Regression Discontinuity Designs

Matias D. Cattaneo and Rocío Titiunik

Princeton University

August 2, 2021

Prepared for the NBER Summer Institute Methods Lectures.
Complementary materials available at `https://rdpackages.github.io/`
NSF support through grants SES-1357561, SES-1459931, SES-1947805, SES-2019432.

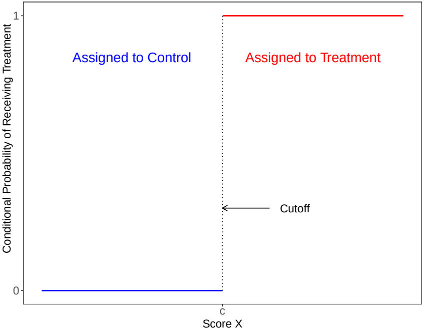
Outline

Causal Inference and Program Evaluation

- 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
  - Easy to falsify and interpret.
  - *Careful*: very local!

Regression Discontinuity Design

- Units receive a **score** (X_i).
- A treatment is assigned based on the score and a *known* **cutoff** (c).
- 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.



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 Interest**.
  - 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

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

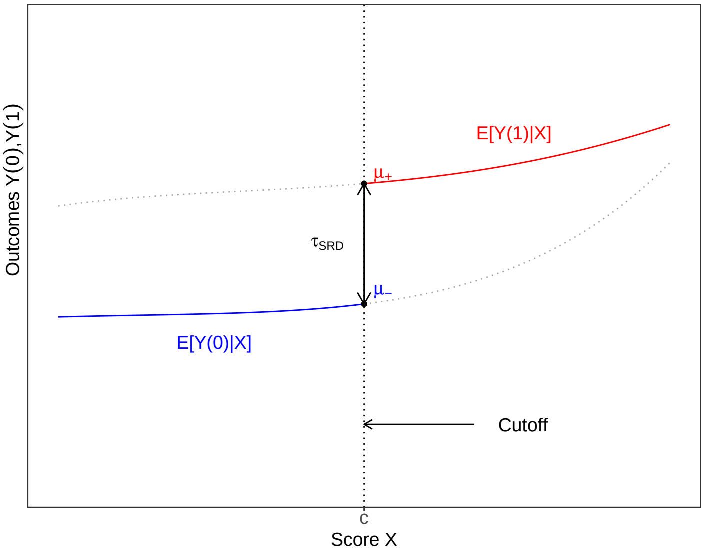
- **Average Treatment Effect at the cutoff** (Continuity-based):

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

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

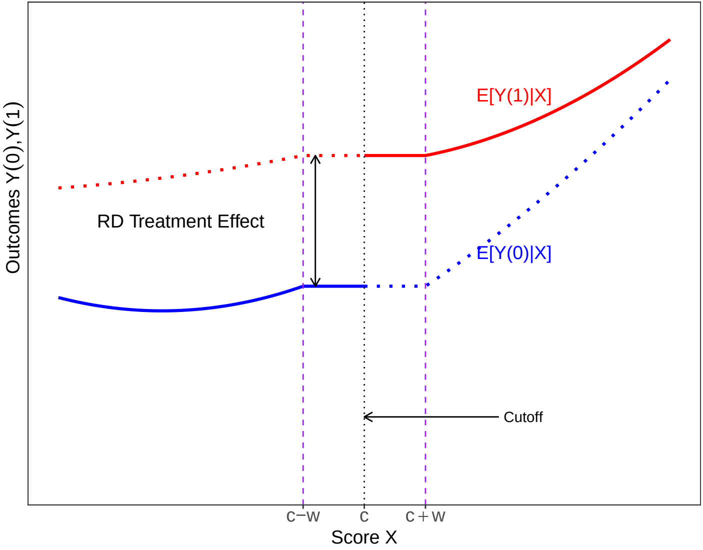
$$\tau_{\texttt{LR}} = \frac{1}{N_{\mathcal{W}}} \sum_{X_i \in \mathcal{W}} \mathbb{E}[Y_i(1) - Y_i(0) | X_i \in \mathcal{W}] = \frac{1}{N_1} \sum_{X_i \in \mathcal{W}, T_i = 1} Y_i - \frac{1}{N_0} \sum_{X_i \in \mathcal{W}, T_i = 0} Y_i$$

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

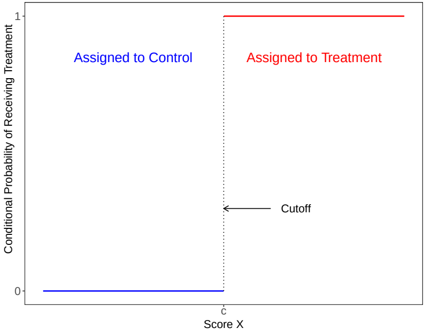


$$T_i \text{ independent of } (Y_i(0), Y_i(1)) \text{ for all } X_i \in \mathcal{W} = [c - w, c + w]$$

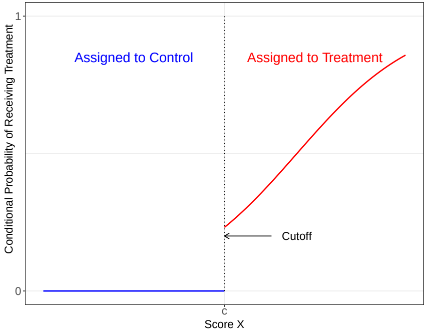
$$+ \text{ exclusion restriction}$$



Fuzzy RD Designs



(a) Sharp RD



(b) Fuzzy RD (one-sided compliance)

Fuzzy RD Designs

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

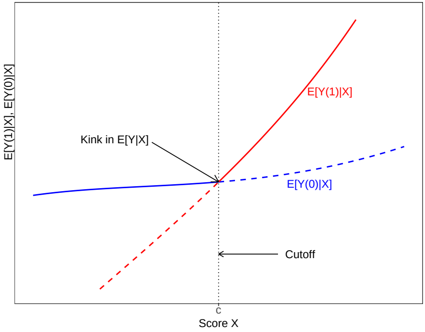
$$\tau_{\mathtt{FRD}} = \frac{\mathbb{E}[(Y_i(1) - Y_i(0)(D_i(1) - D_i(0))|X_i = c]}{\mathbb{E}[D_i(1)|X_i = c] - \mathbb{E}[D_i(0)|X_i = c]}$$

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

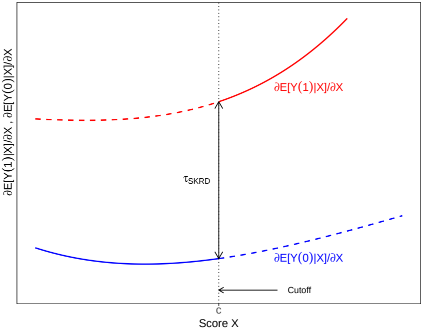
  where $Y_i(t) = Y_i(0)(1 - D_i(t)) + Y_i(1)D_i(t)$ and $D_i(t) = D_i(0)(1 - T_i) + D_i(1)T_i$.

- Similarly for Local Randomization framework.
- Different interpretations under different assumptions.

(Sharp and Fuzzy) Kink RD Designs



(a) Kink RD (levels)



(b) Kink RD (derivatives)

(Sharp and Fuzzy) Kink RD Designs

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

- SKRD Parameter:

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

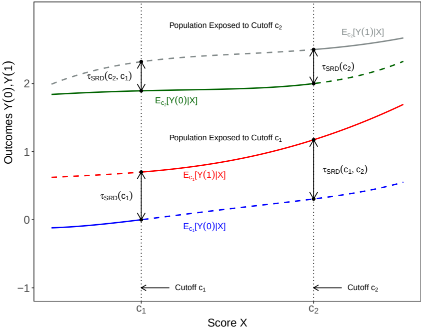
  where $b(x)$ known function inducing "kink".

- FKRD Parameter:

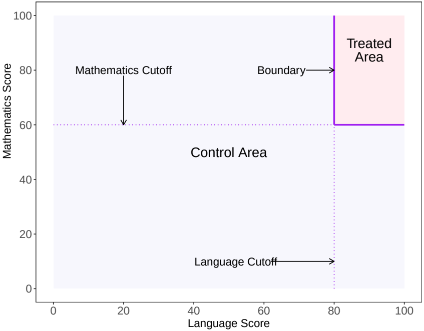
$$\tau_{\texttt{KRD}} = \frac{\lim_{x \downarrow c} \frac{d}{dx}\mathbb{E}[Y_i|X_i = x] - \lim_{x \uparrow c} \frac{d}{dx}\mathbb{E}[Y_i|X_i = x]}{\lim_{x \downarrow c} \frac{d}{dx}\mathbb{E}[D_i|X_i = x] - \lim_{x \uparrow c} \frac{d}{dx}\mathbb{E}[D_i|X_i = x]}$$

- Different interpretation under different assumptions.

Multi-cutoff, Multi-Score, Geographic RD Designs



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

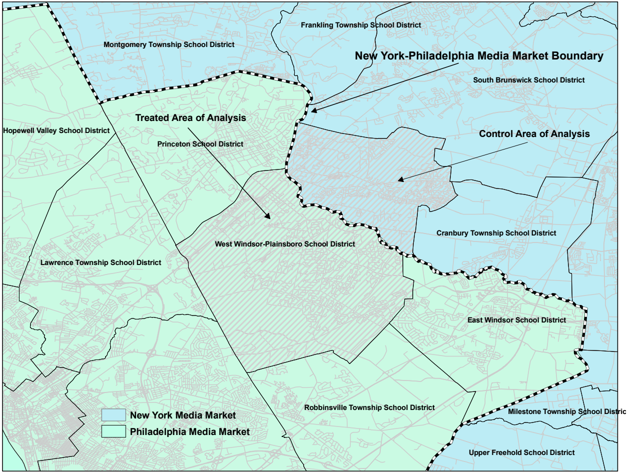


(b) Multi-score:
$\tau_{\text{SRD}}(x_1, x_2) = \mathbb{E}[Y_i(1) - Y_i(0)|X_{1i} = x_1, X_{2i} = x]$

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.

Geographic RD Design



Highlights and Main Takeaways

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

Outline

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.

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**:

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

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

- **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).
- **Continuity/Extrapolation**: Local polynomial approach.
  - Localization: bandwidth selection (trade-off bias and variance).
  - Point estimation: "flexible" (nonparametric).
  - Inference: robust bias-corrected methods.
- 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).

Outline

Local Randomization Approach to RD Design

- **Key assumption**: exists window $\mathcal{W} = [c - w, c + w]$ around cutoff where subjects are as-if randomly assigned to either side of cutoff:

  1. Joint probability distribution of scores for units in the $\mathcal{W}$ is known:

     $$\mathbb{P}[\mathbf{X}_{\mathcal{W}} \leq \mathbf{x}] = F(\mathbf{x}), \qquad \text{for some known joint c.d.f. } F(\mathbf{x}),$$

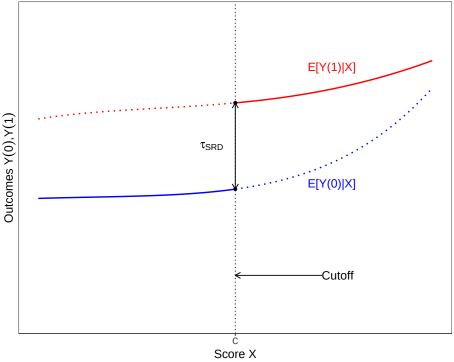
     where $\mathbf{X}_{\mathcal{W}}$ denotes the vector of scores for all i such that $X_i \in \mathcal{W}$.

  2. Potential outcomes not affected by value of the score:

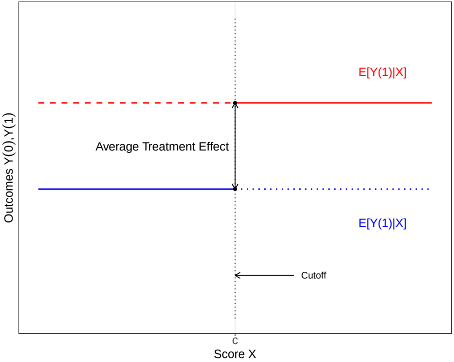
     $$\begin{aligned} Y_i(0, x) &= Y_i(0), \\ Y_i(1, x) &= Y_i(1), \qquad \text{for all } X_i \in \mathcal{W}. \end{aligned}$$

- Note: stronger assumption than continuity-based approach.
  - Potential outcomes are a constant function of the score (can be relaxed).
  - Regression functions are not only continuous at c, but also completely unaffected by the running variable in $\mathcal{W}$.

Experiment versus RD Design

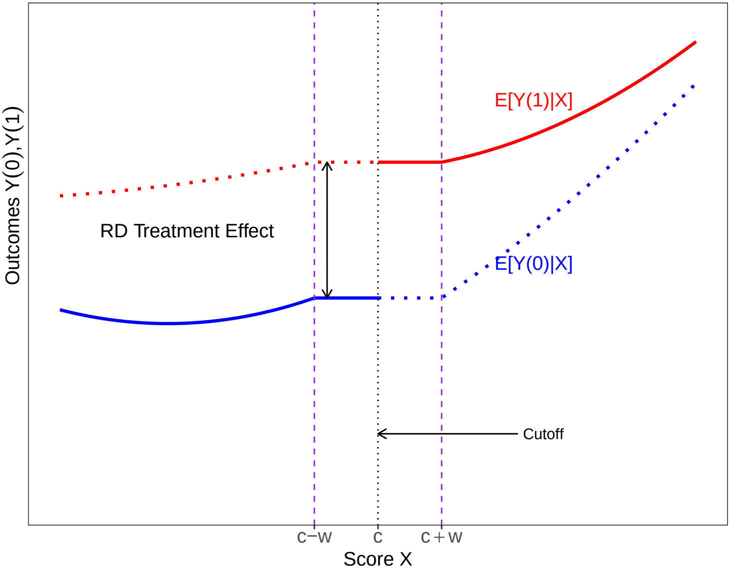


(a) RD Design



(b) Randomized Experiment

Local Randomization RD

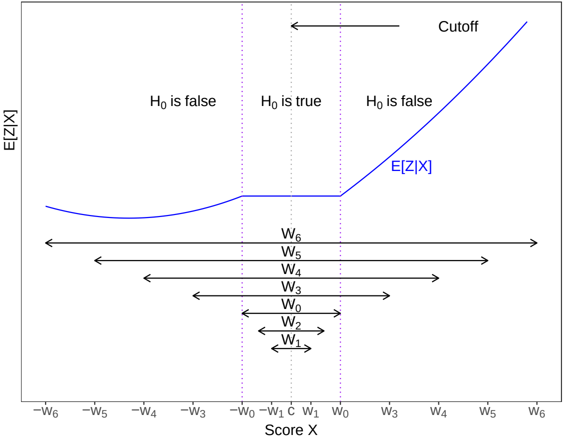


Local Randomization Framework

- **Key idea**: exists window $\mathcal{W} = [c - w, c + w]$ around cutoff where subjects are as-if randomly assigned to either side of cutoff.
- **Two Steps** (analogous to local polynomial methods):
  1. Select window $\mathcal{W}$.
  2. Given window $\mathcal{W}$, perform estimation and inference.
- **Challenges**
  - Window (neighborhood) selection.
  - As-if random assumption good approximation *only very near cutoff*
  - Small sample.

Step 1: Choose the window $\mathcal{W}$

- Find neighborhood where (pre-intervention) covariate-balance holds.
- Find neighborhood where outcome and score independent.
- Domain-specific or application-specific choice.



Step 2: Finite-sample and Large-sample Methods in $\mathcal{W}$

- 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.

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**:

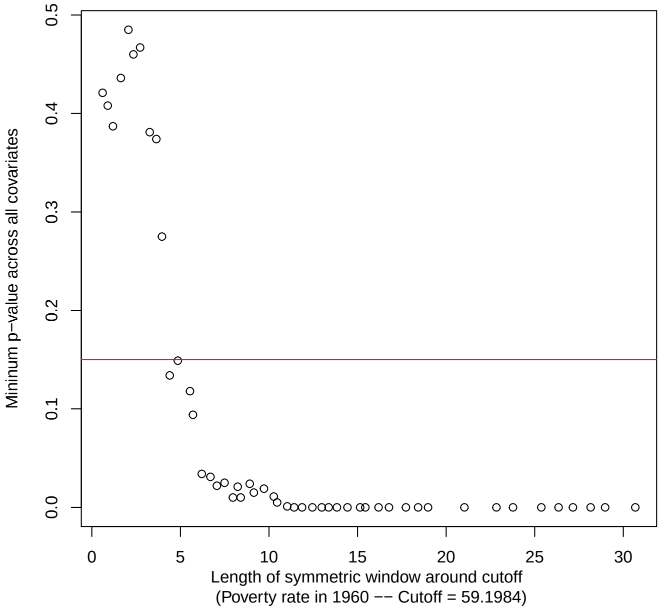
  $Y_i(0) =$ child mortality if **had not received** Head Start

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

- **Causal Inference**:

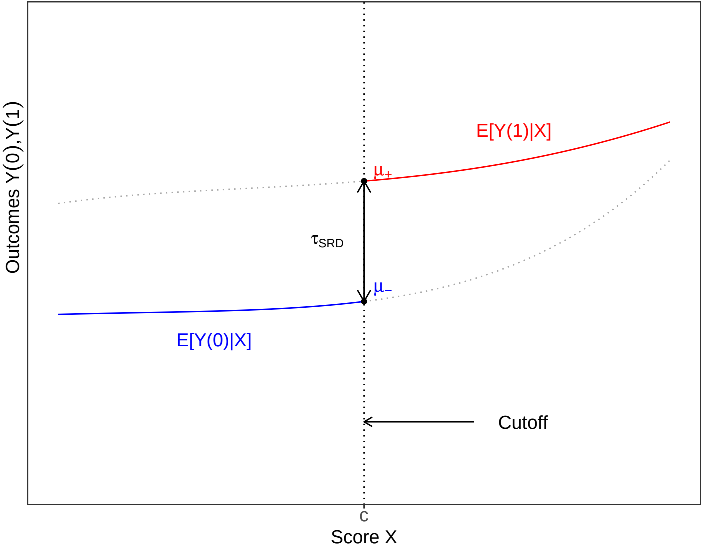
$$Y_i(0) \quad \neq \quad Y_i | T_i = 0 \qquad \text{and} \qquad Y_i(1) \quad \neq \quad Y_i | T_i = 1$$

Empirical Illustration: Window Selector



Outline

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



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 $K(\cdot)$):

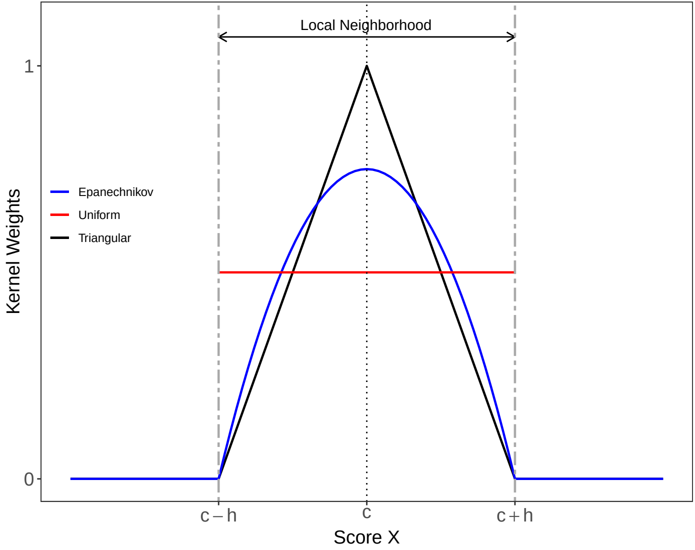
$-h \leq X_i < c$:	$c \leq X_i \leq h$:
$Y_i = \alpha_- + (X_i - c) \cdot \beta_- + \varepsilon_{-,i}$	$Y_i = \alpha_+ + (X_i - c) \cdot \beta_+ + \varepsilon_{+,i}$

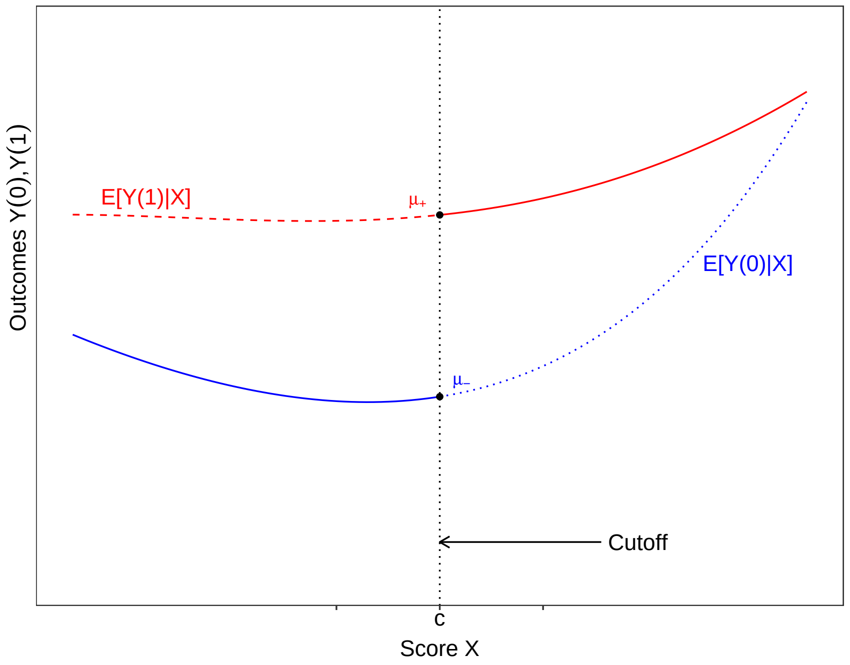
  - Treatment effect (at the cutoff): $\hat{\tau}_{\texttt{SRD}}(h) = \hat{\alpha}_+ - \hat{\alpha}_-$

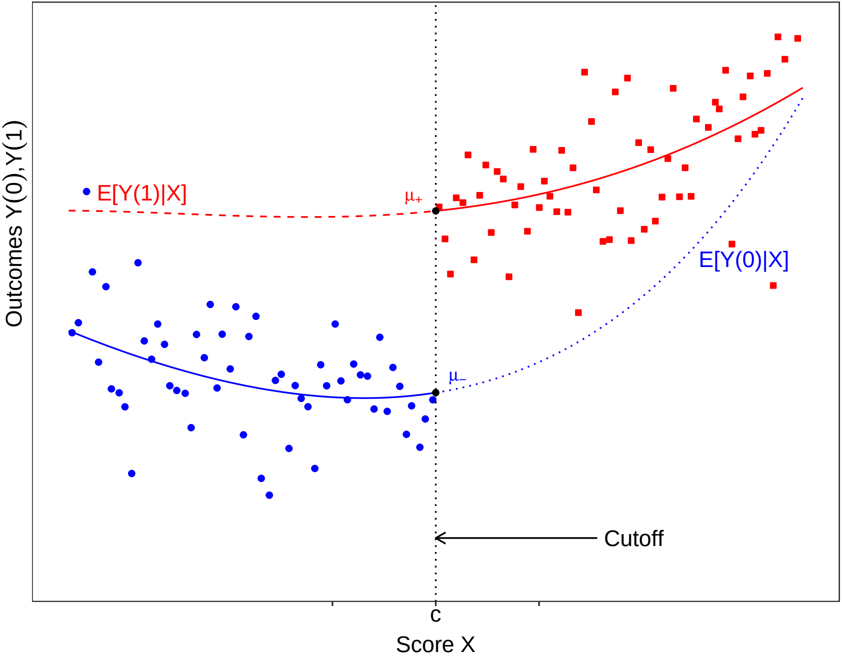
- Can be estimated using linear models (w/ weights $K(\cdot)$):

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

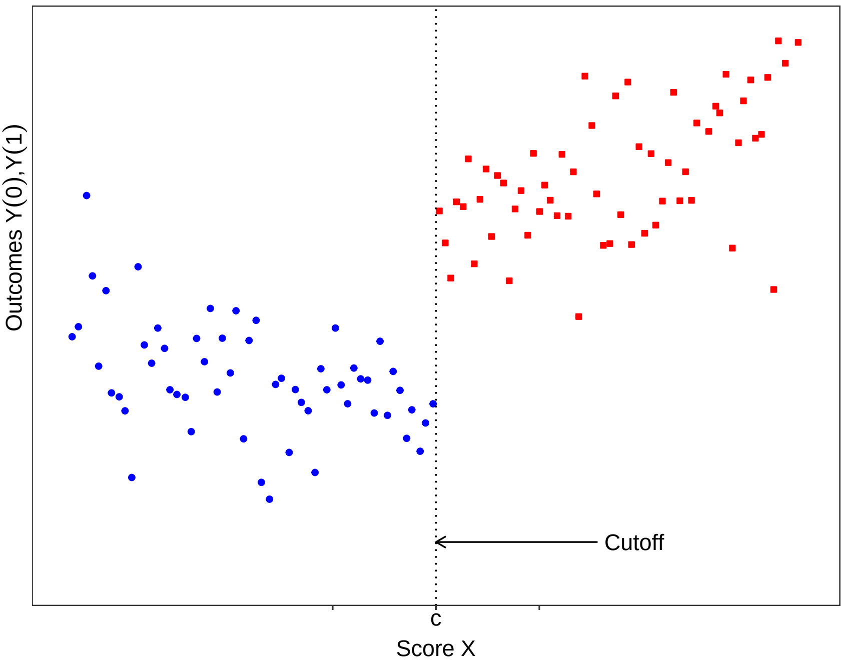
- Given p, K, h chosen $\implies$ weighted least squares estimation.


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

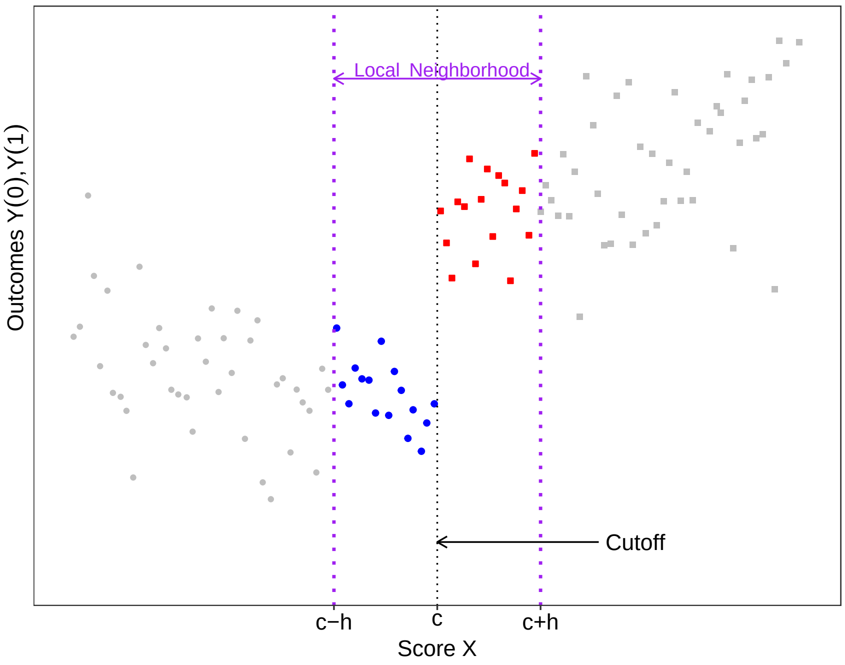

E[Y(1)|X]
μ₊
μ₋
E[Y(0)|X]
Cutoff
Outcomes Y(0),Y(1)
c
Score X



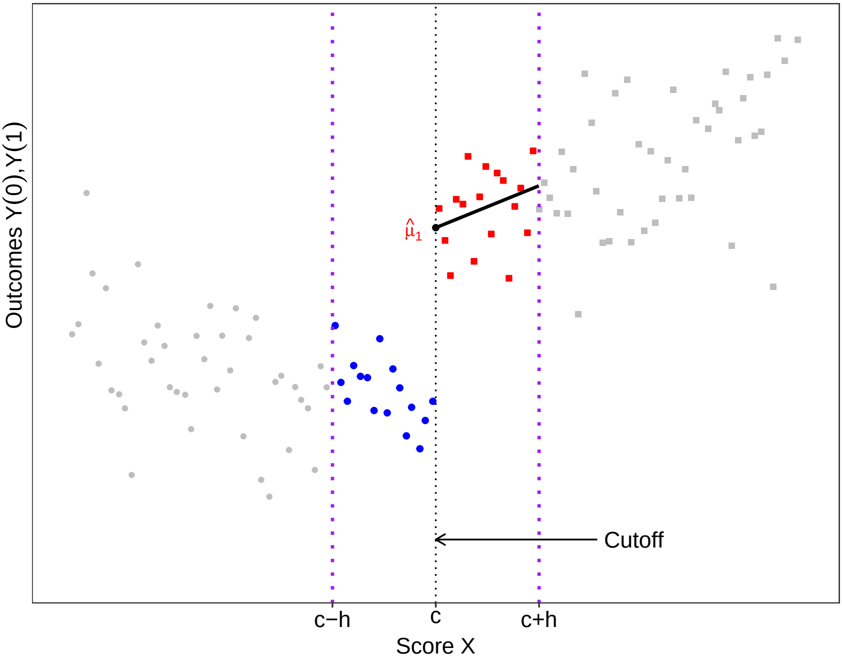
E[Y(1)|X]
E[Y(0)|X]
μ₊
μ₋
Cutoff
Outcomes Y(0),Y(1)
c
Score X



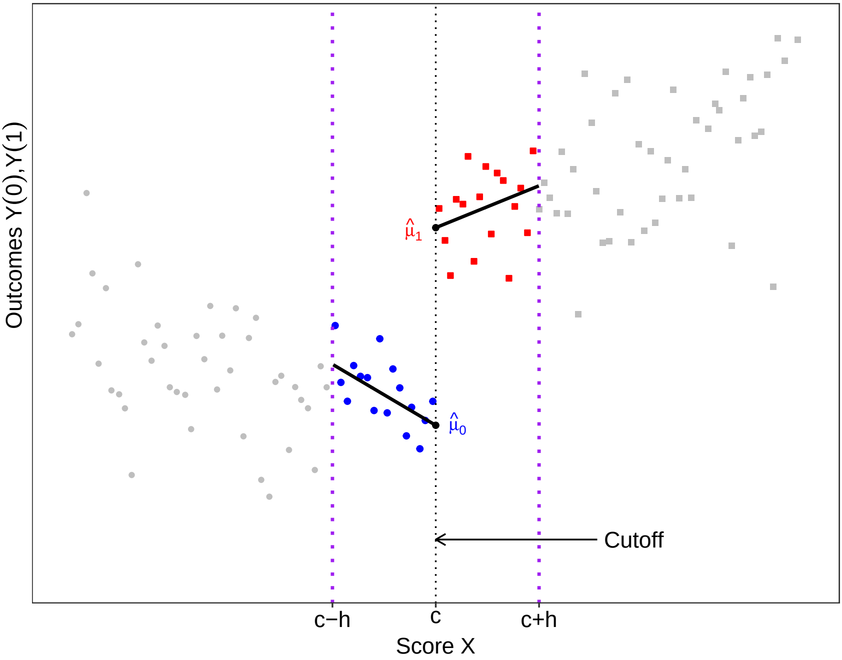
Outcomes Y(0),Y(1)
Cutoff
c
Score X



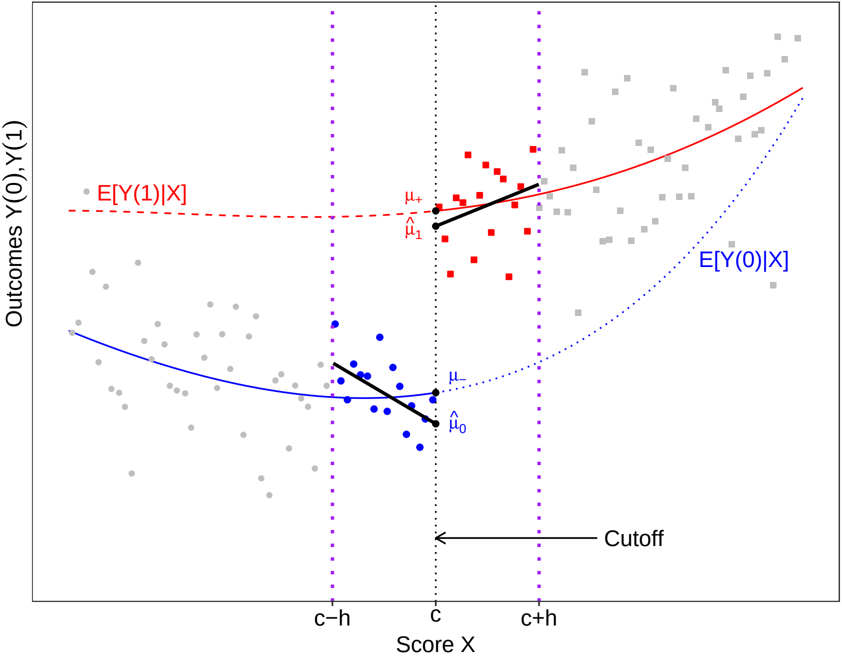
Local Neighborhood
Cutoff
Outcomes Y(0),Y(1)
c−h
c
c+h
Score X

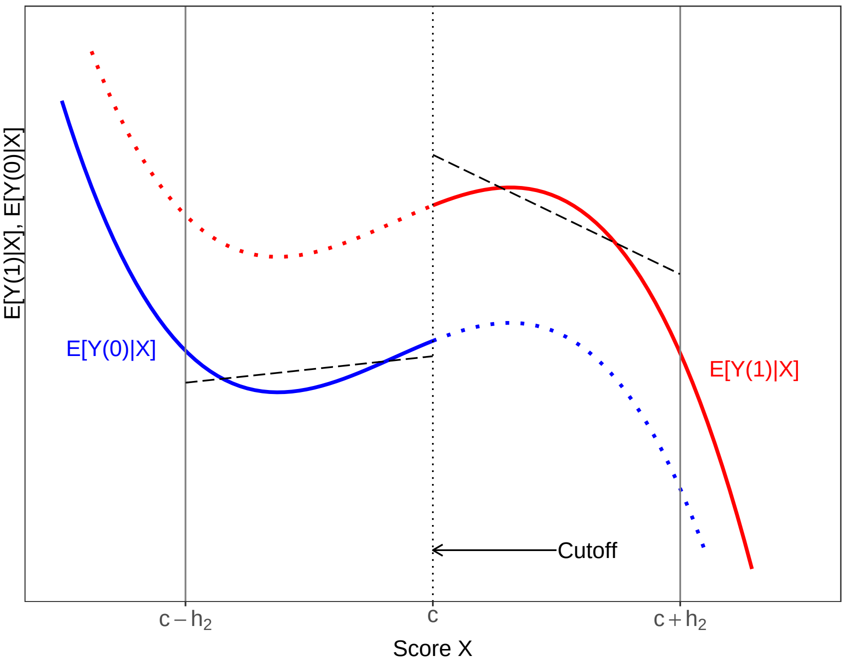


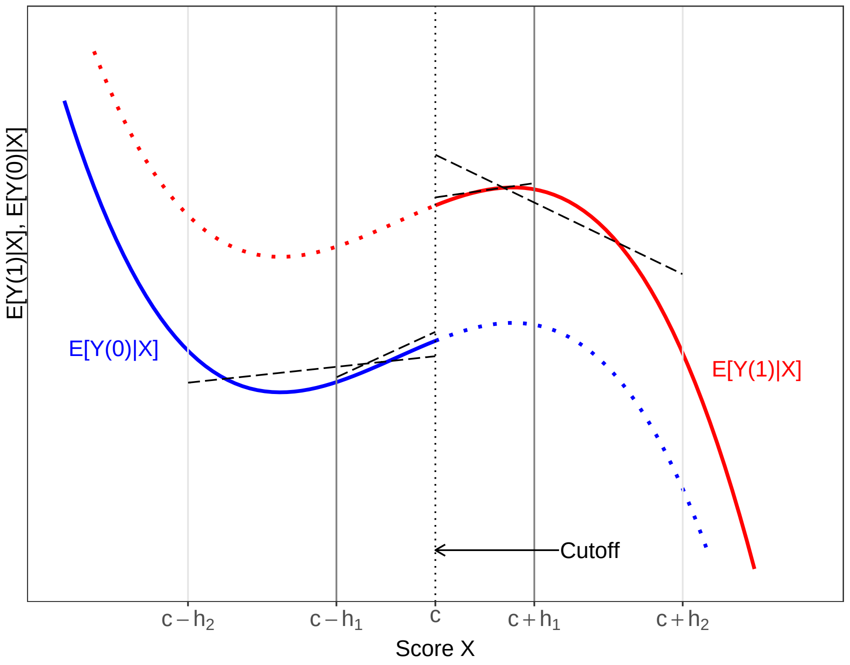
Outcomes Y(0), Y(1)
$\hat{\mu}_1$
Cutoff
c−h
c
c+h
Score X



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


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
$c-h_2$
$c-h_1$
c
$c+h_1$
$c+h_2$
Score X

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

- Mean Square Error Optimal (MSE-optimal).

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

- Coverage Error Optimal (CE-optimal).

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

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

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

  - Implementations: IK first-generation while CCT 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.
  - Rule-of-thumb: $h_{\texttt{CE}} \propto n^{1/20} \cdot h_{\texttt{MSE}}$.

Conventional Inference Approach

- "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): $\hat{\tau}_{\mathtt{SRD}}(h) = \hat{\alpha}_+ - \hat{\alpha}_-$

- 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**:

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

  - B captures bias due to misspecification error.

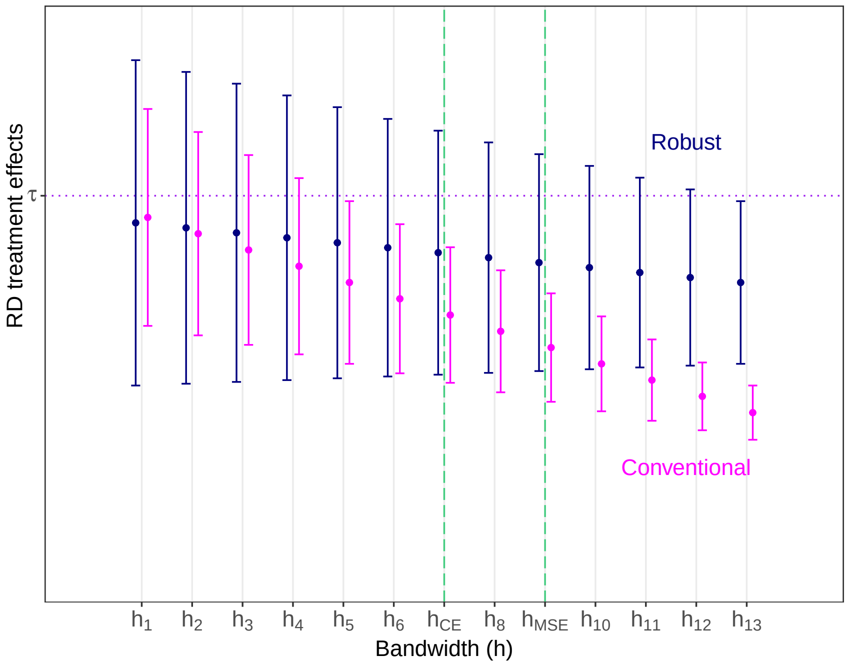
- **RBC distributional approximation**:

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

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

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

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


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**:

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

TABLE III

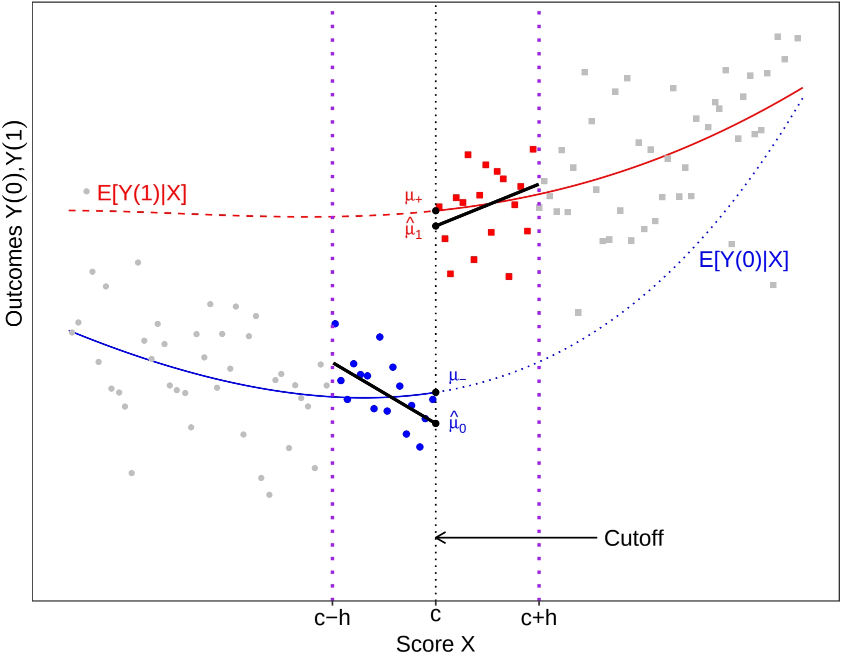
REGRESSION DISCONTINUITY ESTIMATES OF THE EFFECT OF HEAD START ASSISTANCE ON MORTALITY

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]

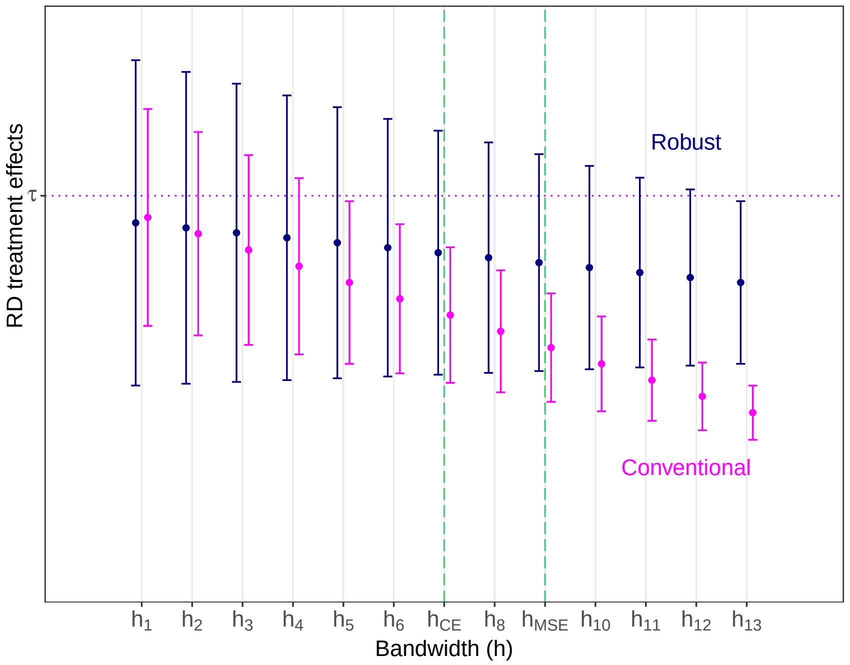
Outline

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**:

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

Thank you!

https://rdpackages.github.io/