Regression Discontinuity Designs

Matias D. Cattaneo

Princeton University

January 4, 2022

Prepared for Amazon / University of Washington.

Complementary materials available at https://rdpackages.github.io/

NSF support through grants SES-1357561, SES-1459931, SES-1947805, SES-2019432.

Outline

Causal Inference & Program Evaluation

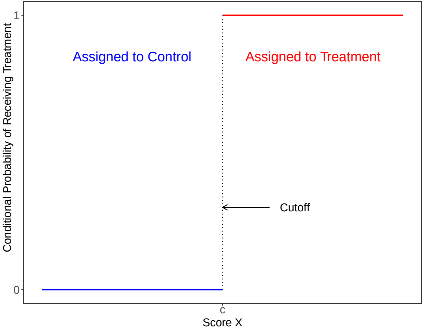
- Main goal: learn about treatment effect of policy or intervention.
- If treatment randomization available, easy to estimate treatment effects.
- If treatment randomization not available, turn to observational studies.
  - Selection on Observables, Instrumental Variables, Selection on Unobservables,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

RD Designs: Building Block

- **Triplet**: *score*, *cutoff*, *treatment*.
  - Units receive a score.
  - A treatment is assigned based on the score and a *known* cutoff.
  - The treatment

    is offered to units whose score is greater than the cutoff.

    is 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 causal treatment effects.

RD Designs: Taxonomy

- **Frameworks**.
  - Identification: Continuity/Extrapolation, Local Randomization.
  - Score: Continuous, Many Repeated, Few Repeated.
- **Settings**.
  - Sharp, Fuzzy, Kink, Kink Fuzzy.
  - Multiple Cutoff, Multiple Scores, Geographic RD.
  - Dynamic, Continuous Treatments, Time, etc.
- **Parameters of Interests**.
  - Average Effects, Quantile/Distributional Effects, Partial Effects.
  - Heterogeneity, Covariate-Adjustment, Differences, Time.
  - Extrapolation.

RCTs vs. (Sharp) RD Designs

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

- **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), $i = 1, 2, \ldots, n$, 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**:

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

RCTs vs. (Sharp) RD Designs

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

- **Treatment**: $T_i \in \{0, 1\}$, $\qquad T_i = \mathbb{1}(X_i \geq c)$, $\qquad c$ cutoff.

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

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

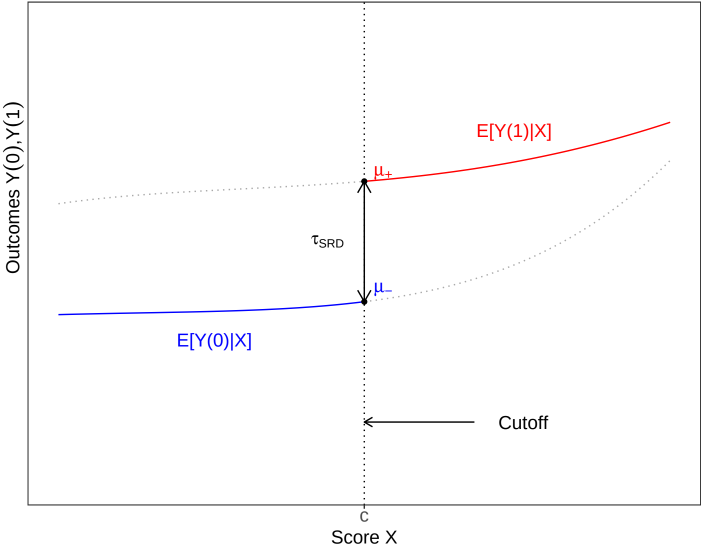
- **Average Treatment Effect at the cutoff** (Continuity-based):

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

- **Average Treatment Effect in a neighborhood** (LR-based):

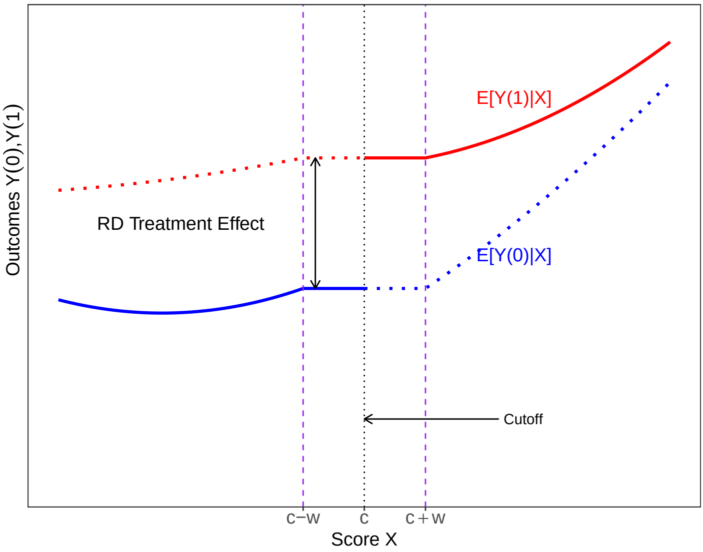
$$\tau_{\texttt{LR}} = \mathbb{E}[Y_i(1) - Y_i(0) | X_i \in \mathcal{W}] = \frac{1}{N_1} \sum_{X_i \in \mathcal{W}, T_i = 1} Y_i - \frac{1}{N_0} \sum_{X_i \in \mathcal{W}, T_i = 0} Y_i$$

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



$$T_i \text{ independent of } (Y_i(0), Y_i(1)) \text{ for all } X_i \in \mathcal{W} = [c-w, c+w]$$

$$+ \text{ exclusion restriction}$$



Fuzzy RD Designs

- **Imperfect compliance.**
  - probability of receiving treatment changes at c, but not necessarily from 0 to 1.

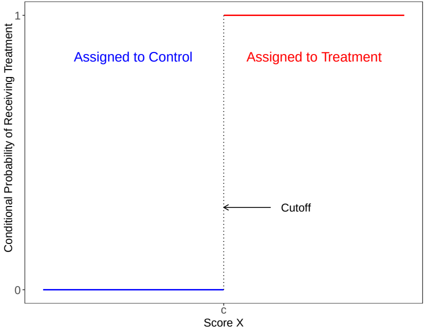
- Canonical Parameter:

$$\tau_{\texttt{FRD}} = \frac{\mathbb{E}[(Y_i(1) - Y_i(0)(D_i(1) - D_i(0)))|X_i = c]}{\mathbb{E}[D_i(1)|X_i = c] - \mathbb{E}[D_i(0)|X_i = c]}$$
$$= \frac{\lim_{x \downarrow c} \mathbb{E}[Y_i|X_i = x] - \lim_{x \uparrow c} \mathbb{E}[Y_i|X_i = x]}{\lim_{x \downarrow c} \mathbb{E}[D_i|X_i = x] - \lim_{x \uparrow c} \mathbb{E}[D_i|X_i = x]}$$

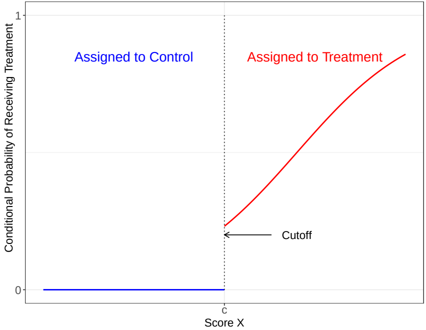
  where $Y_i(t) = Y_i(0)(1 - D_i(t)) + Y_i(1)D_i(t)$ and $D_i(t) = D_i(0)(1 - T_i) + D_i(1)T_i$.

- Similarly for Local Randomization framework.

- Different interpretations under different assumptions.



(a) Sharp RD



(b) Fuzzy RD (one-sided compliance)

(Sharp and Fuzzy) Kink RD Designs

- Treatment assigned via continuous score formula, but slope changes discontinuously at "kink" point (c).

- SKRD Parameter:

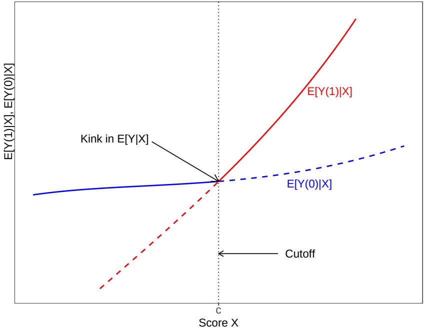
$$\tau_{\texttt{KRD}} = \frac{\lim_{x \downarrow c} \frac{d}{dx}\mathbb{E}[Y_i|X_i = x] - \lim_{x \uparrow c} \frac{d}{dx}\mathbb{E}[Y_i|X_i = x]}{\lim_{x \downarrow c} \frac{d}{dx} b(x) - \lim_{x \uparrow c} \frac{d}{dx} b(x)}$$

  where $b(x)$ known function inducing "kink".

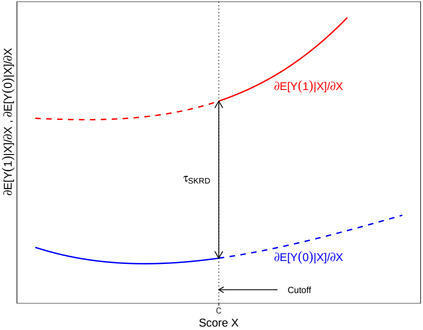
- FKRD Parameter:

$$\tau_{\texttt{KRD}} = \frac{\lim_{x \downarrow c} \frac{d}{dx}\mathbb{E}[Y_i|X_i = x] - \lim_{x \uparrow c} \frac{d}{dx}\mathbb{E}[Y_i|X_i = x]}{\lim_{x \downarrow c} \frac{d}{dx}\mathbb{E}[D_i|X_i = x] - \lim_{x \uparrow c} \frac{d}{dx}\mathbb{E}[D_i|X_i = x]}$$

- Different interpretation under different assumptions.



(a) Kink RD (levels)



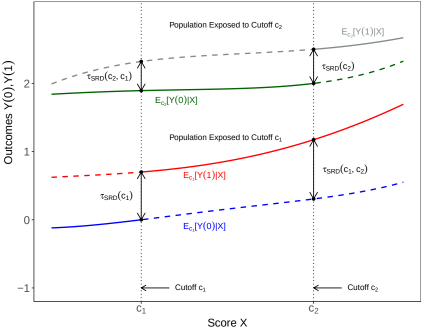
(b) Kink RD (derivatives)

Multi-cutoff, Multi-Score, Geographic RD Designs

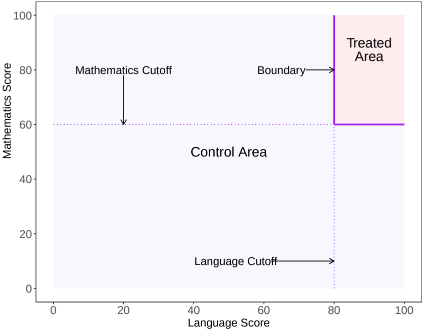
- **Multi-cutoff RD designs**.
  - $C_i \in \mathcal{C}$ with $\mathcal{C} = \{c_1, c_2, \cdots, c_J\}$ or $\mathcal{C} = [\underline{c}, \overline{c}]$.
  - Two strategies: normalize-and-pool ($\tilde{X}_i = X_i - C_i$), or cutoff-by-cutoff analysis.
  - Different interpretation under different assumptions.

- **Multi-score RD designs**.
  - $\mathbf{X}_i = (X_{1i}, X_{2i}, \ldots, X_{di})'$ and $\mathbf{c} = (c_1, c_2, \ldots, c_d)'$.
  - Can always be mapped back to Multi-cutoff RD designs.
  - Leading special cases: Test scores, geography ($d = 2$).
  - Different interpretation under different assumptions.

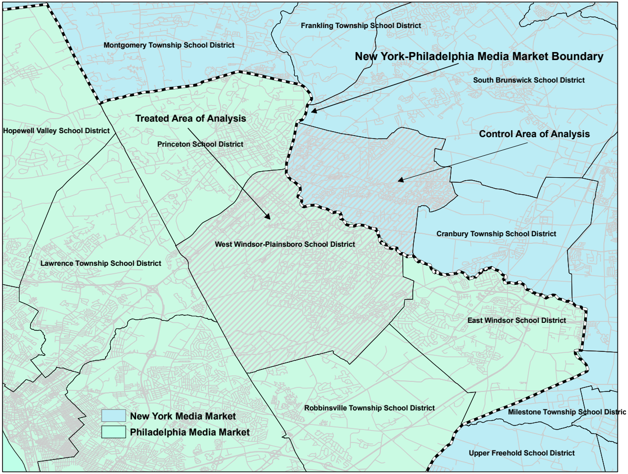
- **Other RD-like designs**.
  - RD in density and bunching designs.
  - RD in time.
  - Dynamic RD designs.
  - etc.



(a) Multi-cutoff: $\tau_{\mathtt{SRD}}(x, c) = \mathbb{E}[Y_i(1) - Y_i(0)|X_i = x, C_i = c]$



(b) Multi-score: $\tau_{\mathtt{SRD}}(x_1, x_2) = \mathbb{E}[Y_i(1) - Y_i(0)|X_{1i} = x_1, X_{2i} = x]$



Frankling Township School District
Montgomery Township School District
New York-Philadelphia Media Market Boundary
South Brunswick School District
Treated Area of Analysis
Hopewell Valley School District
Princeton School District
Control Area of Analysis
Cranbury Township School District
West Windsor-Plainsboro School District
Lawrence Township School District
East Windsor School District
Robbinsville Township School District
Milestone Township School Distric
Upper Freehold School District
New York Media Market
Philadelphia Media Market

Highlights and Main Takeaways

- RD designs exploit “variation” near the cutoff.
- Causal effect is different (in general) than RCT.
- No “overlap” (sharp) so extrapolation is unavoidable (local or global).
- Graphical analysis is both very useful and very dangerous.
- Need to work with data near cutoff $\Longrightarrow$ bandwidth or window selection.
- There exist many design-specific falsification/validation methods.

Outline

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 $(c = 59.1984)$

  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**:

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

RD Packages: Python, R, Stata

`https://rdpackages.github.io/`

- **rdrobust**: estimation, inference and graphical presentation using local polynomials, partitioning, and spacings estimators.
  - `rdrobust`, `rdbwselect`, `rdplot`.
- **rddensity**: discontinuity in density tests (manipulation testing) using both local polynomials and binomial tests.
  - `rddensity`, `rdbwdensity`.
- **rdlocrand**: covariate balance, binomial tests, randomization inference methods (window selection & inference).
  - `rdrandinf`, `rdwinselect`, `rdsensitivity`, `rdrbounds`.
- **rdmulti**: multiple cutoffs and multiple scores.
- **rdpower**: power, sample selection and minimum detectable effect size.

RD Plots

- Main ingredients:
  - Global smooth polynomial fit.
  - Binned discontinuous local-means fit.
- Main goals:
  - Graphical (heuristic) representation.
  - Detention of discontinuities.
  - Representation of variability.
- Tuning parameters:
  - Global polynomial degree.
  - Location (ES or QS) and number of bins.
- **Great to convey ideas but horrible to draw conclusions**.

Estimation and Inference Methods

- **Continuity/Extrapolation**: Local polynomial approach.
  - Localization: bandwidth selection (trade-off bias and variance).
  - Point estimation: "flexible" (nonparametric).
  - Inference: robust bias-corrected methods.
- **Local Randomization**: finite-sample and large-sample inference.
  - Localization: window selection (via local independence implications).
  - Point estimation: parametric, finite-sample (Fisher) or large-sample (Neyman/SP).
  - Inference: randomization inference (Fisher) or large-sample (Neyman/SP).
- Many refinements and other methods exist (EL, Bayesian, Uniformity, etc.).
  - Do not offer much improvements in applications.
  - Can be overly complicated (lack of transparency).
  - Can depend on user-chosen tuning parameters (lack of replicability).

Continuity/Extrapolation: Local Polynomial Methods

- Global polynomial regression: **not recommended**.
  - Runge's Phenomenon, counterintuitive weights, overfitting, lack of robustness.

- Local polynomial regression: captures idea of "localization".

Choose low poly order (p) and weighting scheme ($K(\cdot)$)

⇩

Choose bandwidth h: MSE-optimal or CE-optimal

⇩

Construct point estimator $\hat{\tau}$
(MSE-optimal $h \implies$ optimal estimator)

⇩

Conduct robust bias-corrected inference
(CE-optimal $h \implies$ optimal distributional approximation)

Local Polynomial Methods

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

- "Local-linear" ($p = 1$) estimator (w/ weights $K(\cdot)$):

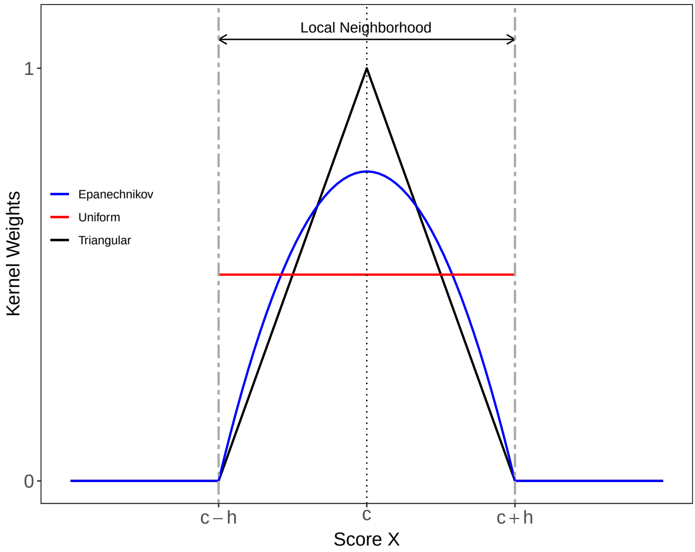
$-h \leq X_i < c:$	$c \leq X_i \leq h:$
$Y_i = \alpha_- + (X_i - c) \cdot \beta_- + \varepsilon_{-,i}$	$Y_i = \alpha_+ + (X_i - c) \cdot \beta_+ + \varepsilon_{+,i}$

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

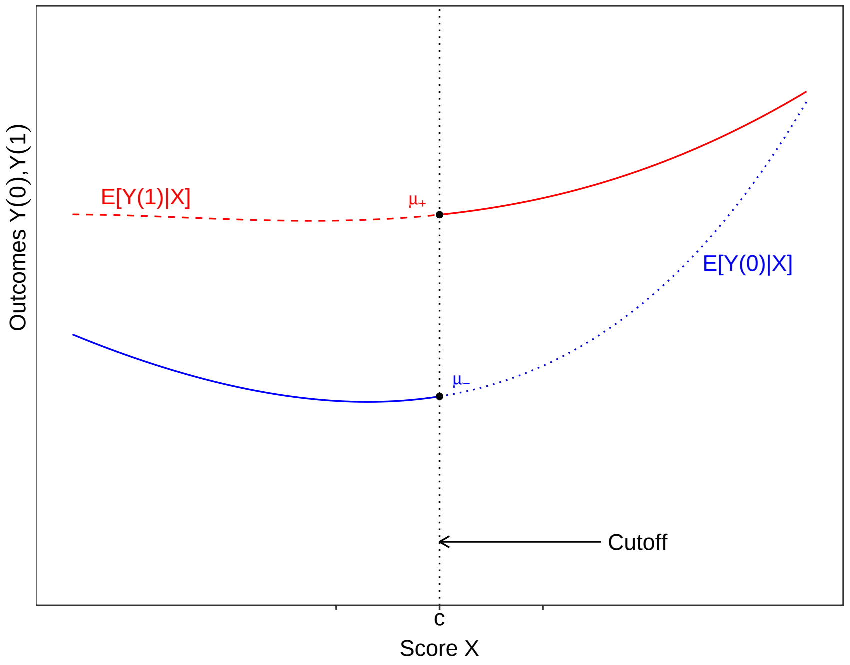
- Can be estimated using linear models (w/ weights $K(\cdot)$):

$$Y_i = \alpha + \tau_{\texttt{SRD}} \cdot T_i + (X_i - c) \cdot \beta_1 + T_i \cdot (X_i - c) \cdot \gamma_1 + \varepsilon_i, \qquad |X_i - c| \leq h$$

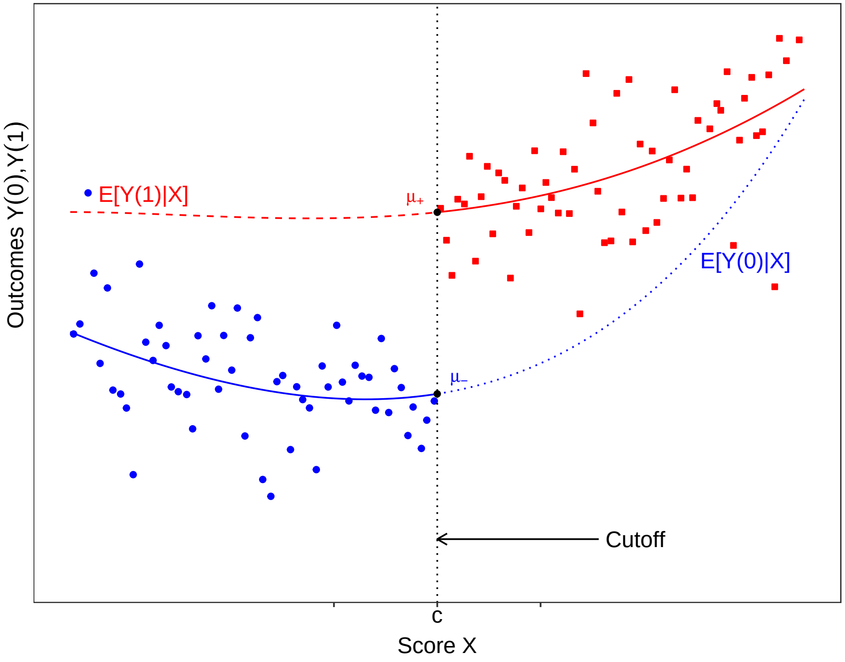
- Given p, K, h chosen $\implies$ weighted least squares estimation.

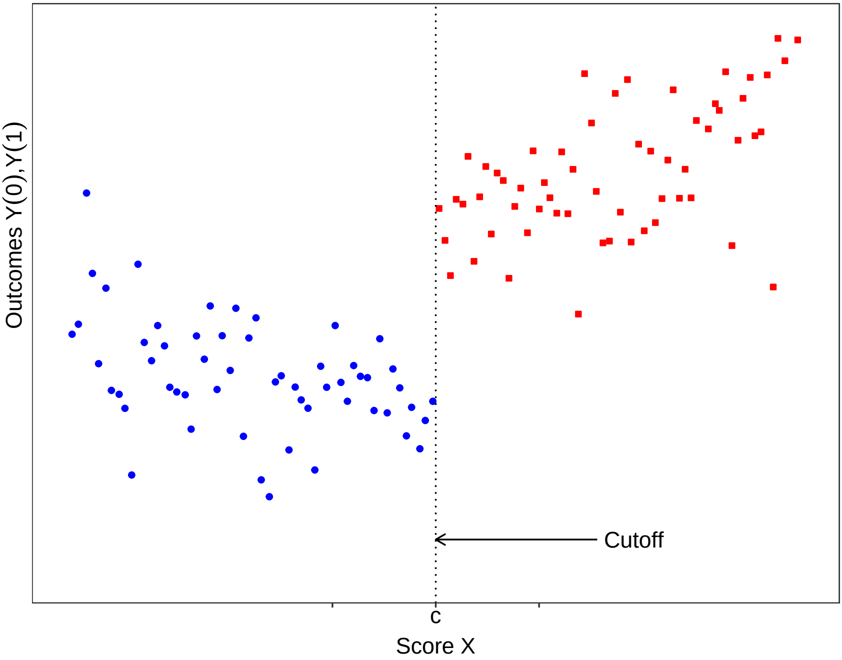


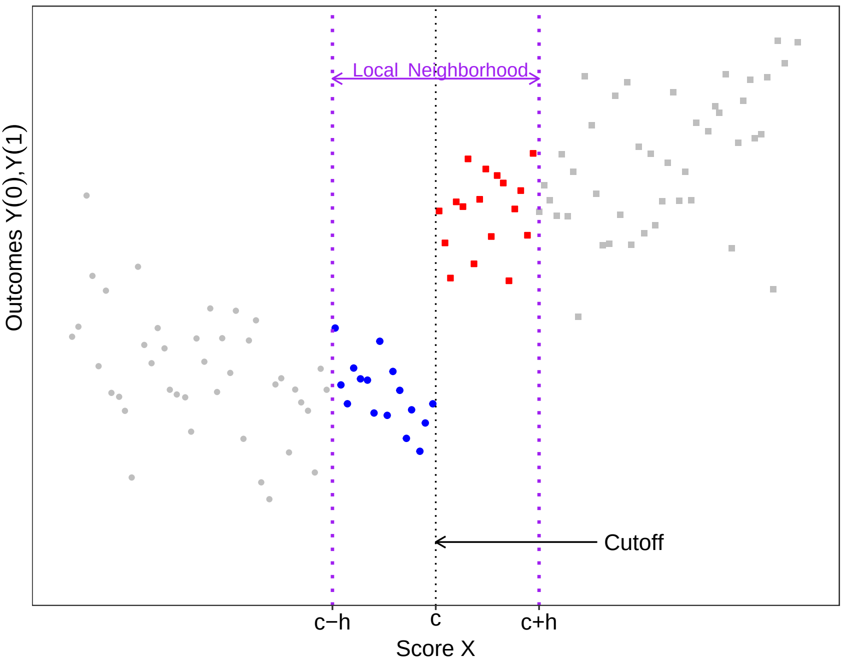
Local Neighborhood
1
0
Kernel Weights
Epanechnikov
Uniform
Triangular
c − h
c
c + h
Score X



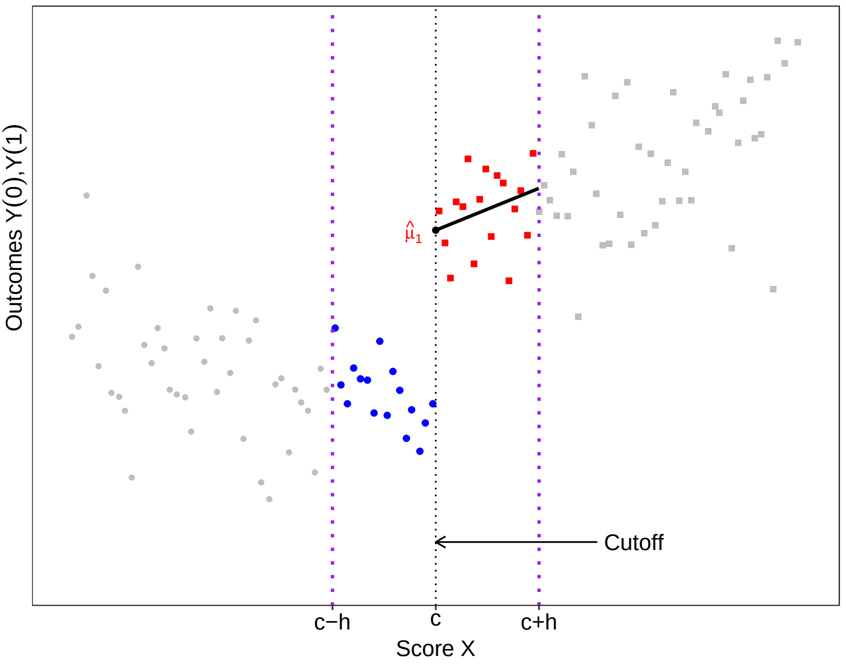
E[Y(1)|X]
μ+
E[Y(0)|X]
μ−
Cutoff
c
Score X
Outcomes Y(0),Y(1)


E[Y(1)|X]
E[Y(0)|X]
μ_+
μ_-
Cutoff
c
Score X
Outcomes Y(0),Y(1)

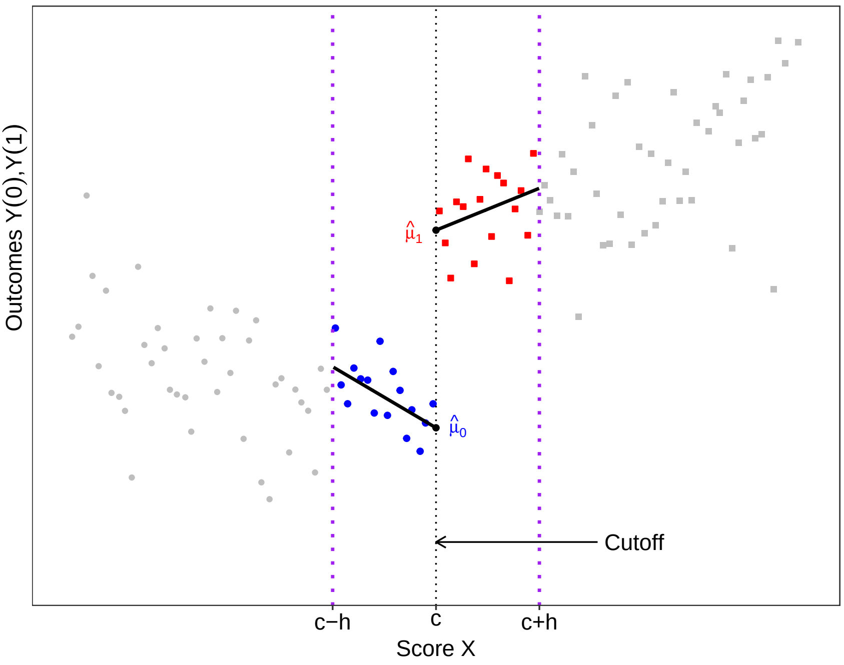

Outcomes Y(0),Y(1)
Cutoff
c
Score X



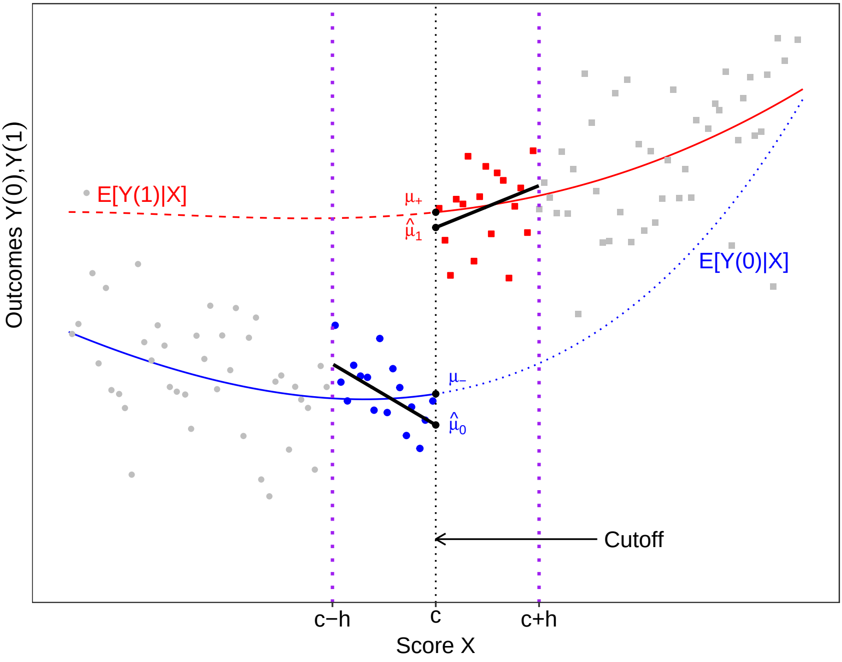
Local Neighborhood
Cutoff
Outcomes Y(0),Y(1)
c−h
c
c+h
Score X



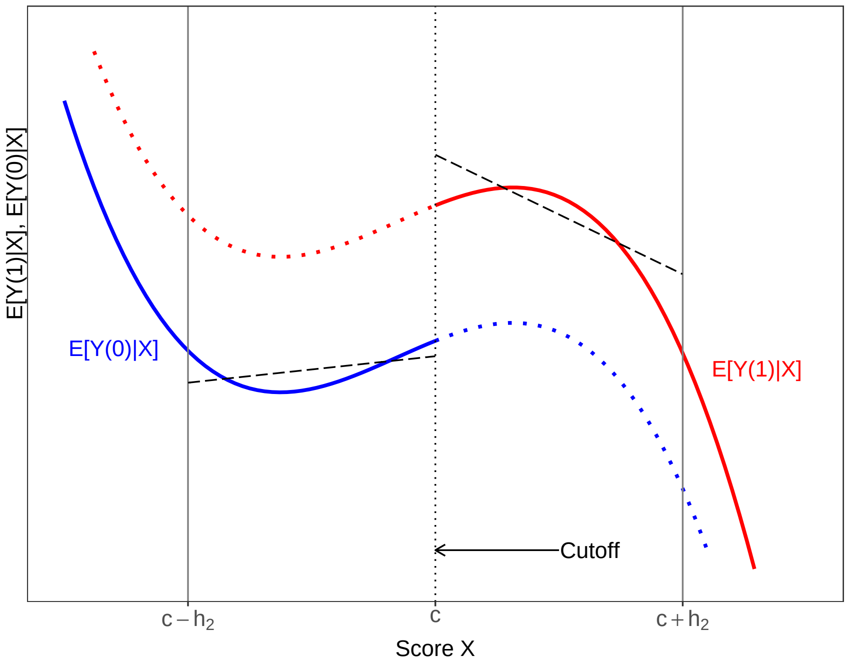
$\hat{\mu}_1$
Cutoff
c−h
c
c+h
Score X
Outcomes Y(0),Y(1)



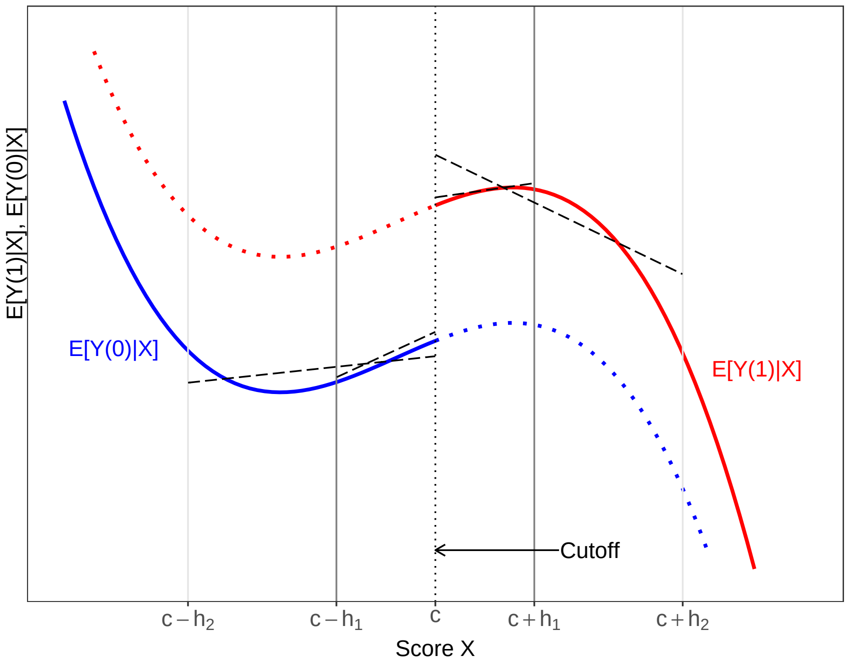
Outcomes Y(0),Y(1)
$\hat{\mu}_1$
$\hat{\mu}_0$
Cutoff
c−h
c
c+h
Score X



E[Y(1)|X]
E[Y(0)|X]
μ_+
$\hat{\mu}_1$
μ_-
$\hat{\mu}_0$
Cutoff
Outcomes Y(0), Y(1)
c−h
c
c+h
Score X



E[Y(1)|X], E[Y(0)|X]
E[Y(0)|X]
E[Y(1)|X]
Cutoff
c − h2
c
c + h2
Score X



E[Y(1)|X], E[Y(0)|X]
E[Y(0)|X]
E[Y(1)|X]
Cutoff
c − h2
c − h1
c
c + h1
c + h2
Score X

Local Polynomial Methods: Choosing bandwidth ($p = 1$)

- Mean Square Error Optimal (MSE-optimal).

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

- Coverage Error Optimal (CE-optimal).

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

- **Key idea**:
  - Trade-off bias and variance of $\hat{\tau}_{\mathtt{SRD}}(h)$. Heuristically:

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

  - Implementations: IK first-generation while CCT second-generation plug-in rule. They differ in the way $\mathsf{Var}(\hat{\tau}_{\mathtt{SRD}})$ and $\mathsf{Bias}(\hat{\tau}_{\mathtt{SRD}})$ are estimated.
  - Rule-of-thumb: $h_{\mathtt{CE}} \propto n^{1/20} \cdot h_{\mathtt{MSE}}$.

Conventional Inference Approach

- "Local-linear" ($p = 1$) estimator (w/ weights $K(\cdot)$):

$$-h \le X_i < c : \qquad\qquad c \le X_i \le h :$$

$$Y_i = \alpha_- + (X_i - c) \cdot \beta_- + \varepsilon_{-,i} \qquad\Bigg|\qquad Y_i = \alpha_+ + (X_i - c) \cdot \beta_+ + \varepsilon_{+,i}$$

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

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

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

- Naïve 95% Confidence interval:

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

Robust Bias Correction Approach

- **Key Problem**:

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

  - B captures bias due to misspecification error.

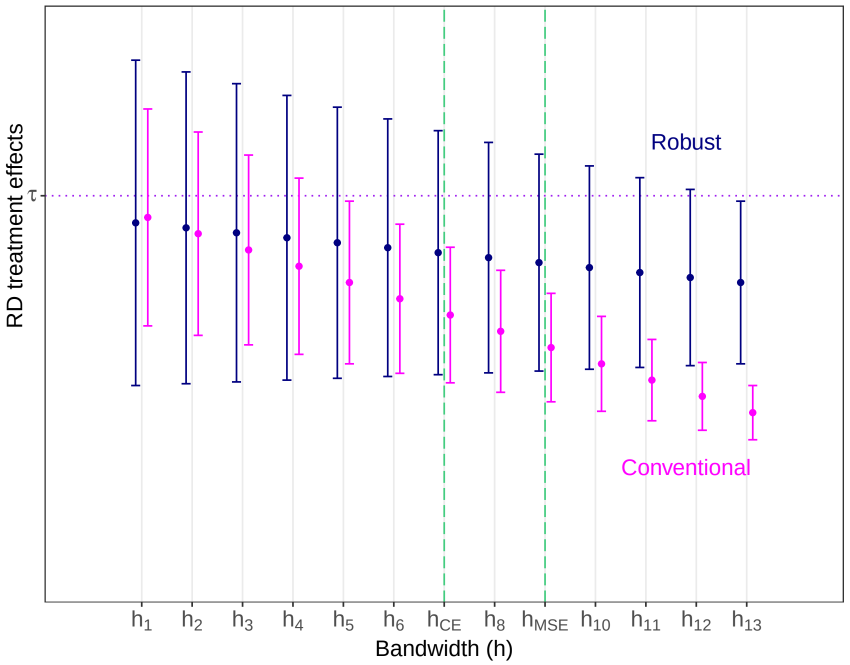
- **RBC distributional approximation**:

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

  - $\hat{\mathsf{B}}$ is constructed to estimate leading bias B, that is, misspecification error.

- **RBC** 95% **Confidence Interval**:

$$I_{\texttt{RBC}} = \left[\ \left(\hat{\tau}_{\texttt{SRD}} - \hat{\mathsf{B}}\right) \ \pm \ 1.96 \cdot \sqrt{\hat{\mathsf{V}} + \hat{\mathsf{W}}} \ \right]$$



RD treatment effects
τ
Robust
Conventional
h1 h2 h3 h4 h5 h6 hCE h8 hMSE h10 h11 h12 h13
Bandwidth (h)

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 $\quad (c = 59.1984)$

  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**:

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

TABLE III
Regression Discontinuity Estimates of the Effect of Head Start Assistance on Mortality

Variable	Control mean	Nonparametric estimator			Parametric: Flexible linear	Parametric: Flexible quadratic
Bandwidth or poverty range		9	18	36	8	16
Number of observations (counties) with nonzero weight		527	961	2,177	484	863
Main results						
Ages 5–9, Head Start-related causes, 1973–1983	3.238	−1.895**	−1.198*	−1.114**	−2.201**	−2.558**
		(0.980)	(0.796)	(0.544)	(1.004)	(1.261)
		[0.036]	[0.081]	[0.027]	[0.022]	[0.021]
Specification checks						
Ages 5–9, injuries, 1973–1983	22.303	0.195	2.426	0.679	−0.164	0.775
		(3.472)	(2.476)	(1.785)	(3.380)	(3.401)
		[0.924]	[0.345]	[0.755]	[0.998]	[0.835]
Ages 5–9, all causes, 1973–1983	40.232	−3.416	0.053	−1.537	−3.896	−2.927
		(4.311)	(3.098)	(2.253)	(4.268)	(4.295)
		[0.415]	[0.982]	[0.558]	[0.317]	[0.505]
Ages 25+, Head Start-related causes, 1973–1983	131.825	2.204	6.016	5.872	2.091	2.574
		(5.719)	(4.349)	(3.338)	(5.581)	(6.415)
		[0.700]	[0.147]	[0.114]	[0.749]	[0.689]

Local Randomization Framework

- **Key idea**: treatment assignment as-if randomly assigned "near" cutoff. There exists window $\mathcal{W} = [-w, w]$, with $-w < c < w$, such that

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

  and possibly other conditions hold (e.g., knowledge of assignment mechanism).

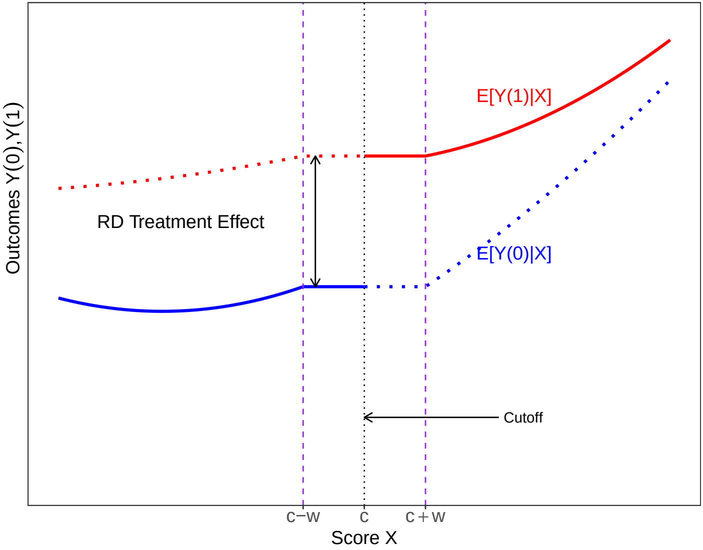
    - Conceptually different from continuity/extrapolation based methods.
    - Challenge: window (neighborhood) selection.
    - Challenge: small sample (estimation and) inference.

- **Two Steps** (analogous to local polynomial methods):

    - Given window $\mathcal{W}$, (estimation and) inference is "standard": superpopulation, large-samples designed-based methods, randomization inference methods.
    - Select windows $\mathcal{W}$ based on idea of local randomization.
    - As-if randomly assigned assumption can be relaxed somewhat, but it is strong.

- **Catch**: as-if random assumption good approximation *only very near cutoff*!

$$T_i \text{ independent of } (Y_i(0), Y_i(1)) \text{ for all } X_i \in \mathcal{W} = [c-w, c+w]$$
$$+ \text{ exclusion restriction}$$



Local Randomization: Finite-sample and Large-sample Methods

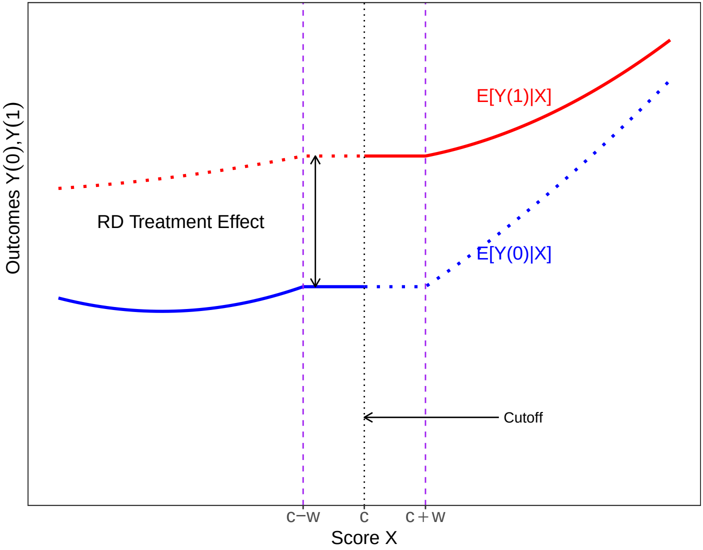
- Given $\mathcal{W}$ where local randomization holds:
  - Randomization inference (Fisher): sharp null, finite-sample exact.
  - Design-based (Neyman): large-sample valid, conservative.
  - Large-sample standard: random potential outcomes, large-sample valid.

- All methods require window ($\mathcal{W}$) selection, and choice of statistic.
  First two also require choice/assumptions assignment mechanism.
  Covariate-adjustments (score or otherwise) possible.

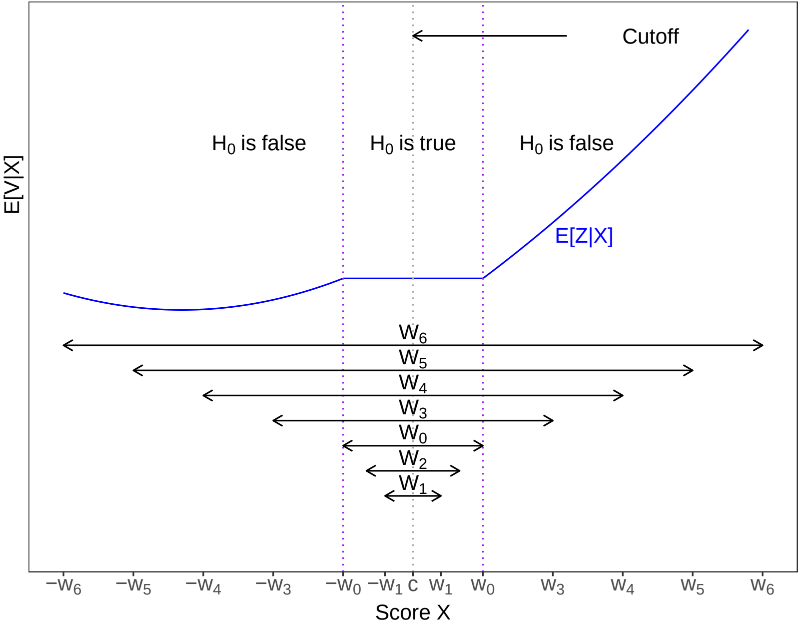
- $\mathcal{W}$ selection:
  - Find neighborhood where (pre-intervention) covariate-balance holds.
  - Find neighborhood where outcome and score independent.
  - Domain-specific or application-specific choice.

$$T_i \text{ independent of } (Y_i(0), Y_i(1)) \text{ for all } X_i \in \mathcal{W} = [c-w, c+w]$$

$$+ \text{ exclusion restriction}$$



Choosing W_0 using predetermined covariate Z



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 $\quad (c = 59.1984)$

  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**:

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

TABLE III

Regression Discontinuity Estimates of the Effect of Head Start Assistance on Mortality

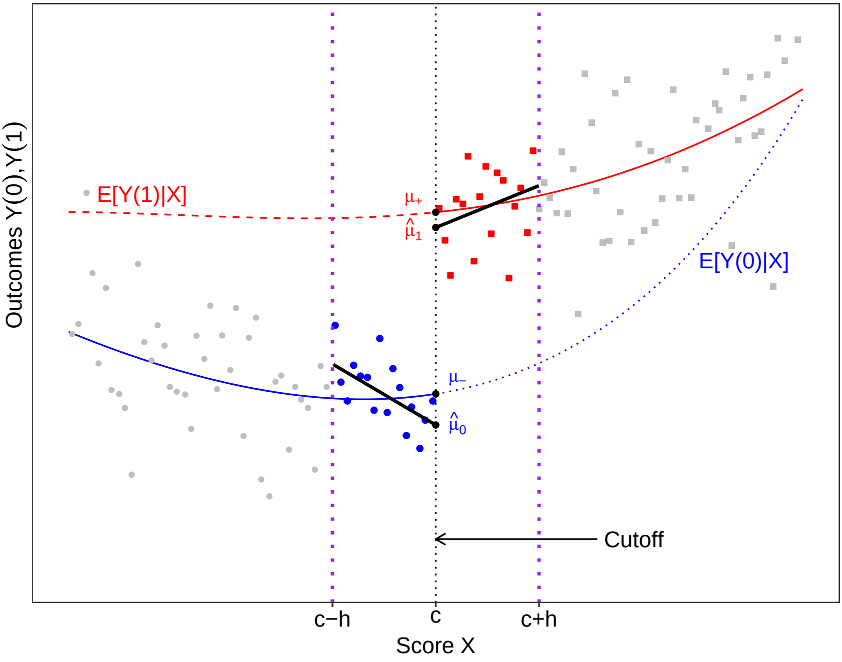
Variable	Control mean	Nonparametric estimator			Parametric: Flexible linear	Parametric: Flexible quadratic
Bandwidth or poverty range		9	18	36	8	16
Number of observations (counties) with nonzero weight		527	961	2,177	484	863
Main results						
Ages 5–9, Head Start-related causes, 1973–1983	3.238	−1.895**	−1.198*	−1.114**	−2.201**	−2.558**
		(0.980)	(0.796)	(0.544)	(1.004)	(1.261)
		[0.036]	[0.081]	[0.027]	[0.022]	[0.021]
Specification checks						
Ages 5–9, injuries, 1973–1983	22.303	0.195	2.426	0.679	−0.164	0.775
		(3.472)	(2.476)	(1.785)	(3.380)	(3.401)
		[0.924]	[0.345]	[0.755]	[0.998]	[0.835]
Ages 5–9, all causes, 1973–1983	40.232	−3.416	0.053	−1.537	−3.896	−2.927
		(4.311)	(3.098)	(2.253)	(4.268)	(4.295)
		[0.415]	[0.982]	[0.558]	[0.317]	[0.505]
Ages 25+, Head Start-related causes, 1973–1983	131.825	2.204	6.016	5.872	2.091	2.574
		(5.719)	(4.349)	(3.338)	(5.581)	(6.415)
		[0.700]	[0.147]	[0.114]	[0.749]	[0.689]

Outline

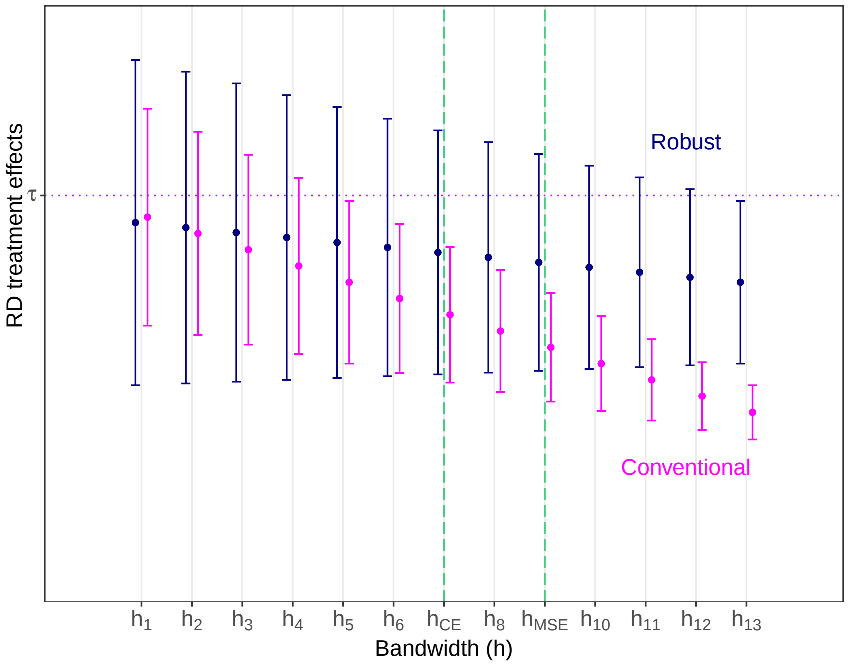
Falsification and Validation

- **RD plots and related graphical methods**:
  - Always plot data: main advantage of RD designs. (Check if RD design!)
  - Plot histogram of X_i (score) and its density. Careful: boundary bias.
  - RD plot $\mathbb{E}[Y_i|X_i = x]$ (outcome) and $\mathbb{E}[Z_i|X_i = x]$ (pre-intervention covariates).
  - Be careful not to oversmooth data/plots.

- **Sensitivity and related methods**:
  - Score density continuity: binomial test and continuity test.
  - Pre-intervention covariate no-effect (covariate balance).
  - Placebo outcomes no-effect.
  - Placebo cutoffs no-effect: informal continuity test away from c.
  - Donut hole: testing for outliers/leverage near c.
  - Different bandwidths: testing for misspecification error.
  - Many other setting-specific (fuzzy, geographic, etc.).



E[Y(1)|X]
E[Y(0)|X]
μ+
$\hat{\mu}_1$
μ−
$\hat{\mu}_0$
Cutoff
c−h
c
c+h
Score X
Outcomes Y(0), Y(1)



RD treatment effects
τ
Robust
Conventional
h_1 h_2 h_3 h_4 h_5 h_6 h_{CE} h_8 h_{MSE} h_{10} h_{11} h_{12} h_{13}
Bandwidth (h)

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 $\quad (c = 59.1984)$

  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

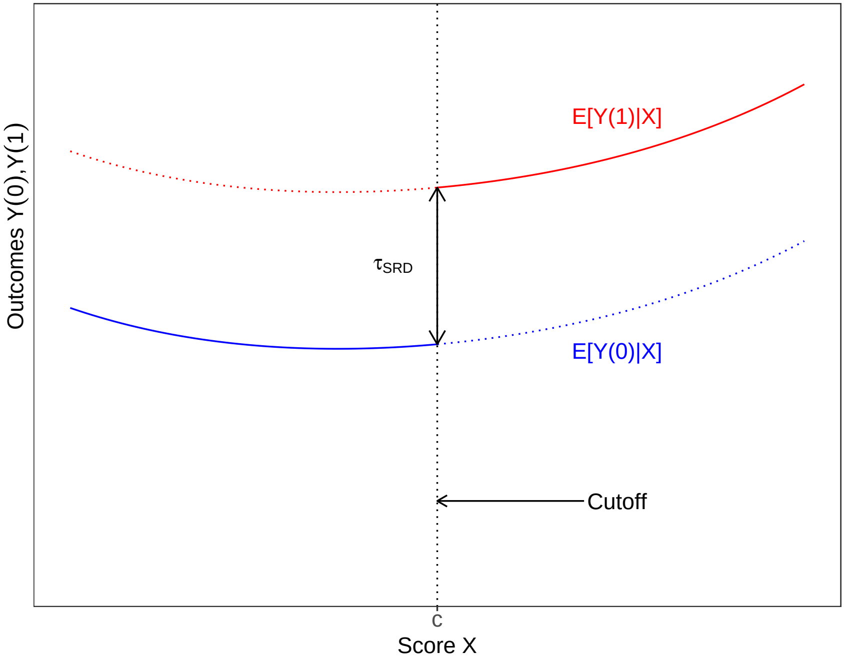
- **Causal Inference**:

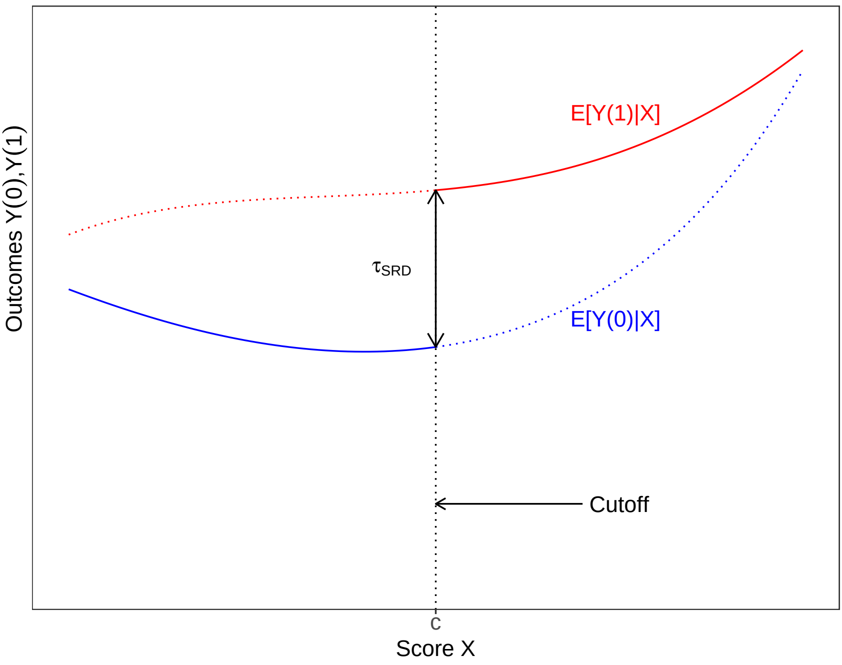
$$Y_i(0) \quad \neq \quad Y_i | T_i = 0 \qquad\qquad \text{and} \qquad\qquad Y_i(1) \quad \neq \quad Y_i | T_i = 1$$

Outline

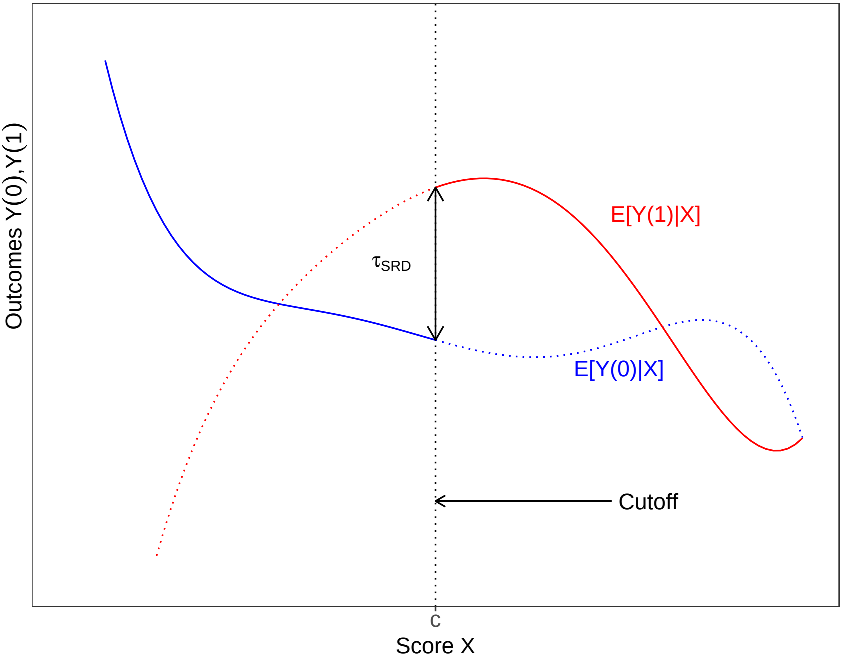
RD Effects Away from the Cutoff

- RD designs are credible, robust and easy to use.
- **Main drawback**: identification is "local" (not even causal if strict!).
- Ongoing research: How to extrapolate RD effects away from c?
  - Internal vs. External validity.
- **Available methods**:
  - Marginal effects: changes at the cutoff.
  - Local randomization: effects near the cutoff.
  - Covariate-adjustment local effects: selection-on-observables near the cutoff.
  - Proxy variables: trace-out evolution of outcome away from cutoff.
  - Setting specific: fuzzy RD, multi-cutoff RD, etc.


E[Y(1)|X]
E[Y(0)|X]
τ_{SRD}
Cutoff
Outcomes Y(0),Y(1)
c
Score X



E[Y(1)|X]
E[Y(0)|X]
τ_{SRD}
Cutoff
Outcomes Y(0),Y(1)
c
Score X



Outcomes Y(0),Y(1)
E[Y(1)|X]
τ_{SRD}
E[Y(0)|X]
Cutoff
c
Score X

Multi-Cutoff RD Designs

- 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: $D_i = \mathbb{1}(X_i \geq c)$.
- 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?

Normalizing-and-Pooling Analysis

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

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

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

Difference between TE across subgroups

- Consider two cutoffs $c_0 < c_1$.
- For a given value of X_i, difference in ATEs has two components:
  - Direct effect: impact of moving a person from one cutoff to the other one.
  - Indirect effect: switching cutoffs shifts the distribution of individual characteristics.
- Example:
  - Treatment is giving fertilizer to poor farmers. Cutoffs differ across regions.
  - Direct effect: regions have different soil characteristics, weather, etc.
  - Indirect effect: farmers in one region may be more entrepreneurial, educated, etc.

Difference between TE across subgroups

- Formally:

$$\begin{aligned}
\tau(c_1, c_0) - \tau(c_1, c_1) &= \mathbb{E}[\tau_i | X_i = c_1, C_i = c_0] - \mathbb{E}[\tau_i | X_i = c_1, C_i = c_1] \\
&= \int \underbrace{[\tau(c_1, c_0, u) - \tau(c_1, c_1, u)]}_{\text{direct effect}} f_{U|X,C}(u|c_1, c_0) d\mu \\
&\quad + \int \tau(c_1, c_1, u) \underbrace{[f_{U|X,C}(u|c_1, c_0) - f_{U|X,C}(u|c_1, c_1)]}_{\text{indirect effect}} d\mu
\end{aligned}$$

Extrapolating RD Treatment Effects using Multiple Cutoffs

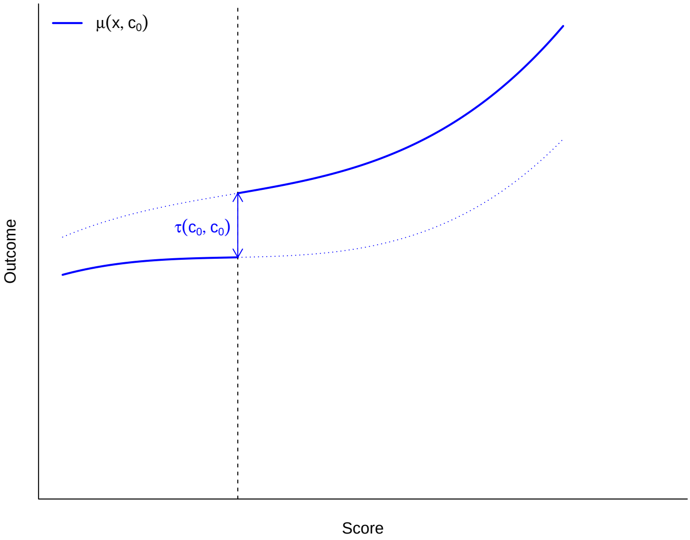
- Two drawbacks of the pooling approach:
  - It discards variation that can identify parameters of interest,
  - Unclear policy relevance: it combines TEs at different cutoffs *for different populations*.
- What are the parameters of interest in this context?
- Potential CEFs:

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

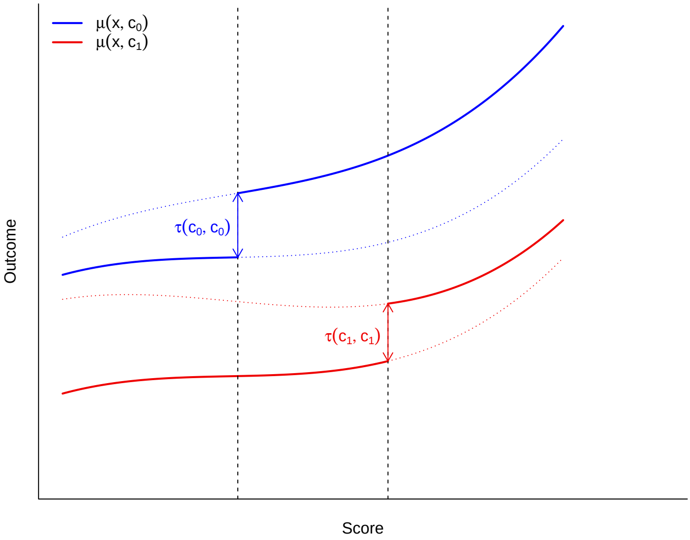
- (Conditional) ATE:

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

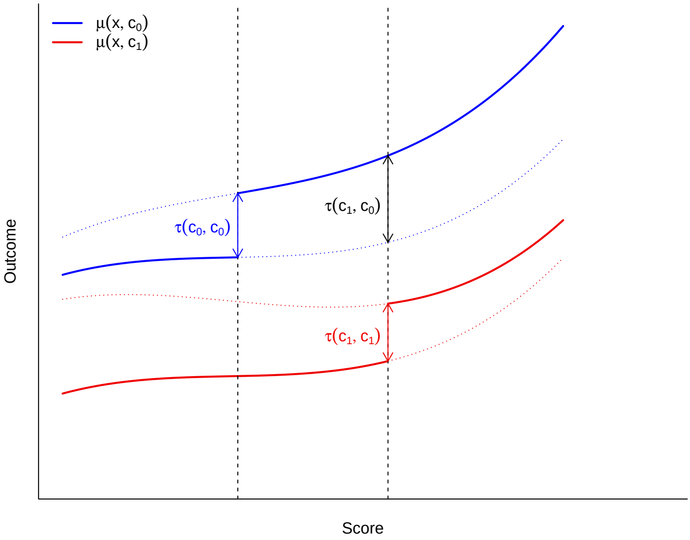
Exploiting multiple cutoffs: parameters of interest



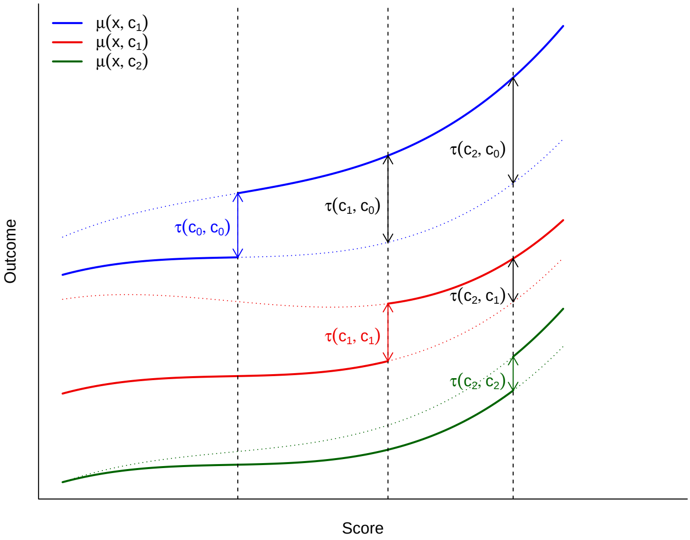
Exploiting multiple cutoffs: parameters of interest



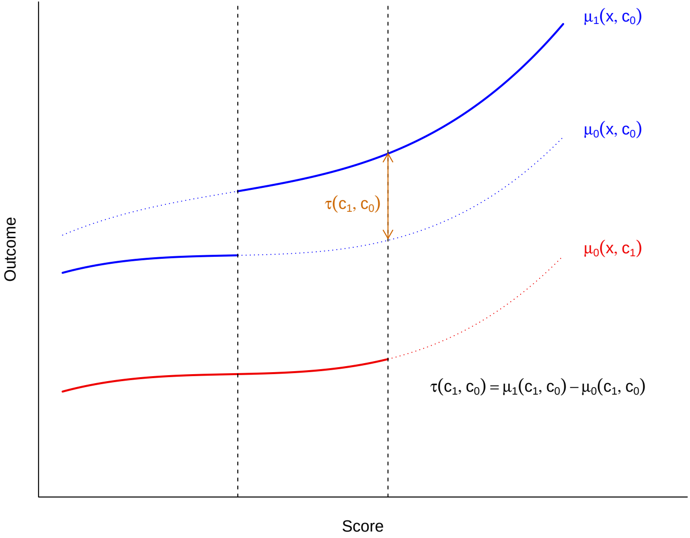
Exploiting multiple cutoffs: parameters of interest



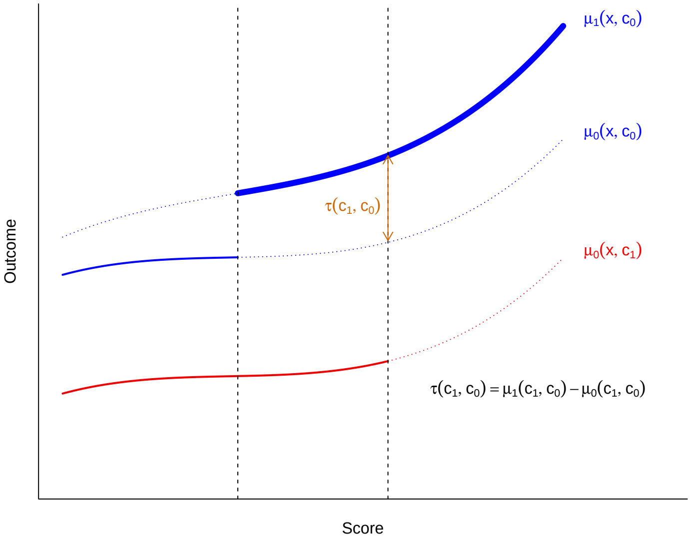
Exploiting multiple cutoffs: parameters of interest



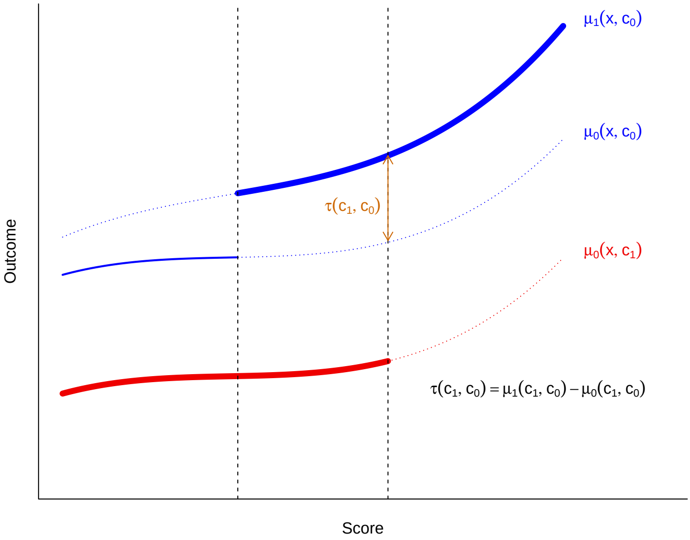
Intuition: two-cutoff case



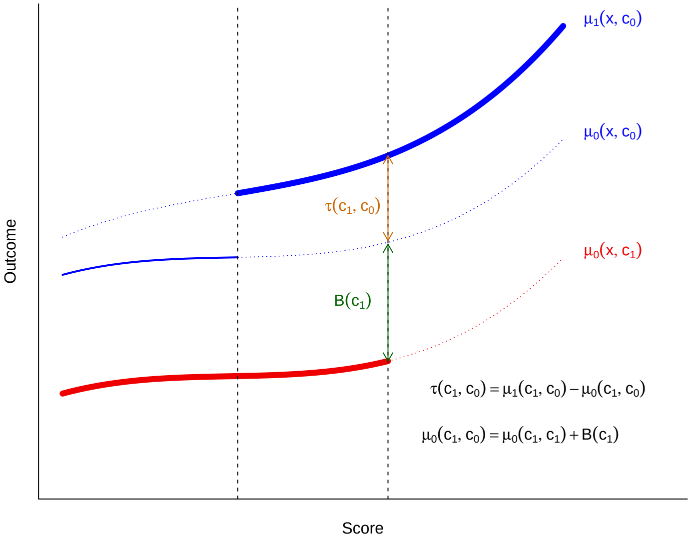
Intuition: two-cutoff case



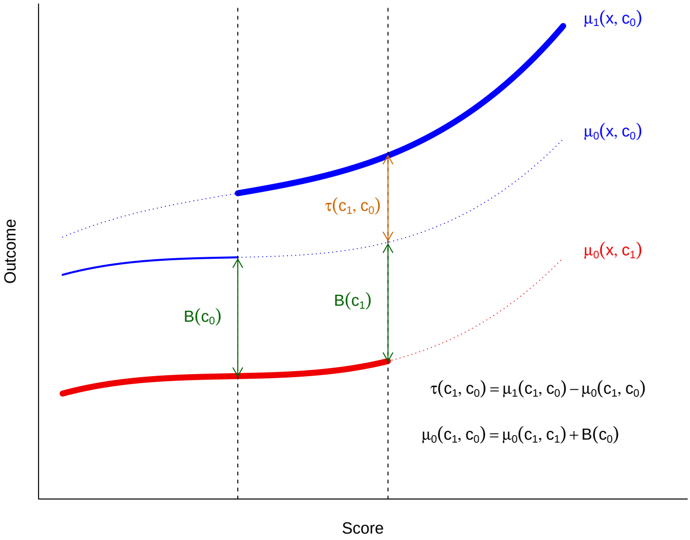
Intuition: two-cutoff case



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.

Thank you!

https://rdpackages.github.io/