

# Tutorial on Regression Discontinuity Designs

Matias D. Cattaneo  
Princeton University

December 8, 2020

# Outline

- ➊ Designs and Frameworks
- ➋ Estimation and Inference
- ➌ Falsification and Validation
- ➍ Extrapolation and Other Topics (if time permits)

# 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!

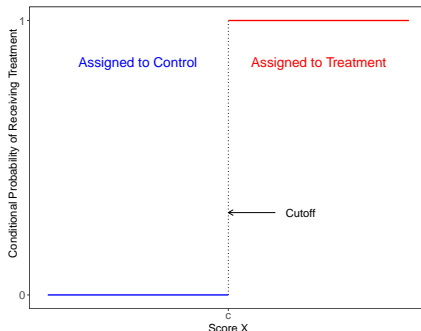
## RD Designs: Building Block

- **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:**  $(Y_i(0), Y_i(1), X_i)$ ,  $i = 1, 2, \dots, n$ .
- **Treatment:**  $T_i \in \{0, 1\}$ ,  $T_i$  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:**

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

## RCTs vs. (Sharp) RD Designs

- **Notation:**  $(Y_i(0), Y_i(1), X_i)$ ,  $i = 1, 2, \dots, n$ ,  $X_i$  score.
- **Treatment:**  $T_i \in \{0, 1\}$ ,  $T_i = \mathbb{1}(X_i \geq c)$ ,  $c$  cutoff.
- **Data:**  $(Y_i, T_i, X_i)$ ,  $i = 1, 2, \dots, n$ , with

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

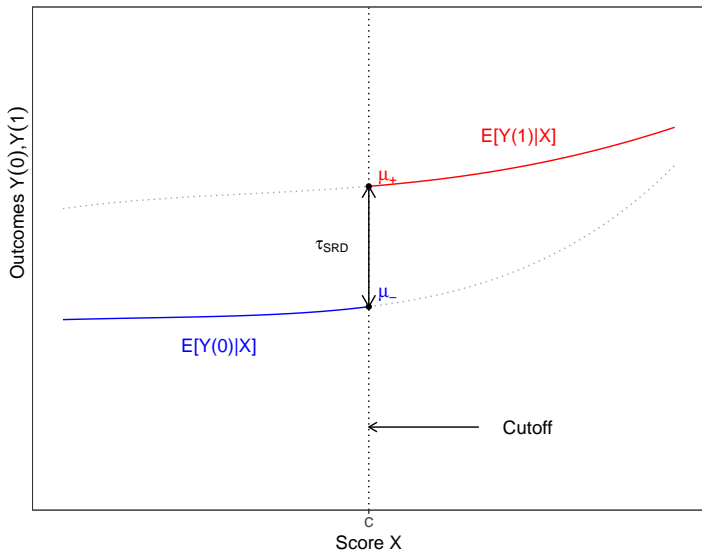
- **Average Treatment Effect at the cutoff** (Continuity-based):

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

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

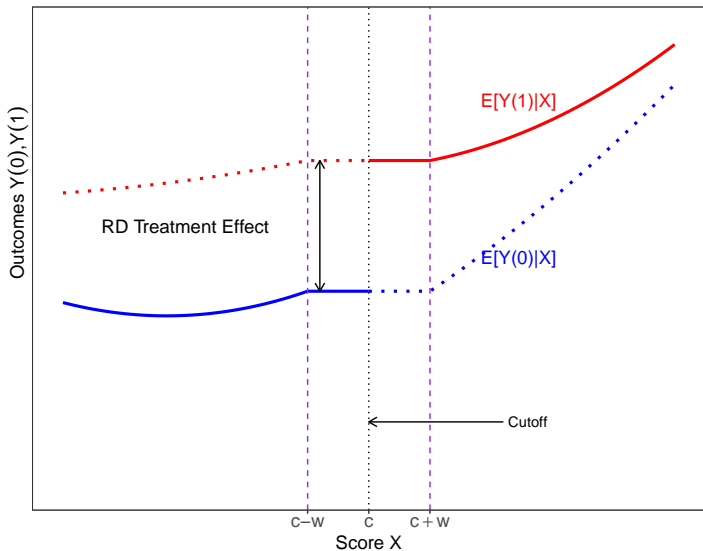
$$\tau_{\text{LR}} = \mathbb{E}[Y_i(1) - Y_i(0) | X_i \in \mathcal{W}] = \frac{1}{N_1} \sum_{X_i \in \mathcal{W}, T_i=1} Y_i - \frac{1}{N_0} \sum_{X_i \in \mathcal{W}, T_i=0} Y_i$$

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





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



- **Imperfect compliance.**

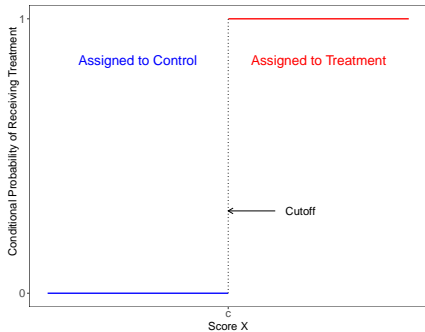
- ▶ probability of receiving treatment changes at  $c$ , but not necessarily from 0 to 1.

- Canonical Parameter:

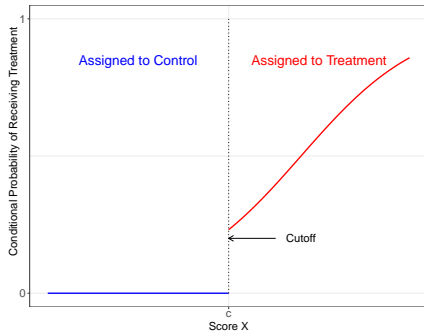
$$\begin{aligned}\tau_{\text{FRD}} &= \frac{\mathbb{E}[(Y_i(1) - Y_i(0))(D_i(1) - D_i(0)) | X_i = c]}{\mathbb{E}[D_i(1) | X_i = c] - \mathbb{E}[D_i(0) | X_i = c]} \\ &= \frac{\lim_{x \downarrow c} \mathbb{E}[Y_i | X_i = x] - \lim_{x \uparrow c} \mathbb{E}[Y_i | X_i = x]}{\lim_{x \downarrow c} \mathbb{E}[D_i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

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:

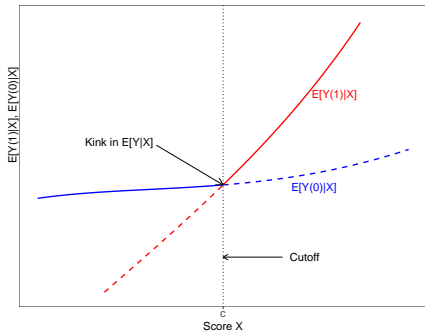
$$\tau_{\text{KRD}} = \frac{\lim_{x \downarrow c} \frac{d}{dx} \mathbb{E}[Y_i | X_i = x] - \lim_{x \uparrow c} \frac{d}{dx} \mathbb{E}[Y_i | X_i = x]}{\lim_{x \downarrow c} \frac{d}{dx} b(x) - \lim_{x \uparrow c} \frac{d}{dx} b(x)}$$

where  $b(x)$  known function inducing “kink”.

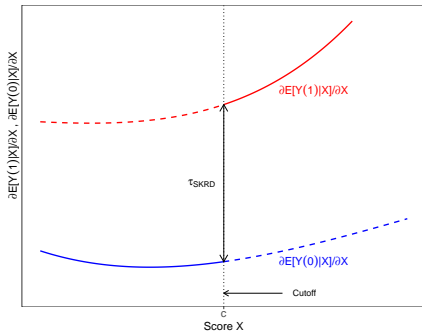
- FKRD Parameter:

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

- Different interpretation under different assumptions.



(a) Kink RD (levels)



(b) Kink RD (derivatives)

# Multi-cutoff, Multi-Score, Geographic RD Designs

- **Multi-cutoff RD designs.**

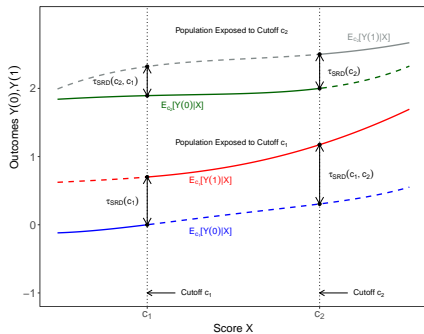
- ▶  $C_i \in \mathcal{C}$  with  $\mathcal{C} = \{c_1, c_2, \dots, c_J\}$  or  $\mathcal{C} = [\underline{c}, \bar{c}]$ .
- ▶ Two strategies: normalize-and-pool ( $\tilde{X}_i = X_i - C_i$ ), or cutoff-by-cutoff analysis.
- ▶ Different interpretation under different assumptions.

- **Multi-score RD designs.**

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

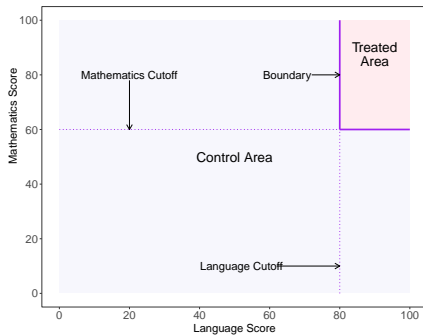
- **Other RD-like designs.**

- ▶ RD in density and bunching designs.
- ▶ RD in time.
- ▶ Dynamic RD designs.
- ▶ etc.



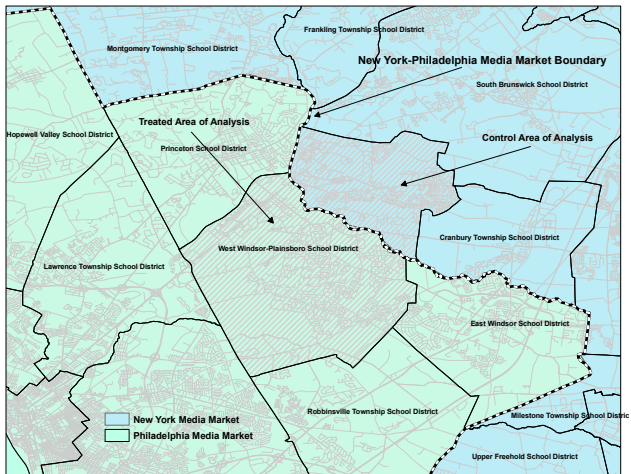
(a) Multi-cutoff:

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



(b) Multi-score:

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





## 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  $\implies$  bandwidth or window selection.
- There exist many design-specific falsification/validation methods.

# Outline

- 1 Designs and Frameworks
- 2 Estimation and Inference
- 3 Falsification and Validation
- 4 Extrapolation and Other Topics (if time permits)

## 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:**

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

# RD Packages

<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.
  - ▶ Detection 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)

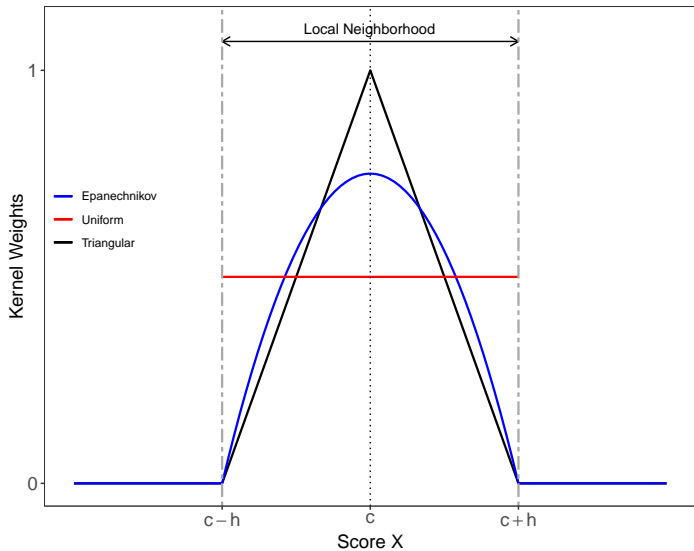
## Local Polynomial Methods

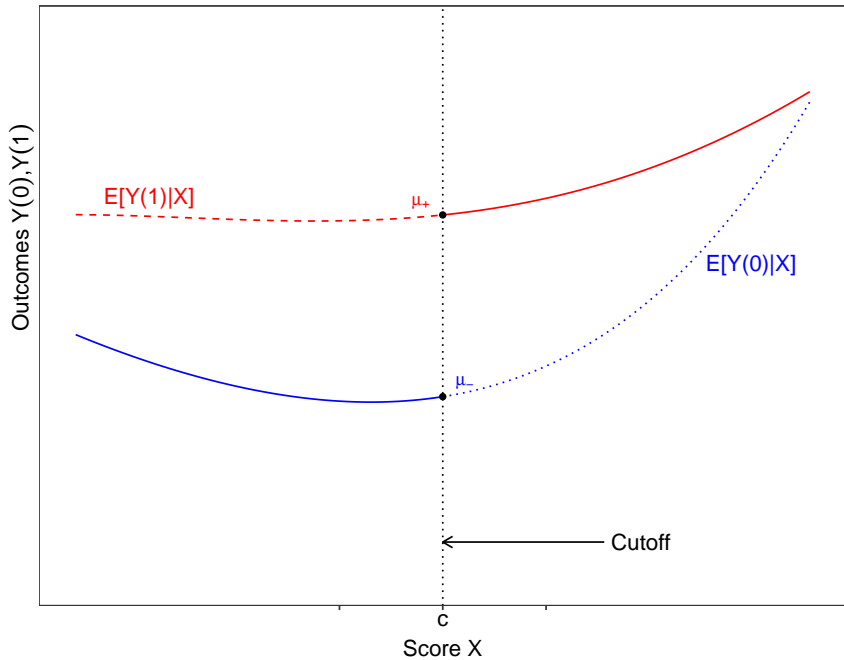
- **Idea:** approximate regression functions for control and treatment units *locally*.
- “Local-linear” ( $p = 1$ ) estimator (w/ weights  $K(\cdot)$ ):

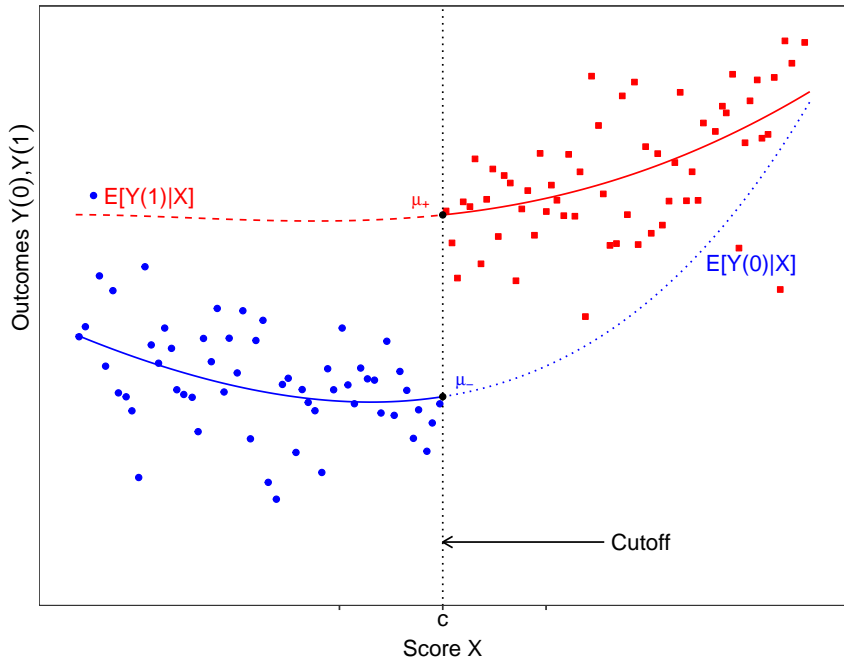
$$\begin{array}{c|c} -h \leq X_i < c : & c \leq X_i \leq h : \\ Y_i = \alpha_- + (X_i - c) \cdot \beta_- + \varepsilon_{-,i} & Y_i = \alpha_+ + (X_i - c) \cdot \beta_+ + \varepsilon_{+,i} \end{array}$$

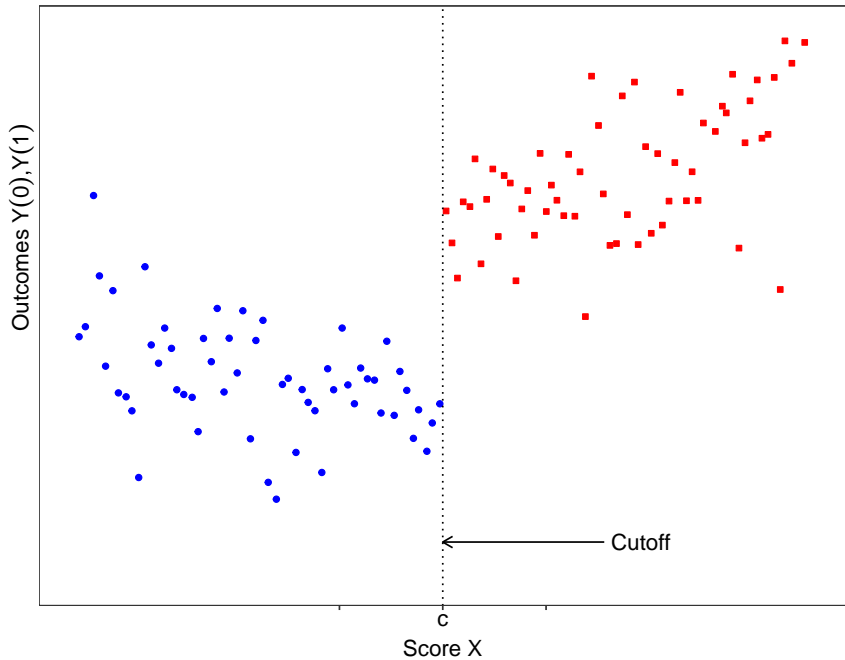
- ▶ Treatment effect (at the cutoff):  $\hat{\tau}_{\text{SRD}}(h) = \hat{\alpha}_+ - \hat{\alpha}_-$
- Can be estimated using linear models (w/ weights  $K(\cdot)$ ):
$$Y_i = \alpha + \tau_{\text{SRD}} \cdot T_i + (X_i - c) \cdot \beta_1 + T_i \cdot (X_i - c) \cdot \gamma_1 + \varepsilon_i, \quad |X_i - c| \leq h$$
- Given  $p, K, h$  chosen  $\implies$  weighted least squares estimation.

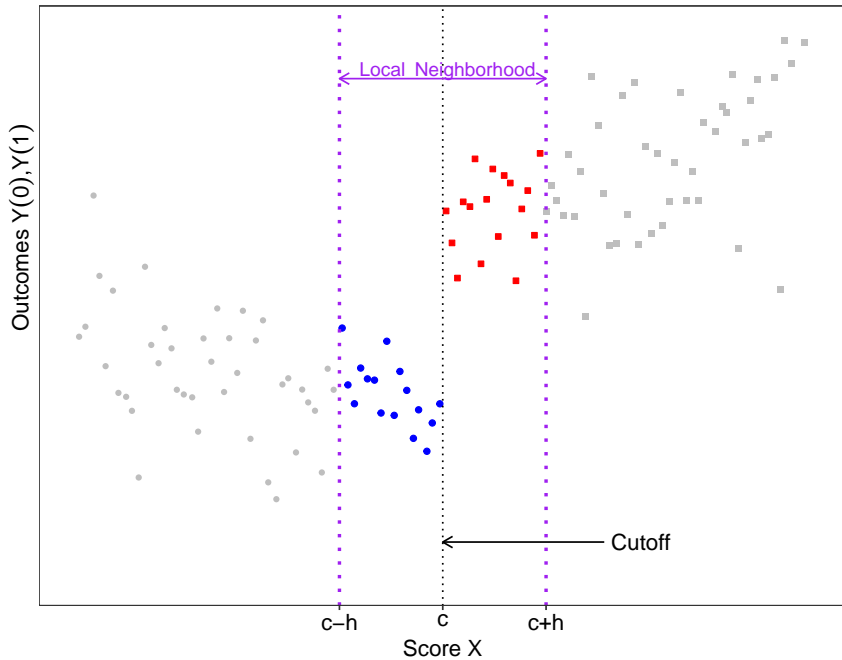


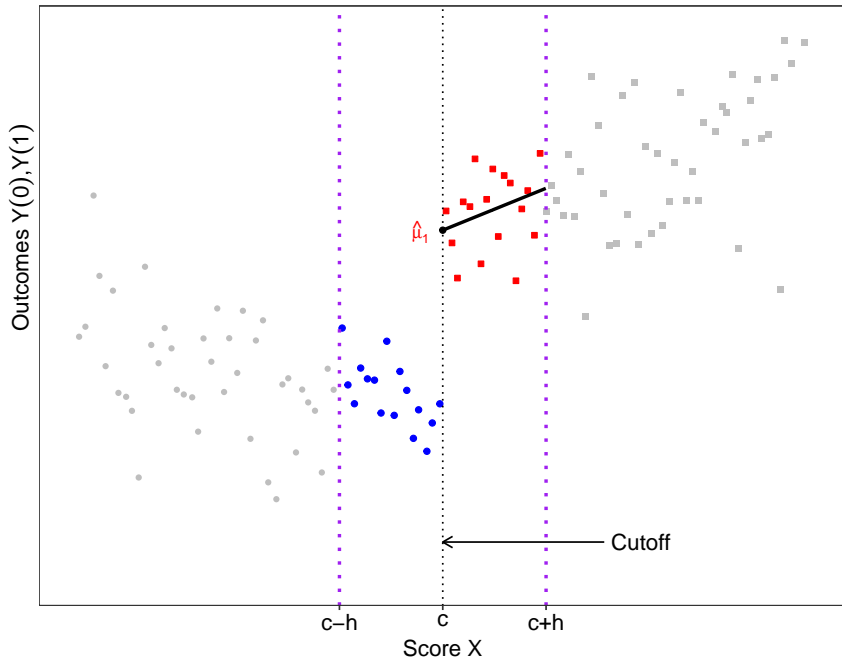


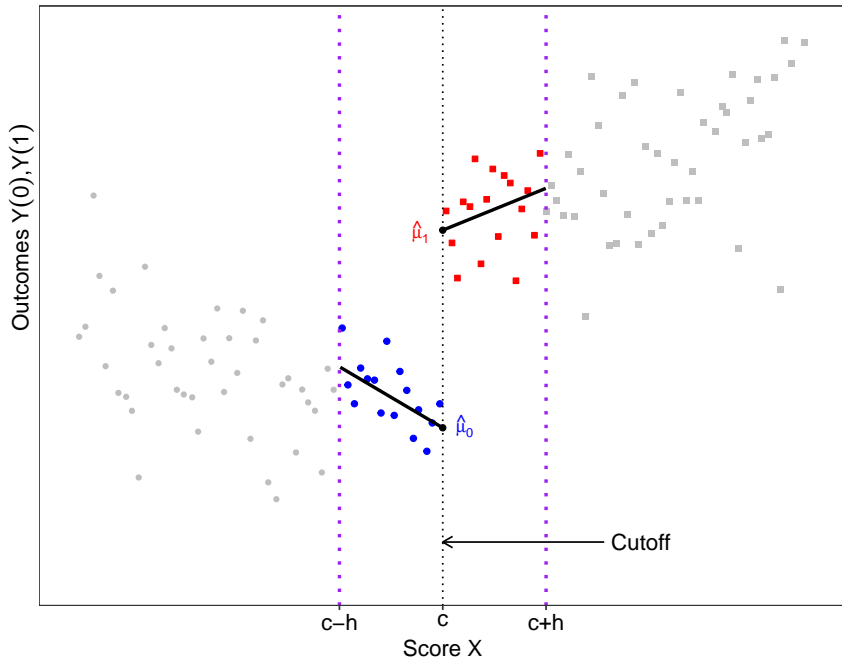


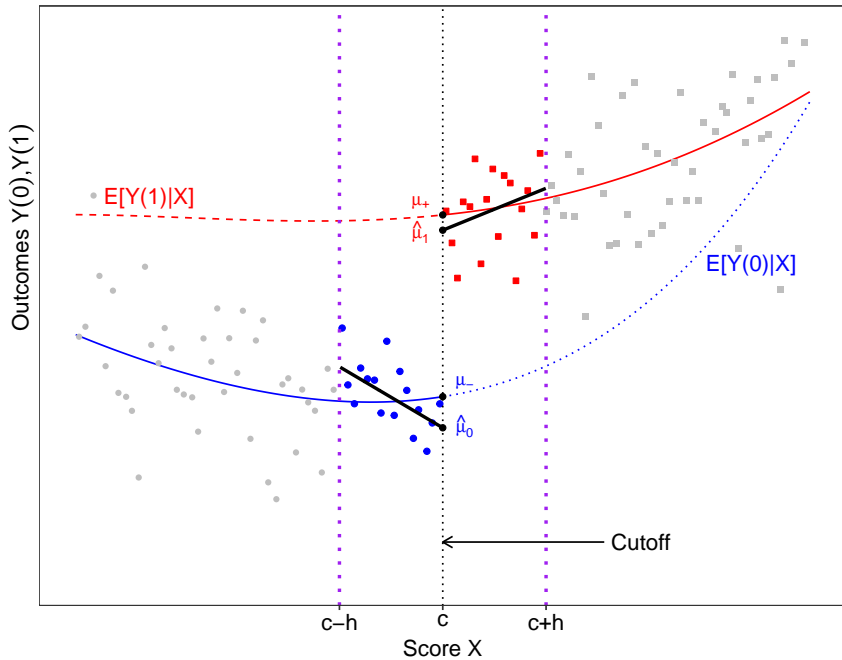




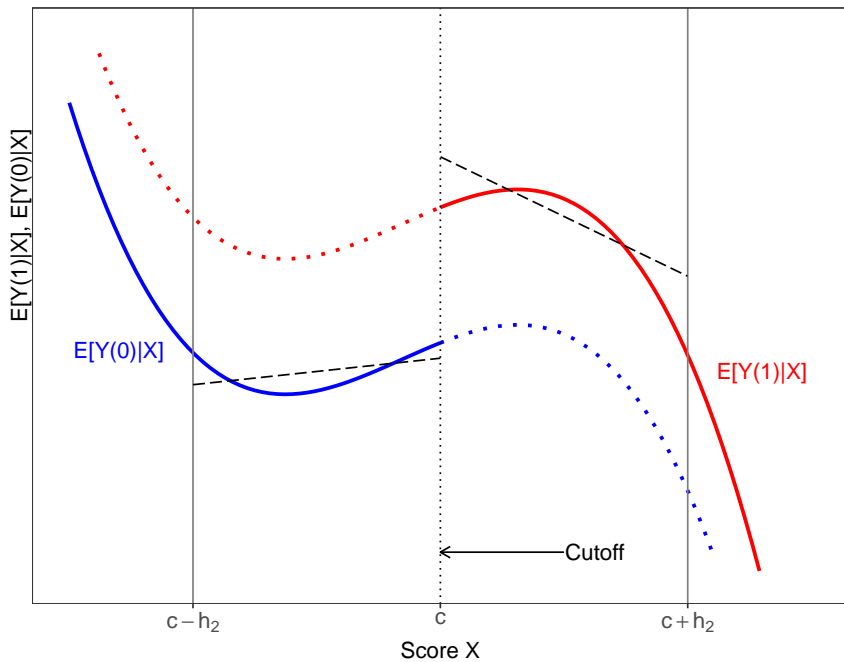


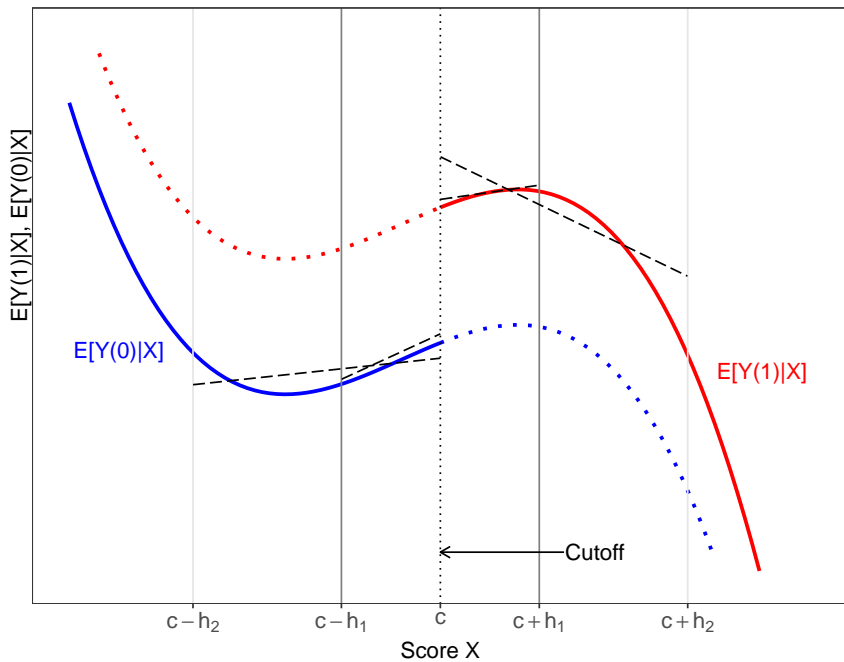












## Local Polynomial Methods: Choosing bandwidth ( $p = 1$ )

- Mean Square Error Optimal (MSE-optimal).

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

- Coverage Error Optimal (CE-optimal).

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

- **Key idea:**

- ▶ Trade-off bias and variance of  $\hat{\tau}_{\text{SRD}}(h)$ . Heuristically:

$$\uparrow \text{Bias}(\hat{\tau}_{\text{SRD}}) \quad \implies \quad \downarrow \hat{h} \quad \text{and} \quad \uparrow \text{Var}(\hat{\tau}_{\text{SRD}}) \quad \implies \quad \uparrow \hat{h}$$

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

## Conventional Inference Approach

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

$$\begin{array}{l|l} -h \leq X_i < c : & c \leq X_i \leq h : \\ Y_i = \alpha_- + (X_i - c) \cdot \beta_- + \varepsilon_{-,i} & Y_i = \alpha_+ + (X_i - c) \cdot \beta_+ + \varepsilon_{+,i} \end{array}$$

- ▶ Treatment effect (at the cutoff):  $\hat{\tau}_{\text{SRD}}(h) = \hat{\alpha}_+ - \hat{\alpha}_-$
- Construct usual t-test. For  $H_0 : \tau_{\text{SRD}} = 0$ ,

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

- Naïve 95% Confidence interval:

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

# Robust Bias Correction Approach

- **Key Problem:**

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

- ▶  $\mathbf{B}$  captures bias due to misspecification error.

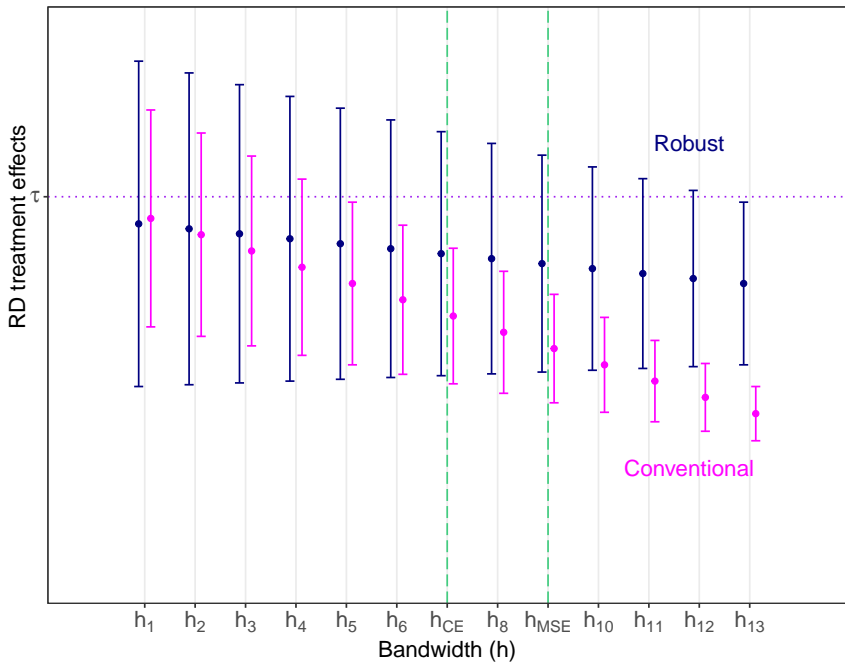
- **RBC distributional approximation:**

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

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

- **RBC 95% Confidence Interval:**

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



## 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:**

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

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

Variable	Control mean	Nonparametric estimator			Parametric	
					Flexible linear	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** (0.980) [0.036]	–1.198* (0.796) [0.081]	–1.114** (0.544) [0.027]	–2.201** (1.004) [0.022]	–2.558** (1.261) [0.021]
Specification checks						
Ages 5–9, injuries, 1973–1983	22.303	0.195 (3.472) [0.924]	2.426 (2.476) [0.345]	0.679 (1.785) [0.755]	–0.164 (3.380) [0.998]	0.775 (3.401) [0.835]
Ages 5–9, all causes, 1973–1983	40.232	–3.416 (4.311) [0.415]	0.053 (3.098) [0.982]	–1.537 (2.253) [0.558]	–3.896 (4.268) [0.317]	–2.927 (4.295) [0.505]
Ages 25+, Head Start-related causes, 1973–1983	131.825	2.204 (5.719) [0.700]	6.016 (4.349) [0.147]	5.872 (3.338) [0.114]	2.091 (5.581) [0.749]	2.574 (6.415) [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

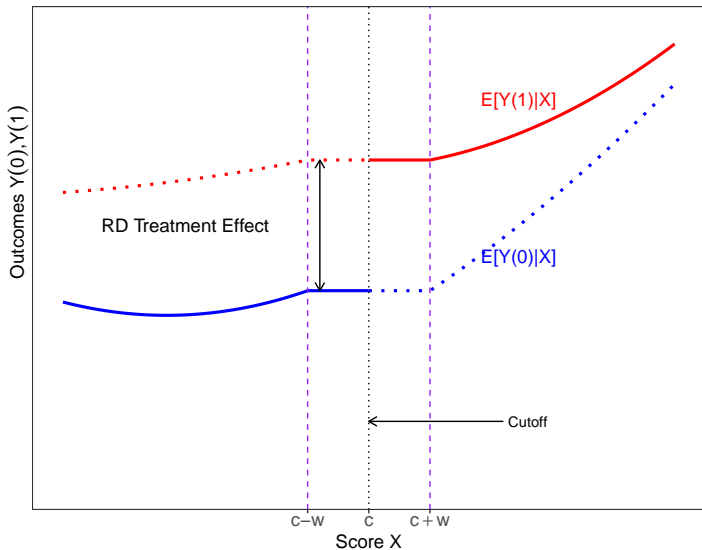
$$\text{for all } X_i \in \mathcal{W} \implies T_i \text{ independent of } (Y_i(0), Y_i(1))$$

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!*

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

+ exclusion restriction

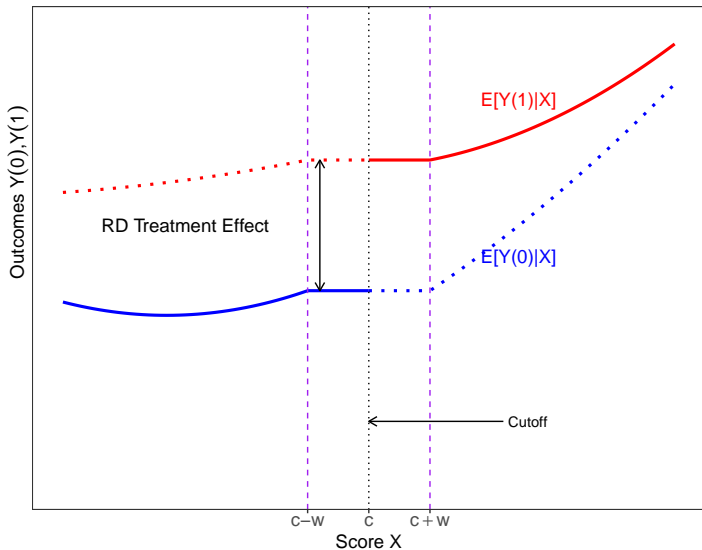


# Local Randomization: Finite-sample and Large-sample Methods

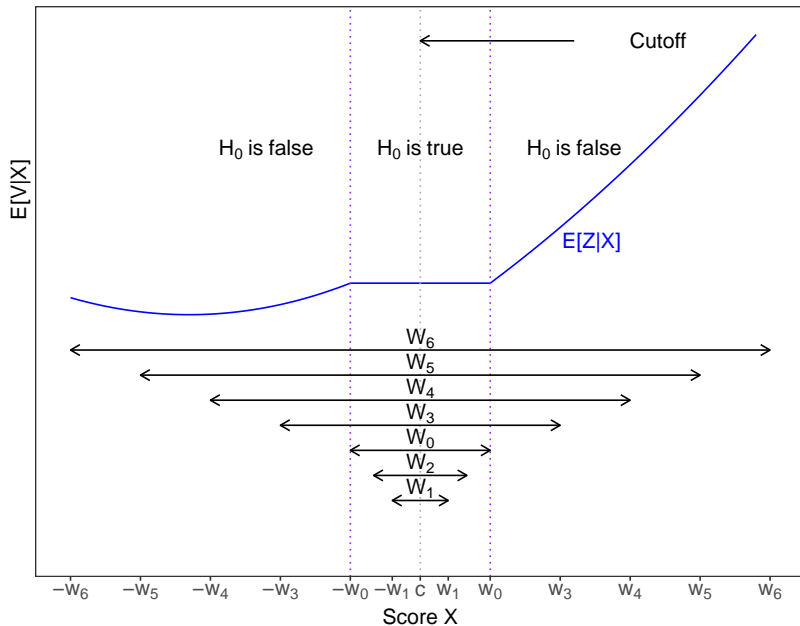
- Given  $\mathcal{W}$  where local randomization holds:
  - ▶ Randomization inference (Fisher): sharp null, finite-sample exact.
  - ▶ Design-based (Neyman): large-sample valid, conservative.
  - ▶ Large-sample standard: random potential outcomes, large-sample valid.
- All methods require window ( $\mathcal{W}$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

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

+ 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 ( $c = 59.1984$ )

$Z_i$  = 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:**

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

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

Variable	Control mean	Nonparametric estimator			Parametric	
					Flexible linear	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** (0.980) [0.036]	–1.198* (0.796) [0.081]	–1.114** (0.544) [0.027]	–2.201** (1.004) [0.022]	–2.558** (1.261) [0.021]
Specification checks						
Ages 5–9, injuries, 1973–1983	22.303	0.195 (3.472) [0.924]	2.426 (2.476) [0.345]	0.679 (1.785) [0.755]	–0.164 (3.380) [0.998]	0.775 (3.401) [0.835]
Ages 5–9, all causes, 1973–1983	40.232	–3.416 (4.311) [0.415]	0.053 (3.098) [0.982]	–1.537 (2.253) [0.558]	–3.896 (4.268) [0.317]	–2.927 (4.295) [0.505]
Ages 25+, Head Start-related causes, 1973–1983	131.825	2.204 (5.719) [0.700]	6.016 (4.349) [0.147]	5.872 (3.338) [0.114]	2.091 (5.581) [0.749]	2.574 (6.415) [0.689]

# Outline

- 1 Designs and Frameworks
- 2 Estimation and Inference
- 3 Falsification and Validation**
- 4 Extrapolation and Other Topics (if time permits)



# 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.).

## 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:**

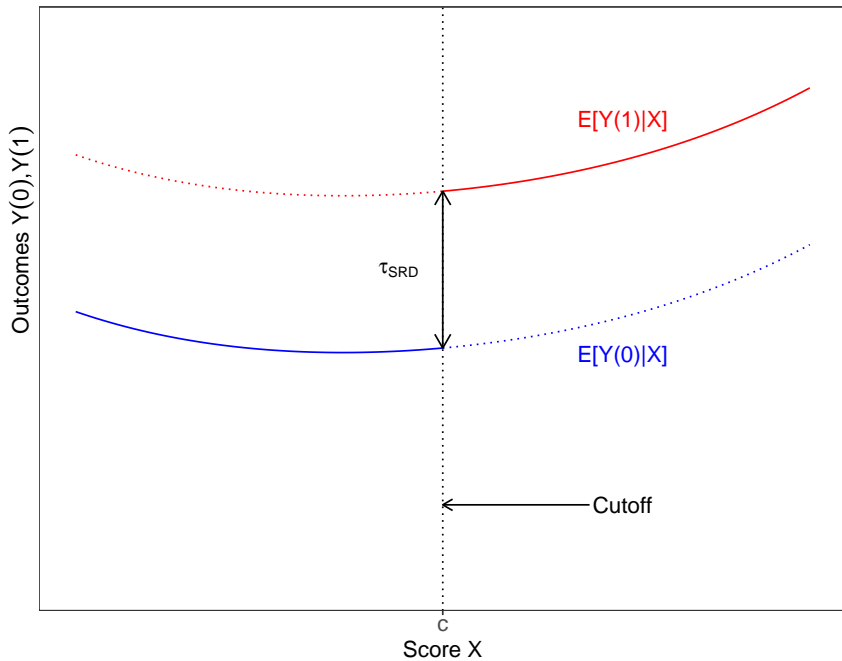
$$Y_i(0) \neq Y_i|T_i = 0 \quad \text{and} \quad Y_i(1) \neq Y_i|T_i = 1$$

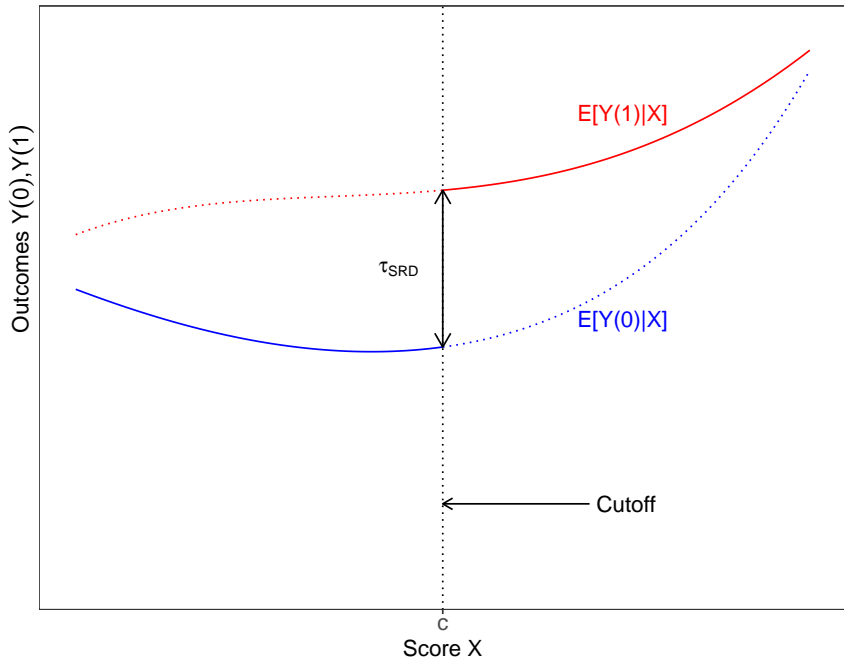
# Outline

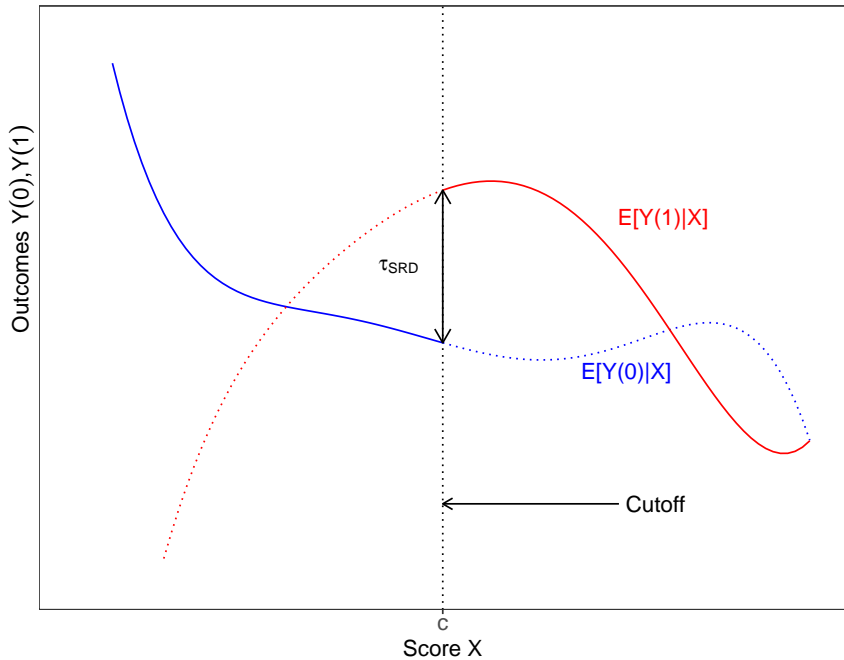
- 1 Designs and Frameworks
- 2 Estimation and Inference
- 3 Falsification and Validation
- 4 Extrapolation and Other Topics (if time permits)

# 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.







Thank you!

<https://rdpackages.github.io/>