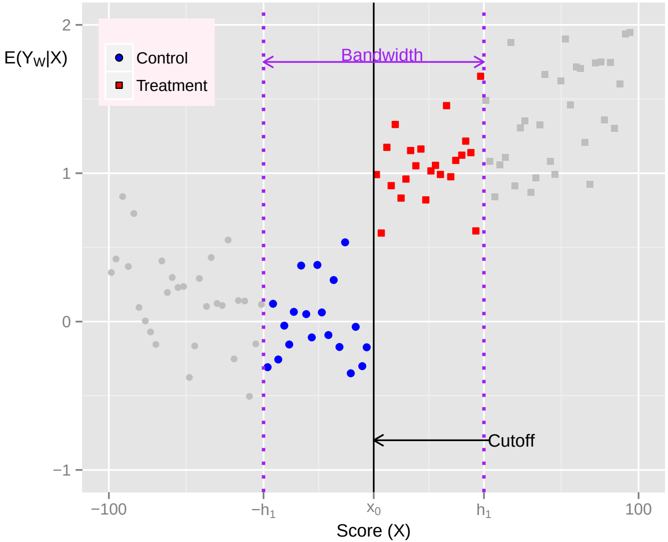
Local Polynomial Estimation



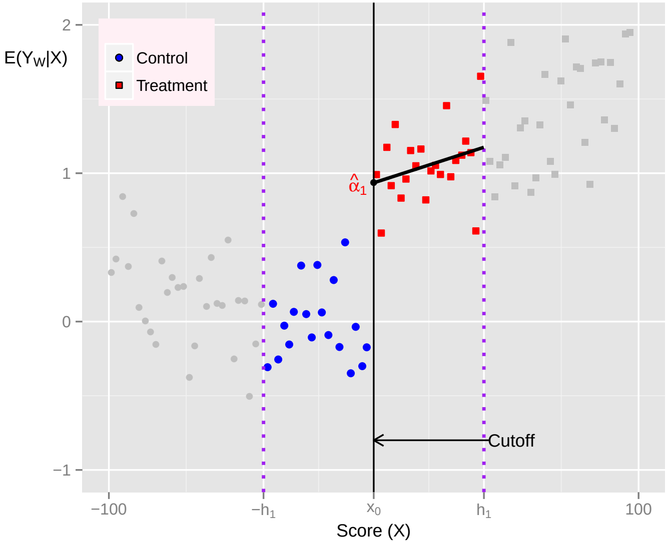
Local Polynomial Estimation



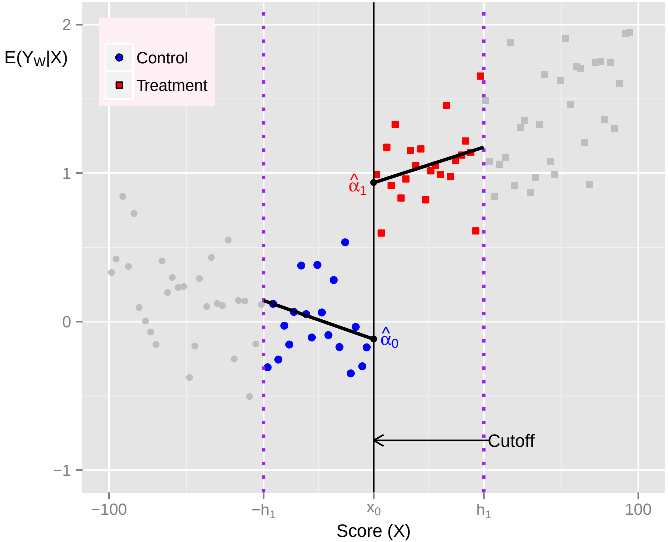
Local Polynomial Estimation



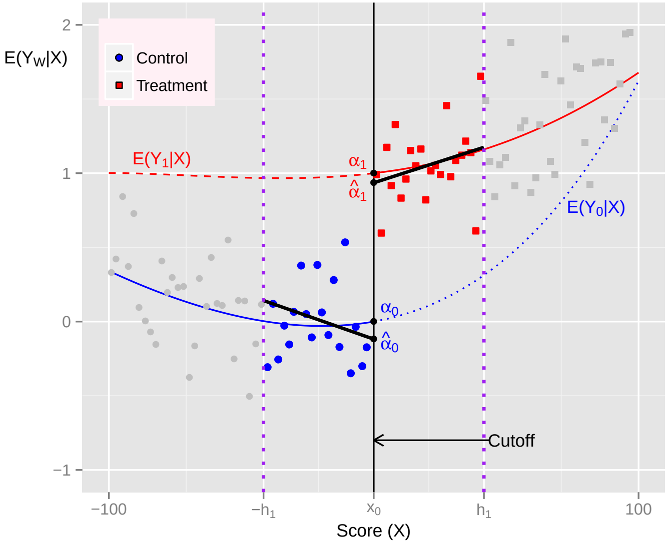
Local Polynomial Estimation



Local Polynomial Estimation



Local Polynomial Estimation



Conventional Local Polynomial Approach: Choosing bandwidth

- How to choose h_n?

- Mean Square Error Optimal (MSE-optimal) Choice
[Imbens & Kalyanaraman 2012 REStud, Calonico, Cattaneo & Titiunik 2014 ECMA):

$$\hat{h}_{\texttt{mse}} = \hat{C}_{\texttt{mse}} \cdot n^{-1/5}$$

- Coverage Error Rate Optimal (CER-optimal) Choice
[Calonico, Cattaneo & Farrell, 2017, JASA]

$$\hat{h}_{\texttt{cer}} = n^{-1/20} \cdot \hat{h}_{\texttt{mse}}$$

- **Key idea**: trade-off bias and variance of $\hat{\tau}_{\texttt{SRD}}(h_n)$. Heuristically:

$$\uparrow \ \mathsf{Bias}(\hat{\tau}_{\texttt{SRD}}) \quad \Longrightarrow \quad \downarrow \hat{h} \qquad \text{and} \qquad \uparrow \ \mathsf{Var}(\hat{\tau}_{\texttt{SRD}}) \quad \Longrightarrow \quad \uparrow \hat{h}$$

Local Polynomial Methods: Bandwidth Selection

- Two main methods: MSE & CER. Useful for different goals.

- MSE-Optimal (Calonico, Cattaneo & Titiunik 2014 ECMA):

$$h_{\texttt{mse}} = C_{\texttt{mse}}^{1/5} \cdot n^{-1/5} \qquad\qquad C_{\texttt{mse}} = C(K) \cdot \frac{\mathsf{Var}(\hat{\tau}_{\texttt{SRD}})}{\mathsf{Bias}(\hat{\tau}_{\texttt{SRD}})^2}$$

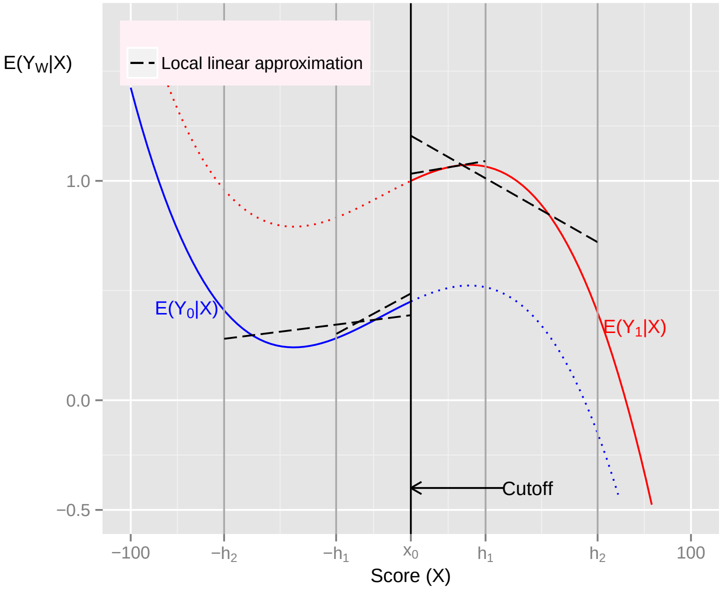
  - IK implementation: first-generation plug-in rule.
  - CCT implementation: second-generation plug-in rule.
  - They differ in the way $\mathsf{Var}(\hat{\tau}_{\texttt{SRD}})$ and $\mathsf{Bias}(\hat{\tau}_{\texttt{SRD}})$ are estimated.

- CER-optimal (Calonico, Cattaneo & Farrell 2016+ JASA):

$$h_{\texttt{cer}} = C_{\texttt{cer}}^{1/4} \cdot n^{-1/4} \qquad\qquad C_{\texttt{cer}} \propto C(K) \cdot \frac{\mathsf{Var}(\hat{\tau}_{\texttt{SRD}})}{|\mathsf{Bias}(\hat{\tau}_{\texttt{SRD}})|}$$

  - Trade-offs bias and variance differently.
  - CCF suggest simple rescaling: $n^{1/20}$.

Figure: Bias in Local Approximations—Choosing bandwidth



Conventional Approach to RD

- "Local-linear" estimator (w/ weights $K(\cdot)$):

$-h_n \leq X_i < \bar{x}$:	$\bar{x} \leq X_i \leq h_n$:
$Y_i = \alpha_- + (X_i - \bar{x}) \cdot \beta_- + \varepsilon_{-,i}$	$Y_i = \alpha_+ + (X_i - \bar{x}) \cdot \beta_+ + \varepsilon_{+,i}$

  - Treatment effect (at the cutoff): $\hat{\tau}_{\mathtt{SRD}} = \hat{\alpha}_+ - \hat{\alpha}_-$

- Construct usual t-test. For $\mathsf{H}_0 : \tau_{\mathtt{SRD}} = 0$,

$$\hat{T}(h_n) = \frac{\hat{\tau}_{\mathtt{SRD}}}{\sqrt{\hat{\mathsf{V}}_n}} = \frac{\hat{\alpha}_+ - \hat{\alpha}_-}{\sqrt{\hat{\mathsf{V}}_{+,n} + \hat{\mathsf{V}}_{-,n}}} \approx_d \mathcal{N}(0,1)$$

- 95% Confidence interval:

$$\hat{I}(h_n) = \left[\ \hat{\tau}_{\mathtt{SRD}} \ \pm \ 1.96 \cdot \sqrt{\hat{\mathsf{V}}_n}\ \right]$$

Bias-Correction Approach to RD

- Note well: for usual t-test,

$$\hat{T}(h_{\texttt{MSE}}) = \frac{\hat{\tau}_{\texttt{SRD}}}{\sqrt{\hat{\mathsf{V}}_n}} \approx_d \mathcal{N}(\mathsf{B}, 1) \quad \neq \quad \mathcal{N}(0,1), \qquad \mathsf{B} > 0$$

  - Bias B in RD estimator captures "curvature" of regression functions.

- Undersmoothing/"Small Bias" Approach: Choose "smaller" h_n... Perhaps $\hat{h}_n = 0.5 \cdot \hat{h}_{\texttt{IK}}$?

$\Longrightarrow$ Not clear guidance & power loss!

- Bias-correction Approach:

$$\hat{T}^{\mathsf{bc}}(h_n, b_n) = \frac{\hat{\tau}_{\texttt{SRD}} - \hat{\mathsf{B}}_n}{\sqrt{\hat{\mathsf{V}}_n}} \approx_d \mathcal{N}(0,1)$$

$\Longrightarrow$ 95% Confidence Interval: $\hat{I}^{\mathsf{bc}}(h_n, b_n) = \left[\ \left(\hat{\tau}_{\texttt{SRD}} - \hat{\mathsf{B}}_n\right) \ \pm \ 1.96 \cdot \sqrt{\hat{\mathsf{V}}_n} \ \right]$

- How to choose b_n? Same ideas as before... $\hat{b}_n = \hat{C} \cdot n^{-1/7}$

Robust Bias-Correction Approach to RD

- **Recall**:

$$\hat{T}(h_n) = \frac{\hat{\tau}_{\mathtt{SRD}}}{\sqrt{\hat{\mathsf{V}}_n}} \approx_d \mathcal{N}(0,1) \qquad \text{and} \qquad \hat{T}^{\mathtt{bc}}(h_n, b_n) = \frac{\hat{\tau}_{\mathtt{SRD}} - \hat{\mathsf{B}}_n}{\sqrt{\hat{\mathsf{V}}_n}} \approx_d \mathcal{N}(0,1)$$

  - $\hat{\mathsf{B}}_n$ is constructed to estimate leading bias B.

- **Robust approach**:

$$\hat{T}^{\mathtt{bc}}(h_n, b_n) = \frac{\hat{\tau}_{\mathtt{SRD}} - \hat{\mathsf{B}}_n}{\sqrt{\hat{\mathsf{V}}_n}} = \underbrace{\frac{\hat{\tau}_{\mathtt{SRD}} - \mathsf{B}_n}{\sqrt{\hat{\mathsf{V}}_n}}}_{\approx_d \, \mathcal{N}(0,1)} + \underbrace{\frac{\mathsf{B}_n - \hat{\mathsf{B}}_n}{\sqrt{\hat{\mathsf{V}}_n}}}_{\approx_d \, \mathcal{N}(0,\gamma)}$$

- **Robust bias-corrected t-test**:

$$\hat{T}^{\mathtt{rbc}}(h_n, b_n) = \frac{\hat{\tau}_{\mathtt{SRD}} - \hat{\mathsf{B}}_n}{\sqrt{\hat{\mathsf{V}}_n + \hat{\mathsf{W}}_n}} = \frac{\hat{\tau}_{\mathtt{SRD}} - \hat{\mathsf{B}}_n}{\sqrt{\hat{\mathsf{V}}_n^{\mathtt{bc}}}} \approx_d \mathcal{N}(0,1)$$

$\Longrightarrow$ 95% Confidence Interval:

$$\hat{I}^{\mathtt{rbc}}(h_n, b_n) = \left[\ \left(\hat{\tau}_{\mathtt{SRD}} - \hat{\mathsf{B}}_n\right) \ \pm \ 1.96 \cdot \sqrt{\hat{\mathsf{V}}_n^{\mathtt{bc}}} \ \right], \qquad \hat{\mathsf{V}}_n^{\mathtt{bc}} = \hat{\mathsf{V}}_n + \hat{\mathsf{W}}_n$$

Local-Polynomial Methods: Robust Inference

- Approach 1: *Undersmoothing/"Small Bias"*.

$$\hat{I}(h_n) = \left[\; \hat{\tau}_{\texttt{SRD}} \;\pm\; 1.96 \cdot \sqrt{\hat{\mathsf{V}}_n} \; \right]$$

- Approach 2: *Bias correction* (**not recommended**).

$$\hat{I}^{\mathsf{bc}}(h_n, b_n) = \left[\; \left(\hat{\tau}_{\texttt{SRD}} - \hat{\mathsf{B}}_n\right) \;\pm\; 1.96 \cdot \sqrt{\hat{\mathsf{V}}_n} \; \right]$$

- Approach 3: *Robust Bias correction*.

$$\hat{I}^{\mathtt{rbc}}(h_n, b_n) = \left[\; \left(\hat{\tau}_{\texttt{SRD}} - \hat{\mathsf{B}}_n\right) \;\pm\; 1.96 \cdot \sqrt{\hat{\mathsf{V}}_n + \hat{\mathsf{W}}_n} \; \right]$$

Estimation and Inference Methods

- Local polynomial methods.
  - Standard Approaches.
  - Bandwidth selection.
  - Robust Bias-correction.
  - Confidence intervals.
- Local randomization methods (not today).
  - Interpreting RD as a local randomization in a window around the cutoff.
  - Conceptual differences with local polynomial estimation.
  - Window selection.
  - Estimation and inference using randomization-based methods.

Outline

Goal

- RDD with multiple cutoffs are common in practice.
- Researchers usually pool cutoffs by re-centering the running variable.
- Questions:
  - What parameter is identified when pooling?
  - What are the parameters of interest in this context?
  - Can variation in cutoffs be exploited to identify them?

"Classical" Regression Discontinuity designs

- Potential outcomes: $(Y_i(1), Y_i(0))$, with treatment effect:

$$\tau_i = Y_i(1) - Y_i(0)$$

- Running variable (*score*): X_i.
- Treatment indicator: $D_i = D_i(X_i) = 1$ if treated, 0 otherwise.
- Observed outcome: $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$.
- Sharp design: $D_i = \mathbf{1}(X_i \geq c)$.
- Under smoothness,

$$\mathbb{E}[\tau_i \mid X_i = c] = \lim_{x \to c^+} \mathbb{E}[Y_i \mid X_i = x] - \lim_{x \to c^-} \mathbb{E}[Y_i \mid X_i = x]$$

RD with multiple cutoffs: motivation

- Frequently, programs or policies have multiple cutoffs.

Table: Progresa (Mexico)

Region	Cutoff	Obs
27	691.0	828
6	751.0	541
5	751.5	3,116
4	753.0	1,189
3	759.4	933
28	853.3	175

Table: P-900 (Chile)

Region	Cutoff	Obs
7	42.4	157
6,8	43.4	497
13	46.4	959
9	47.4	197
2,5,10	49.4	560
1,3,4	51.4	190

RD with multiple cutoffs

- Common empirical approach: pooling.
  - $C_i \in \mathcal{C}$ (random) cutoff faced by unit i.
  - Discrete cutoffs: $\mathcal{C} = \{c_0, c_1, ..., c_J\}$.
  - Re-centered running variable: $\tilde{X}_i = X_i - C_i$.
  - Pooled estimand:

$$\tau^p = \lim_{x \to 0^+} \mathbb{E}[Y_i \mid \tilde{X}_i = x] - \lim_{x \to 0^-} \mathbb{E}[Y_i \mid \tilde{X}_i = x]$$

- What parameter is this approach identifying?

Identification under the pooling approach

$$\tau^p = \lim_{x \to 0^+} \mathbb{E}[Y_i \mid \tilde{X}_i = x] - \lim_{x \to 0^-} \mathbb{E}[Y_i \mid \tilde{X}_i = x]$$

Identification under pooling

If the CEFs and $f_{X|C}(x|c)$ are continuous at the cutoffs,

$$\tau^p = \sum_{c \in \mathcal{C}} \mathbb{E}[\tau_i \mid X_i = c, C_i = c]\omega(c)$$

where

$$\omega(c) = \frac{f_{X|C}(c|c)\mathbb{P}[C_i = c]}{\sum_{c \in \mathcal{C}} f_{X|C}(c|c)\mathbb{P}[C_i = c]}$$

Exploiting multiple cutoffs

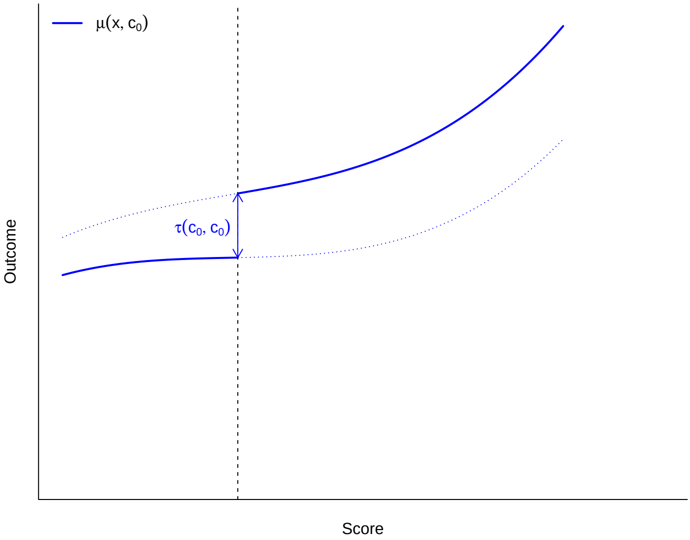
- Two drawbacks of the pooling approach:
  - It discards variation that can identify parameters of interest,
  - Unclear policy relevance: it combines TEs at different cutoffs *for different populations*.
- What are the parameters of interest in this context?
- Potential CEFs:

$$\mu_d(x,c) := \mathbb{E}[Y_i(d)|X_i = x, C_i = c], \qquad d \in \{0,1\}$$

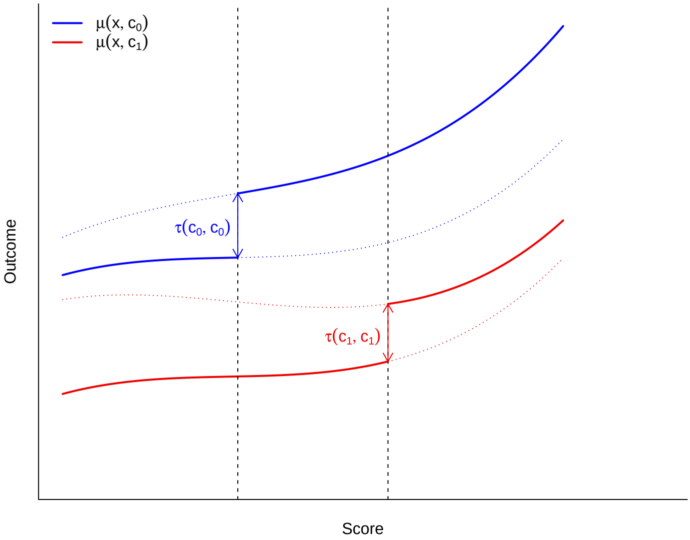
- (Conditional) ATE:

$$\tau(x,c) := \mathbb{E}[\tau_i \mid X_i = x, C_i = c] = \mu_1(x,c) - \mu_0(x,c)$$

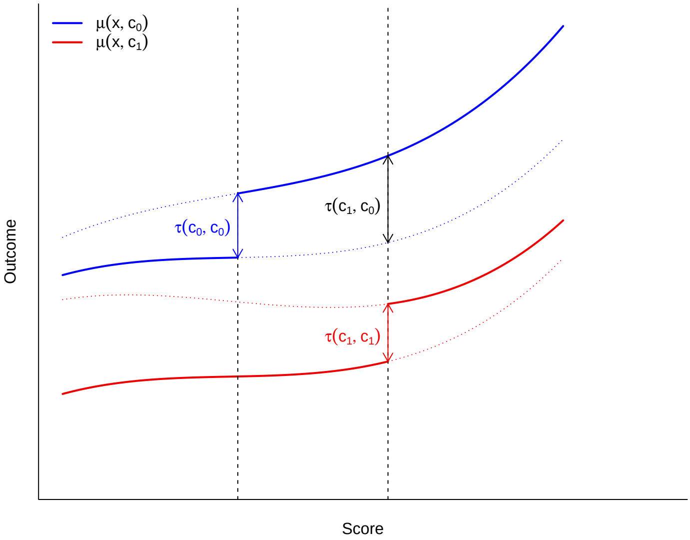
Exploiting multiple cutoffs: parameters of interest



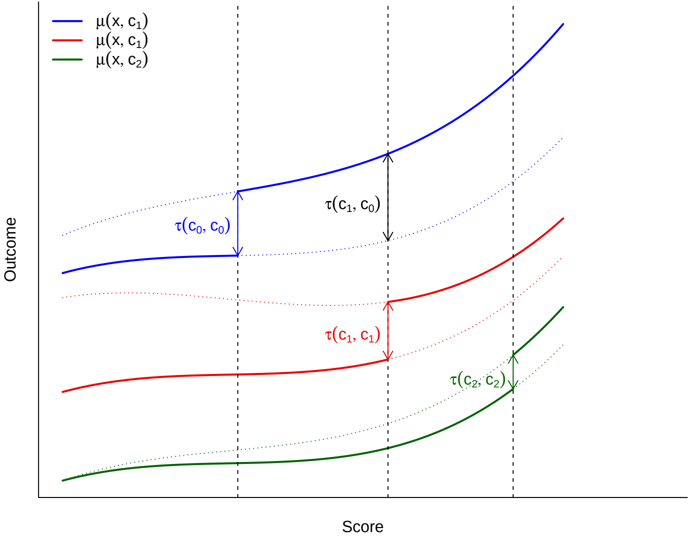
Exploiting multiple cutoffs: parameters of interest



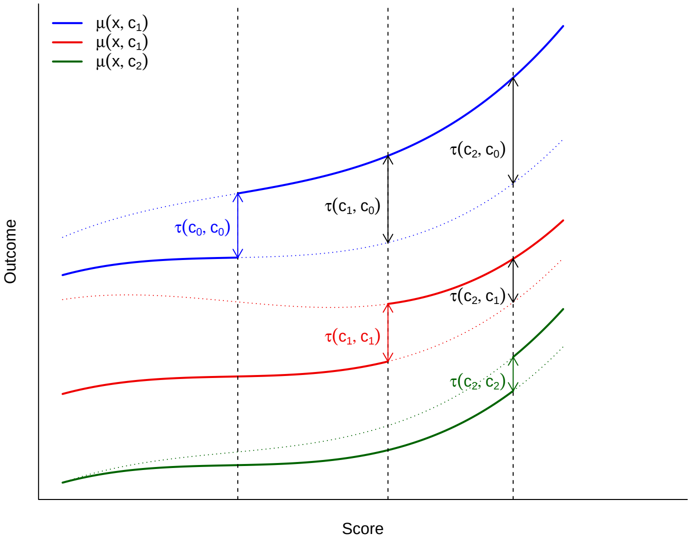
Exploiting multiple cutoffs: parameters of interest



Exploiting multiple cutoffs: parameters of interest



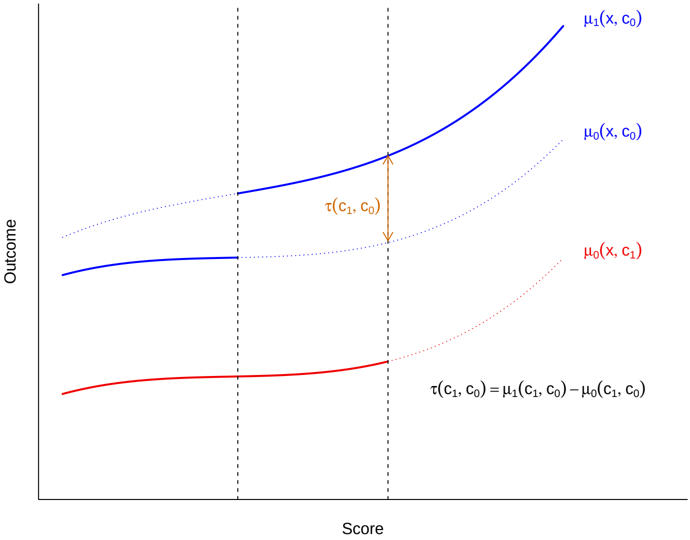
Exploiting multiple cutoffs: parameters of interest



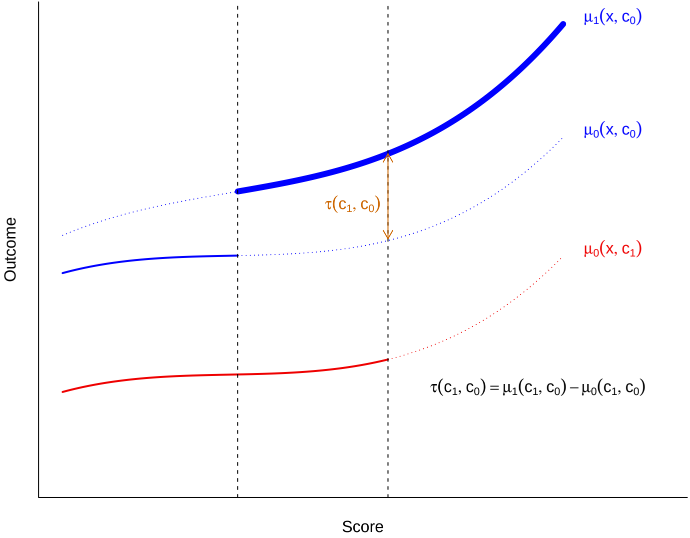
Intuition: two-cutoff case

- Can we exploit variation in cutoffs to identify magnitudes of interest?
- Intuition: focus on $\tau(c_1, c_0)$.
- No untreated units among $C_i = c_0$...
- But we can try to use use units with $C_i = c_1$.

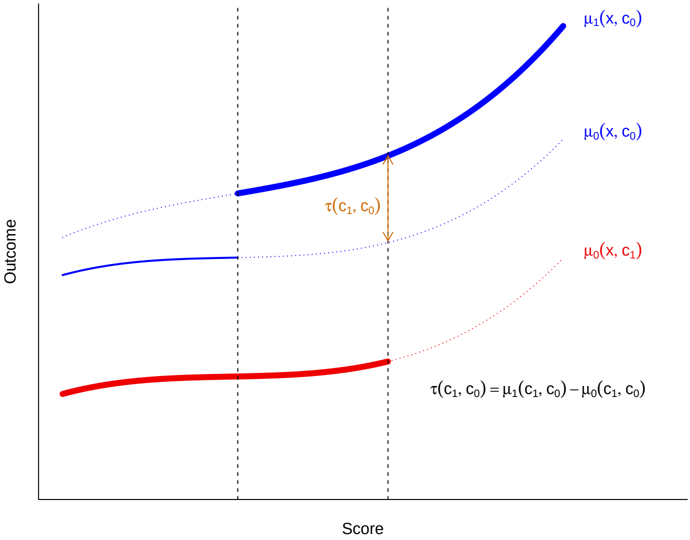
Intuition: two-cutoff case



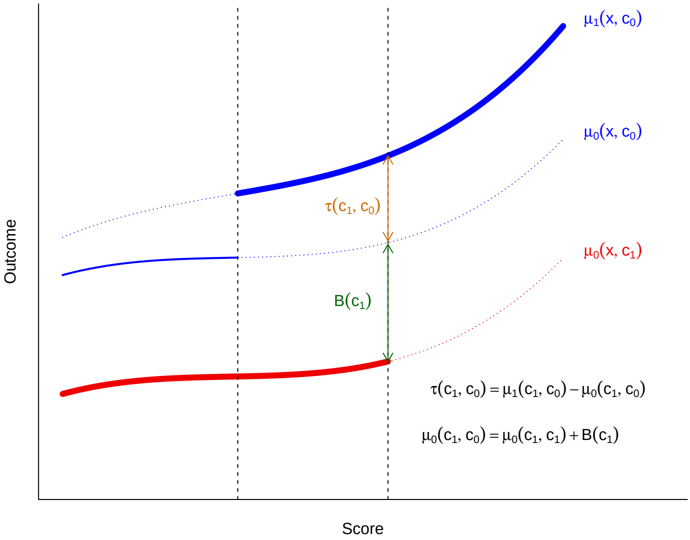
Intuition: two-cutoff case



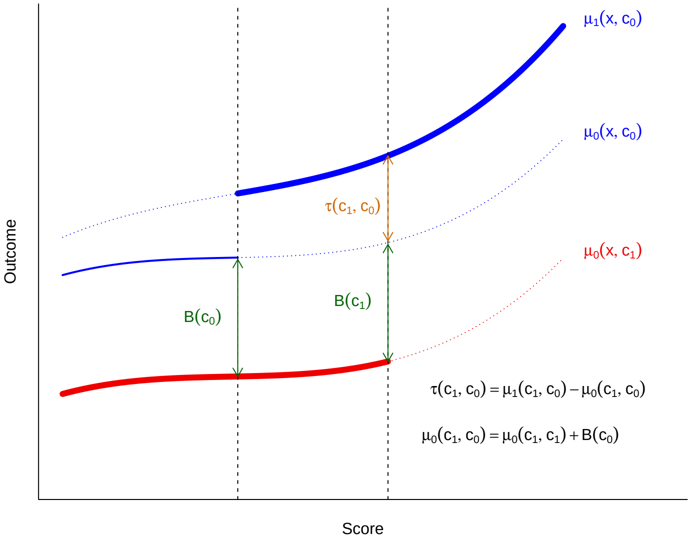
Intuition: two-cutoff case



Intuition: two-cutoff case



Intuition: two-cutoff case



Summary

- Researchers usually pool different cutoffs together.
  - Identified parameter is a weighted average of TEs.
  - Not necessarily what the researcher was looking for.
  - Discards potentially useful information.
- We can exploit variation in cutoffs to study TE heterogeneity.
  - Response function: how the TE changes with X.
  - External validity: how the TE changes across subpopulations.

Extensions

- Generalization of constant-bias assumption:

$$B(c_1) \approx B(c_0) + \sum_{s=1}^{p} \frac{1}{s!} B^{(s)}(c_0) \cdot [c_1 - c_0]^s$$

$\rightarrow$ account for differences in slopes, curvature, etc.

- Implementation with more than two cutoffs: "fixed effects" model.

$$\mu_0(x, c_j) = g(x) + \theta_j$$

- Combining both approaches:

$$\mu_0(x, c_j) = g(x) + p_k(x)'\theta_j$$