Short Course in Regression Discontinuity Designs

Rocío Titiunik
Department of Politics
Princeton University
titiunik@princeton.edu

October 3-5, 2023
Mixtape Sessions, Virtual Course

Overview

Part 1: Introduction to Causal Inference and Policy Evaluation

Part 2: Introduction to Regression Discontinuity Designs

Part 3: Graphical illustration of RD models

Part 4: RD Designs: Local Polynomial Analysis

Part 5: RD Local Randomization Methods

Part 5: Fuzzy RD Designs

Part 7: RD Falsification Analysis

Part 1:

Introduction to Causal Inference and Policy Evaluation

Overview

- Introduction to Causal Inference and Policy Evaluation
  - Potential Outcomes and Assignment Mechanisms
  - Finite and Large Sample Inference in Randomized Experiments
    - Fisher's exact P-values Approach

Causal Inference

- The goal of program evaluation is to assess the causal effect of program or policy interventions. Examples:
  - Class size on test scores
  - Minimum wage on employment
  - Literacy intervention on kindergartners's reading ability
- In addition, we may be interested in the effect of variables that do not represent policy interventions. Examples:
  - Incentive scheme on employer productivity
  - Terrorist risk on economic behavior

Causes of effects vs. effects of causes

Important distinction between effect and cause

Cause: an event that generates some phenomenon

Effect: the consequence (or one of the consequences) of the cause

Crucial asymmetry in the difficulty of learning about the cause of an effect versus learning about the effect of a cause

Program evaluation focuses on effect-of-cause questions

- Not: why do younger citizens vote at lower rates?
- Rather: what is the effect of same-day registration on youth turnout?

Key Ideas

- Assignment mechanism is the procedure that determines which units are selected for treatment intake.
  - Examples include:
    1. random assignment
    2. selection on observables
    3. selection on unobservables
- Typically, treatment effects models attain identification by restricting the assignment mechanism in some way.
- Causality is defined by potential outcomes, not by realized (observed) outcomes.
- Observed association is neither necessary nor sufficient for causation.
- Estimation of causal effects of a treatment (usually) starts with studying the assignment mechanism.

Causal Inference Framework

Two essential ingredients:

1. Potential Outcomes: each individual has a different outcome corresponding to each level that the treatment takes

2. Assignment Mechanism: each individual is assigned treatment based on some mechanism, and this mechanism guides how estimation and inference will be conducted

1. Potential Outcomes: Causation as Manipulation

- For causal analysis, it is essential that "each unit be potentially exposable to any one of the causes" [Holland, 1986].
  - If units could have been exposed to cause but were not: no problem.
  - If units could not have been exposed to cause: might not really be a cause
  - Example: worker's education level versus worker's gender.
- Each i has as many potential outcomes as different possible treatments:
  - If treatment is binary, potential outcomes for unit i are $Y_i(1)$ and $Y_i(0)$
  - Called "potential" outcomes because only one of them is observed.
  - Observed outcome: outcome corresponding to level of treatment actually selected by (or assigned to) the unit.

1. Potential Outcomes: Causation as Manipulation

- This introduces the idea of counterfactual:

  What would the outcome of this unit look like if the unit had been exposed to a different treatment?

- **Key ideas**:
  - Non-manipulable attributes versus manipulable causes.
  - Pre-exposure ("pre-treatment") versus post-exposure ("post-treatment").

2. The assignment mechanism

$$\mathbb{P}(\mathbf{T}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1))$$

- Conditional probability of full assignment $\mathbf{T}$ given potential outcomes and covariates.
- The process by which each unit selected or was assigned the particular treatment condition that it received
- Two important cases

  Known, independent of $Y_i(0), Y_i(1)$: random assignment

  Unknown, (conditionally) independent of $Y_i(0), Y_i(1)$: unconfounded assignment

- Opposite of conventional focus on distribution of observed outcomes given covariates, $Y_i|X_i, T_i$
- $p_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \sum_{\mathbf{T}:T_i=1} \mathbb{P}(\mathbf{T}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1))$ (i's assignment prob)

Potential Outcomes Framework: Notation

Basic Binary Treatment Setup

- Each unit exposed to binary treatment $\Rightarrow$ two potential outcomes
  - $T_i = 1$ if i receives treatment; $T_i = 0$ if i receives control
  - $Y_i(1)$: outcome that would occur if i were exposed to treatment
  - $Y_i(0)$: outcome that would occur if i were exposed to control
- Observed data: $(Y_i, T_i)'$ where

$$Y_i = Y_i(T_i) = T_i \cdot Y_i(1) + (1 - T_i) \cdot Y_i(0)$$

More General Setup

- Multiple treatments: $T_i \in \mathcal{T}$ and $\{Y_i(t) : t \in \mathcal{T}\}$, with $\mathcal{T}$ finite, countable or uncountable.
- Throughout we assume: $\mathcal{T} = \{0, 1, 2, \cdots, J\}$
- Observed data: $Y_i = Y_i(T) = \sum_{t=0}^{J} \mathbb{1}(T_i = t) \cdot Y_i(t)$

Stable Unit Treatment Value Assumption (SUTVA)

- Key (implicit) assumption: $Y_i(t)$ depends only on i's treatment status
  - More general would be: $Y_i(\mathbf{t})$ with $\mathbf{t} = (t_1, t_2, \cdots, t_n)' \in \mathbb{R}^n$.
- Implies that potential outcomes of unit i are unaffected by treatment status of unit j.
- Rules out "interference", "spillovers", etc., across units, such as
  - Effect of fertilizer on plot yield.
  - Effect of flu vaccine on hospitalization.
- SUTVA may be problematic:
  - Choose the units of analysis to minimize interference across units!
  - Address "interference", "spillovers", etc., explicitly.

Treatment Effects of Interest

Treatment Effect with Binary Treatments:

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

- Effect of treatment cause (relative to control cause) on unit i.
- τ_i depends on potential outcomes, not observed outcomes.
- **Fundamental Problem of Causal Inference**:
  - For unit i, we observe either $Y_i(1)$ or $Y_i(0)$, but never both!
  - Impossible to learn about individual causal effect in general.

Define Aggregate Treatment Effects:

- Average Treatment Effect: $\tau_{\texttt{ATE}} := \mathbb{E}[Y_i(1) - Y_i(0)]$
- Quantile Treatment Effect: $\tau_{\texttt{QTE}}(p) := F^{-1}_{Y_i(1)}(p) - F^{-1}_{Y_i(0)}(p)$
  - Can be defined for sample or (super) population.
  - Can be defined for subpopulations (on T or observables covariates).

Random Assignment of Treatment

Restrictions on the assignment mechanism

- Probabilistic: $0 < p_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) < 1$
- Unconfounded: $\Pr(\mathbf{T}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \Pr(\mathbf{T}|\mathbf{X}, \mathbf{Y}(0)', \mathbf{Y}(1)')$
- Individualistic: $p_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = q(X_i, Y_i(0), Y_i(1))$

Classical Randomized Experiment

- A classical randomized experiment is an assignment mechanism that
  1. is probabilistic
  2. is individualistic
  3. is unconfounded
  4. has functional form that is known and controlled by the researcher

Classical Randomized Experiments

- **Taxonomy**
  - Bernoulli trials: $\mathbb{P}[\mathbf{T} = \mathbf{t}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)] = p^n$ with $p \in (0, 1)$.
  - Fixed margins: $\mathbb{P}[\mathbf{T} = \mathbf{t}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)] = \binom{n_1}{n}^{-1}$ where $n_1 = \sum_{i=1}^{n} \mathbb{1}(T_i = 1)$
  - Stratified randomized experiments: fixed margins by subgroups.
  - Paired randomized experiments: stratified experiments with $n_j = 2$ for all j.

- All of the above designs satisfy the main characteristics of a classical randomized experiment:
  - individualistic assignment mechanism
  - probabilistic assignment mechanism
  - unconfounded assignment mechanism
  - known assignment mechanism

Analysis of Randomized Experiments

Two frameworks for analysis:

- Conventional or super-population framework
  - Data is sample from a larger population
  - Potential outcomes are random variables
  - Inference relies on large-sample approximations

- Fisherian framework
  - The units in the sample are the population: no sampling, no approximations
  - Potential outcomes are fixed quantities
  - Inference relies on exact finite-sample distribution of treatment assignment

Analysis of Experiments: Conventional Framework

- When potential outcomes are random and treatment is randomly assigned, treatment and potential outcomes are statistically independent

$$T_i \perp\!\!\!\perp (Y_i(0), Y_i(1))$$

- This implies

$$T_i \perp\!\!\!\perp Y_i(0) \quad \text{and} \quad T_i \perp\!\!\!\perp (Y_i(1) - Y_i(0))$$

and

$$\begin{aligned}
\mathbb{E}[Y_i(1)] &= \mathbb{E}[Y_i(1)|T_i = 1] \\
\mathbb{E}[Y_i(1)] &= \mathbb{E}[Y_i(1)|T_i = 0] \\
\mathbb{E}[Y_i(0)] &= \mathbb{E}[Y_i(0)|T_i = 1] \\
\mathbb{E}[Y_i(0)] &= \mathbb{E}[Y_i(0)|T_i = 0] \\
\mathbb{E}[\mathbf{X}_i] &= \mathbb{E}[\mathbf{X}_i|T_i = 1] \\
\mathbb{E}[\mathbf{X}_i] &= \mathbb{E}[\mathbf{X}_i|T_i = 0]
\end{aligned}$$

for (predetermined) covariates $\mathbf{X}_i$

Analysis of Experiments: Conventional Framework

- We can use observed outcomes to learn about potential outcomes

$$\begin{aligned}\mathbb{E}[Y_i(1)] &= \mathbb{E}[Y_i(1)|T_i = 1] = \mathbb{E}[Y_i|T_i = 1] \\ \mathbb{E}[Y_i(0)] &= \mathbb{E}[Y_i(0)|T_i = 0] = \mathbb{E}[Y_i|T_i = 0]\end{aligned}$$

- The average treatment effect is recovered from *observed* outcomes:

$$\mathbb{E}[Y_i(1) - Y_i(0)] = \mathbb{E}[Y_i|T_i = 1] - \mathbb{E}[Y_i|T_i = 0]$$

- Furthermore, ATE and ATET are equal because

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

Analysis of Experiments: Conventional Framework

- Also, QTE and QTET are identified and equal

$$\tau_{\texttt{QTE}}(q) = F^{-1}_{Y_i(1)}(q) - F^{-1}_{Y_i(0)}(q) = F^{-1}_{Y_i(1)|T_i=1}(q) - F^{-1}_{Y_i(0)|T_i=1}(q) = \tau_{\texttt{QTET}}(q)$$

  - However, RCTs do not identify the quantiles of the effect: $F^{-1}_{Y_i(1)-Y_i(0)}(q)$

Analysis of Experiments: Conventional Framework

- Suppose n units randomly assigned to two treatments. Consider $\tau_{\texttt{ATE}} = \mathbb{E}[Y_i(1) - Y_i(0)]$.
- Takes potential outcomes as random.
- Plug-in approach (analogy approach), we may construct:

$$\hat{\tau} = \bar{Y}_1 - \bar{Y}_0$$

with $\bar{Y}_1 = \frac{\sum_{i=1}^n T_i Y_i}{\sum_{i=1}^n T_i} = \frac{\sum_{i=1}^n T_i Y_i}{N_1}$, $\bar{Y}_0 = \frac{\sum_{i=1}^n (1-T_i) Y_i}{\sum_{i=1}^n (1-T_i)} = \frac{\sum_{i=1}^n (1-T_i) Y_i}{N_0}$

- $\hat{\tau}$ is unbiased for $\tau_{\texttt{ATE}}$, and $\hat{\tau}$ is consistent for $\tau_{\texttt{ATE}}$.

Analysis of Experiments: Conventional Framework

- Suppose n units randomly assigned to two treatments. Consider $\tau_{\texttt{ATE}} = \mathbb{E}[Y_i(1) - Y_i(0)]$.
- Distribution theory: under $H_0 : \tau_{\texttt{ATE}} = \tau_0$,

$$W_n = \frac{\hat{\tau} - \tau_0}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_0^2}{N_0}}} \to_d \mathcal{N}(0,1), \qquad \sigma_t^2 = \frac{1}{N_t - 1} \sum_{i=1}^{n} D_i(t) \left(Y_i - \bar{Y}_t\right)^2$$

- We reject H_0 (against $H_1 : \tau_{\texttt{ATE}} \neq \tau_0$) at level $\alpha \in (0,1)$ iff $|W_n| > \Phi_{1-\alpha/2}$.
- $(1-\alpha)$ Confidence interval for τ_0 :

$$\mathsf{CI}_{1-\alpha}(\tau_0) = \left[\hat{\tau} - \Phi_{1-\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_0^2}{N_0}}\,,\ \hat{\tau} + \Phi_{1-\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_0^2}{N_0}}\right]$$

Analysis of Experiments: Fisherian Framework

- Potential outcomes are fixed quantities
- The units in the sample are seen as the population: sample size is also fixed
- Hypothesis of interest is a sharp null hypothesis that allows for imputation of full profile of potential outcomes
- Inferences are based on the known randomization distribution of the treatment assignment

Analysis of Experiments: Fisherian Framework

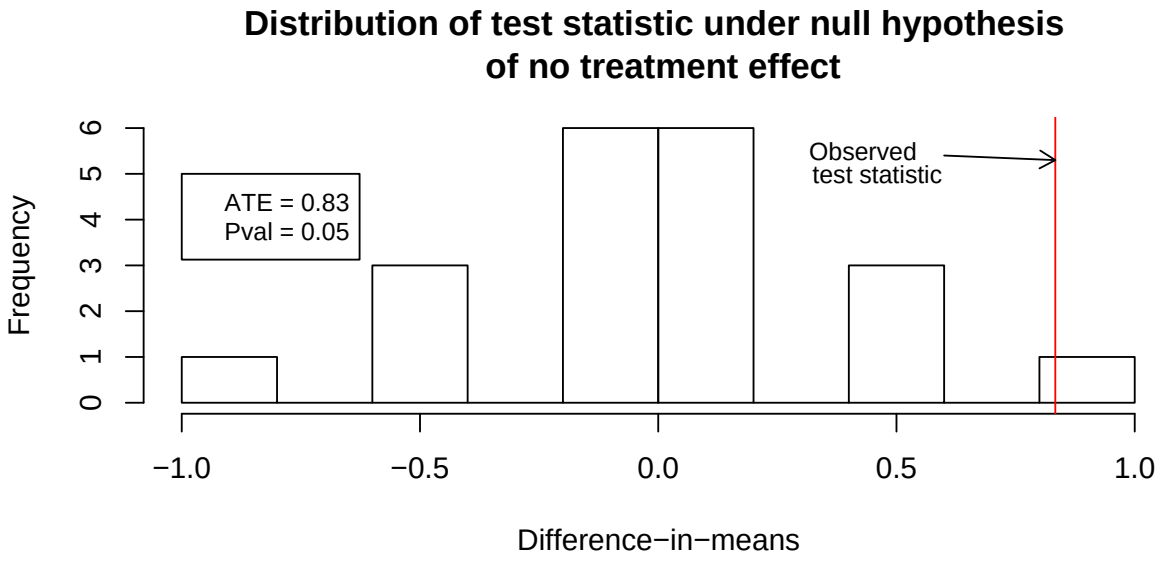
- Because we ran an experiment, the randomization mechanism that assigned units to treatment and control is *known*
- Since this assignment is entirely known, the distribution of the treatment assignment is known
- Under the sharp null, the only source of randomness is the treatment assignment
- Therefore, we can use the known distribution of the random variable $\mathbf{T}$ to derive the distribution of *any* test-statistic $s(\mathbf{T}, \mathbf{Y})$

Fisherian Framework: Example

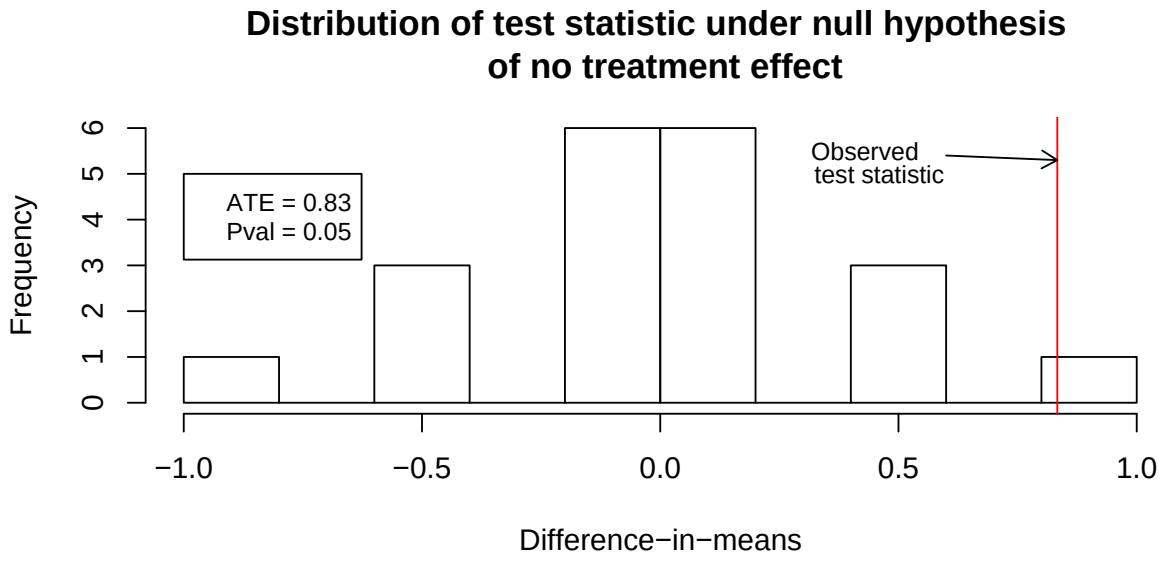
- Six subjects assigned binary treatment: 3 treated, 3 control
- Realized treatment assignment $\mathbf{T} = [1, 0, 0, 1, 1, 0]$
- $\binom{6}{3} = 20$ possible treatment assignments
- Obtain the distribution of the test-statistic under the null hypothesis of no treatment effect

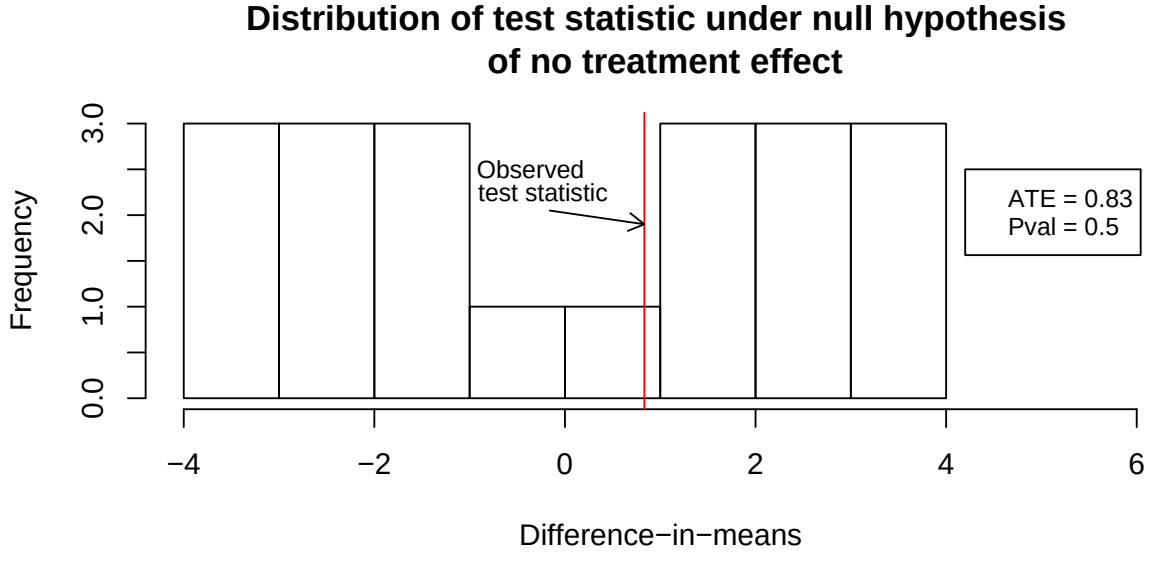
$$
\begin{cases}
\tilde{T}_1 = [1,1,1,0,0,0] \Longrightarrow \tilde{s}_1 = \frac{Y_1+Y_2+Y_3}{3} - \frac{Y_4+Y_5+Y_6}{3} \\
\tilde{T}_2 = [1,1,0,0,0,1] \Longrightarrow \tilde{s}_2 = \frac{Y_1+Y_2+Y_6}{3} - \frac{Y_3+Y_4+Y_5}{3} \\
\vdots \qquad\qquad\qquad\qquad\qquad\qquad \vdots \\
\tilde{T}_{20} = [0,0,0,1,1,1] \Longrightarrow \tilde{s}_{20} = \frac{Y_4+Y_5+Y_6}{3} - \frac{Y_1+Y_2+Y_3}{3}
\end{cases}
$$

Fisherian Framework: Example



Fisherian Framework: Example





Analysis of Experiments: Fisherian Framework

- Let $\mathbf{T}$ be an n-dimensional column vector whose elements are the T_i for all units
- We randomize $\mathbf{T}$: this randomization mechanism is by known by definition
- We collect all observed outcomes Y_i in $\mathbf{Y}$
- We test sharp null hypothesis

$$H_0 : Y_i(1) = Y_i(0) \quad \text{for } i = 1, 2, \ldots N$$

- Under H_0, $Y_i = Y_i(1) = Y_i(0)$ for every i: both potential outcomes are known

Analysis of Experiments: Fisherian Framework

- In order to test the null hypothesis, we define a *test-statistic* $s(\mathbf{T}, \mathbf{Y})$
- Example: difference in means

$$s(\mathbf{T}, \mathbf{Y}) = \frac{\sum_{i=1}^{n} T_i Y_i}{\sum_{i=1}^{n} T_i} - \frac{\sum_{i=1}^{n} (1 - T_i) Y_i}{\sum_{i=1}^{n} (1 - T_i)}$$

- Since $\mathbf{Y}$ is fixed under the null, $\mathbb{P}\left[S(\mathbf{Y}, \mathbf{T}) \leq s\right]$ is fully known because the law of $\mathbf{T}$ is known
- The (one-sided) exact p-value is the probability of seeing a value of $s(\mathbf{T}, \mathbf{Y})$ equal to or greater than the observed value $s(\mathbf{t}, \mathbf{Y})$

$$\text{p-value} = \Pr(s(\mathbf{T}, \mathbf{Y}) \geq s(\mathbf{t}, \mathbf{Y})) = \sum_{\mathbf{t} \in \Omega} \mathbb{1}\left\{s(\mathbf{t}, \mathbf{Y}) \geq S\right\} \cdot \Pr(\mathbf{T} = \mathbf{t})$$

- Thus, the p-value is measure of how unusual the observed value of the test-statistic is given its null distribution

Analysis of Experiments: Fisherian Framework

- We collect all possible realizations of $\mathbf{T}$ in the set Ω
- The treatment assignment mechanism must give a positive probability of receiving treatment and control to every unit
- Many mechanisms satisfy this requirement
- Common choice: "complete" or "fixed-margins" randomization
  - Fix number of treatments at n_t and number of controls at n_c
  - In this case, the set Ω has $\binom{n}{n_t} = \frac{n!}{n_t!(n-n_t)!}$ elements
  - All elements of Ω equally likely, $\Pr(\mathbf{T} = \mathbf{t}) = \frac{1}{\binom{n}{n_t}}$
  - p-value $= \Pr(s(\mathbf{T}, \mathbf{Y}) \geq s(\mathbf{t}, \mathbf{Y})) = \sum_{\mathbf{t} \in \Omega} \mathbb{1}\{s(\mathbf{t}, \mathbf{Y}) \geq S\} \cdot \frac{1}{\binom{n}{n_t}}$

Example: Job Training Partnership Act (JTPA)

- Largest randomized training evaluation ever undertaken in the U.S.
  - started in 1983 at 649 sites throughout the country.
- Sample: Disadvantaged persons in labor market (previously unemployed or low earnings)
- T_i: Assignment to one of three general service strategies
  - classroom training in occupational skills
  - on-the-job training and/or job search assistance
  - other services (eg. probationary employment)
- Y_i: earnings 30 months following assignment
- $\mathbf{X}_i$: Characteristics measured before assignment
  - age, gender, previous earnings, race, etc.

Discussion and Final Remarks

Threats to the Validity of Randomized Experiments

- *Internal validity*: can we estimate treatment effect in our sample?
  - Fails when there are differences between treated and controls (other than the treatment itself) that affect the outcome and that we cannot control for.
- *External validity*: can we extrapolate our estimates to other populations?
  - Fails when the treatment effect is different outside the evaluation environment.

Most Common Threats to Internal Validity

- Failure of randomization.
- Non-compliance with experimental protocol.
- Attrition.

Most Common Threats to External Validity

- Non-representative sample
- Non-representative program
  - The treatment differs in actual implementations.
  - Scale effects.
  - Actual implementations are not randomized (nor full scale).

Discussion and Final Remarks

Analysis and Falsification of Randomized Experiments

- Covariate Balance.
  - Randomization balances observed but also unobserved characteristics between treatment and control group.
  - Can check random assignment using so called "balance tests" (e.g., t-tests) to see if distributions of the observed covariates, **X**, are the same in the treatment and control groups.
  - **X** are pre-treatment variables that are measured prior to treatment assignment (i.e., at "baseline").

- Placebo Analysis.
  - Treatment does not affect all possible outcomes.
  - Can be used to check credibility of research designs.

Part 2:

Introduction to Regression Discontinuity Designs

Overview

The RD Design: Definition and Taxonomy

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

Causal Inference

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

Regression Discontinuity Design

Defined by the triplet: score, cutoff, treatment.

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

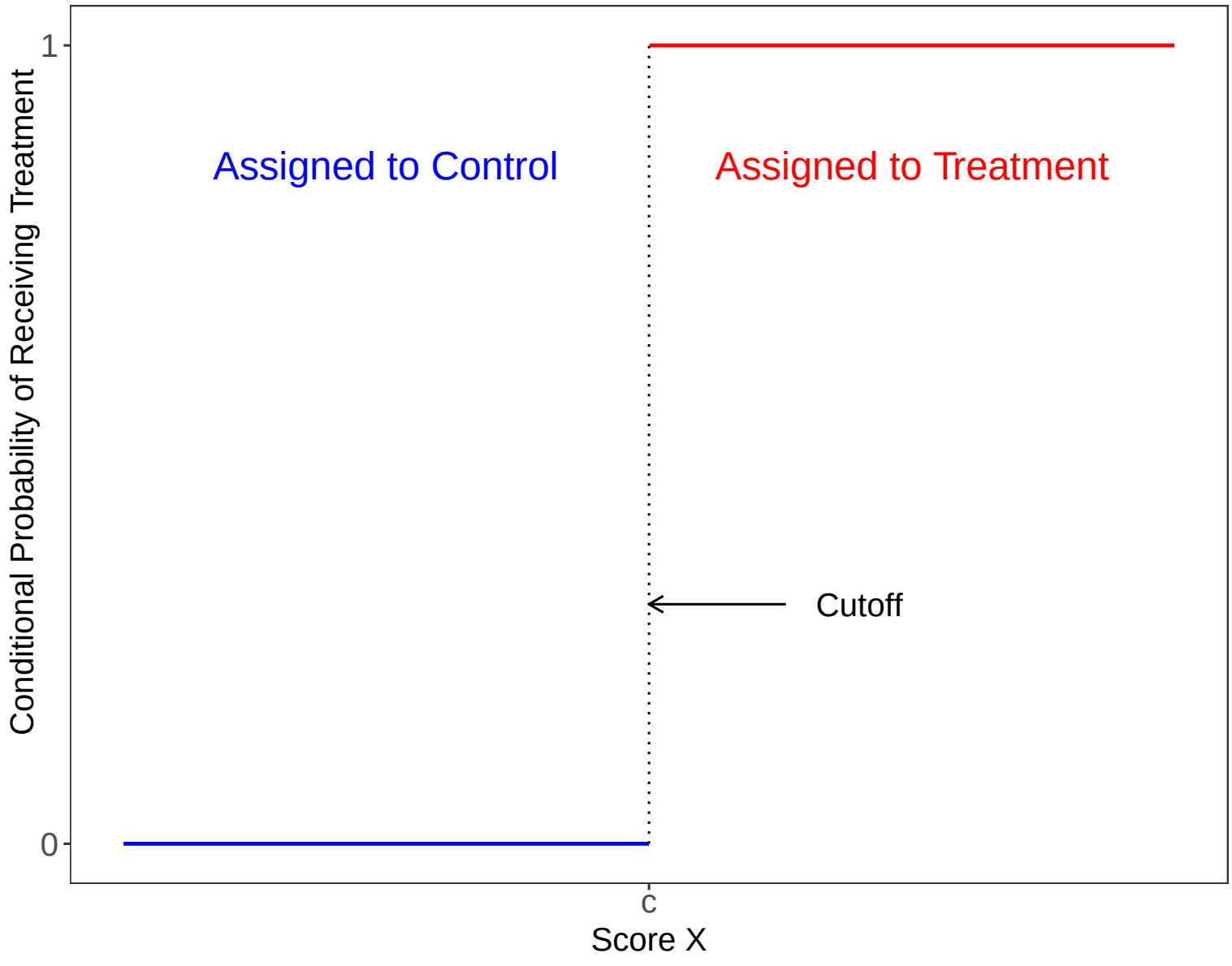
Regression Discontinuity Design

Defined by the triplet: score, cutoff, treatment.

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

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

Treatment Assignment in (Sharp) RD Design



Sharp Regression Discontinuity Design

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

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

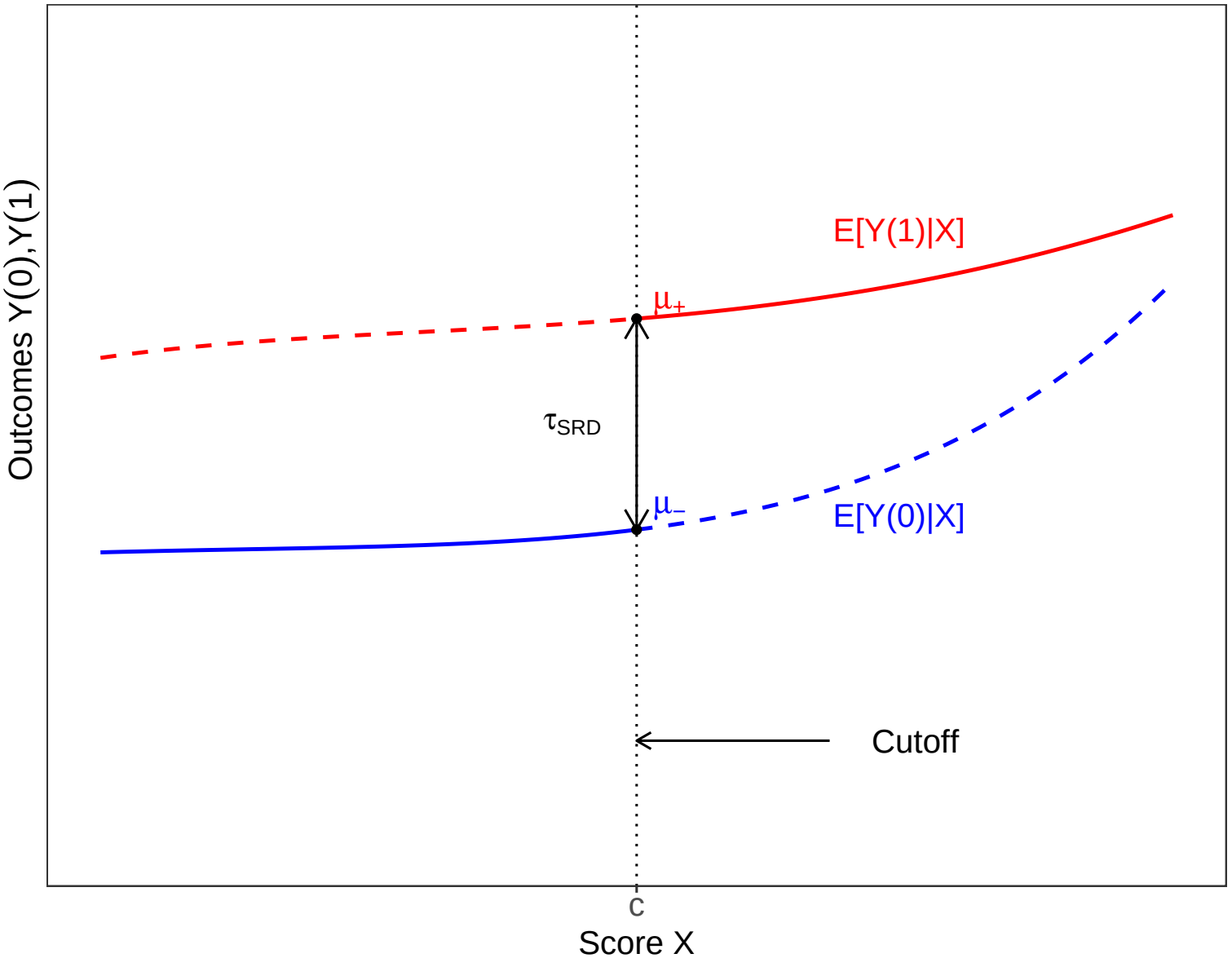
  $Y_i(0)$: outcome that would be observed if i received control

- The *observed* outcome is

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

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

RD Treatment Effect in Sharp RD Design



Fundamental Missing Data Problem

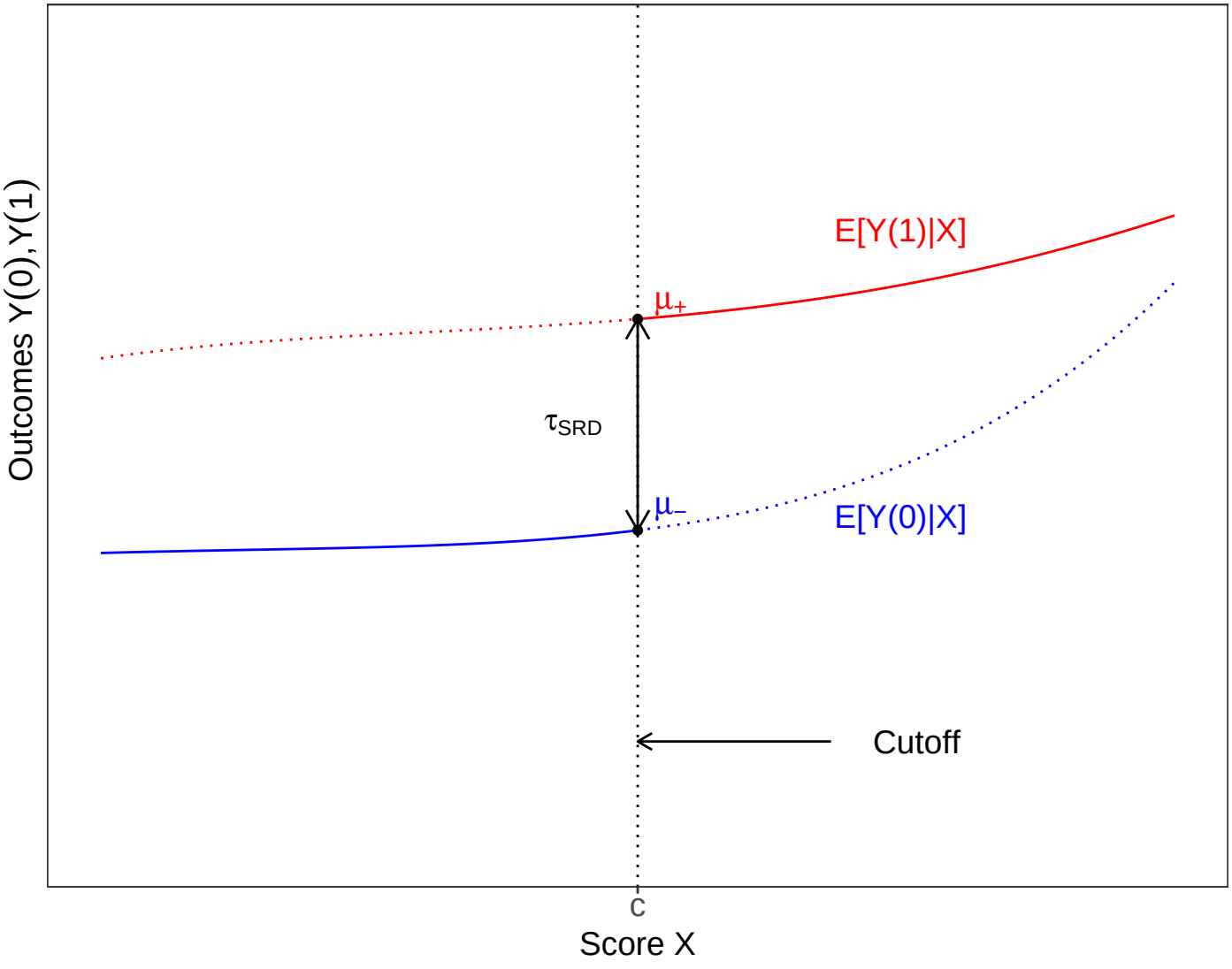
- A special situation occurs at the cutoff $X = c$, the only point at which we may "almost" observe both curves
- Imagine two groups of units:

  with score equal to c, $X_i = c \rightarrow$ treated

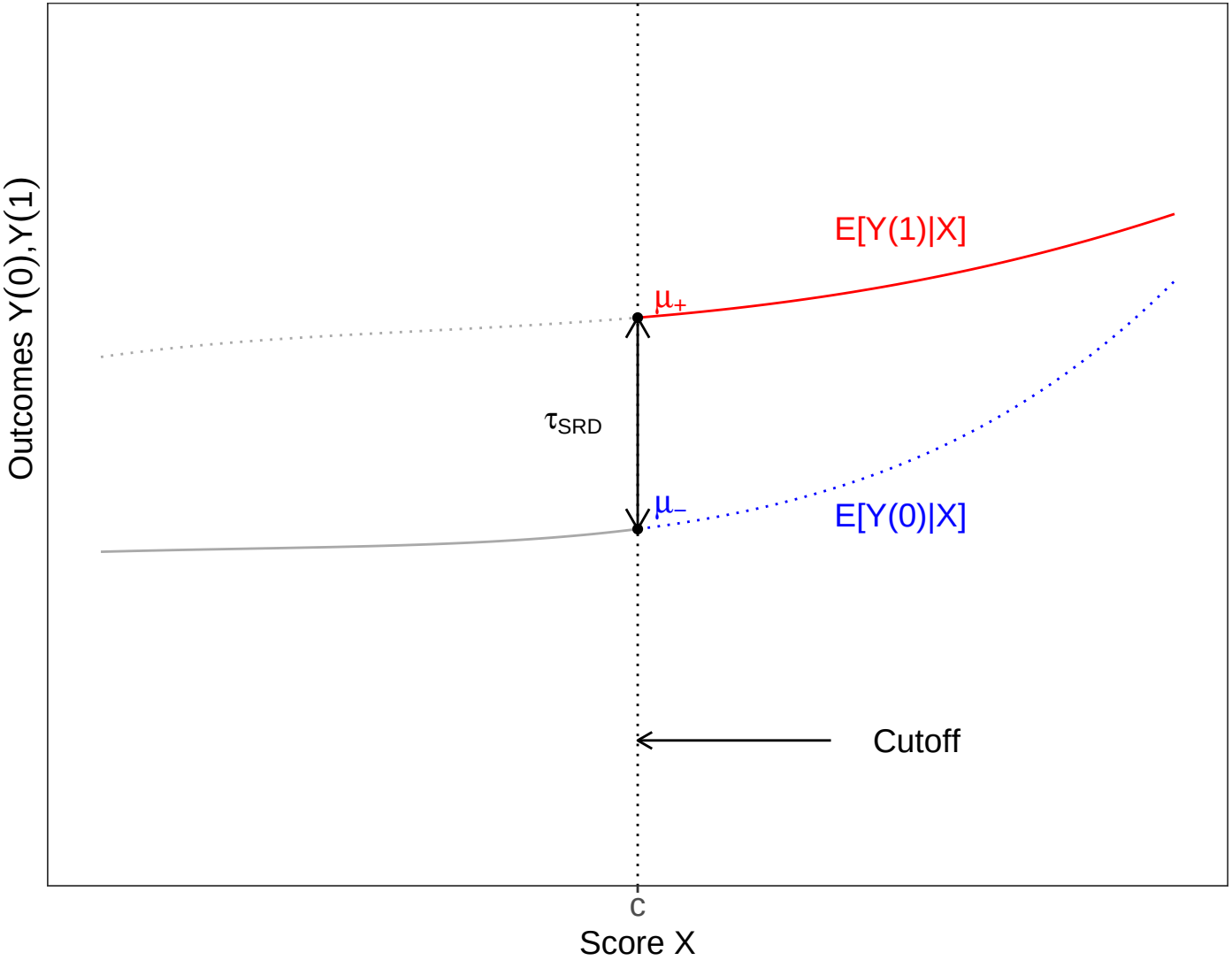
  with with score barely below c, $X = c - \varepsilon \rightarrow$ control

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

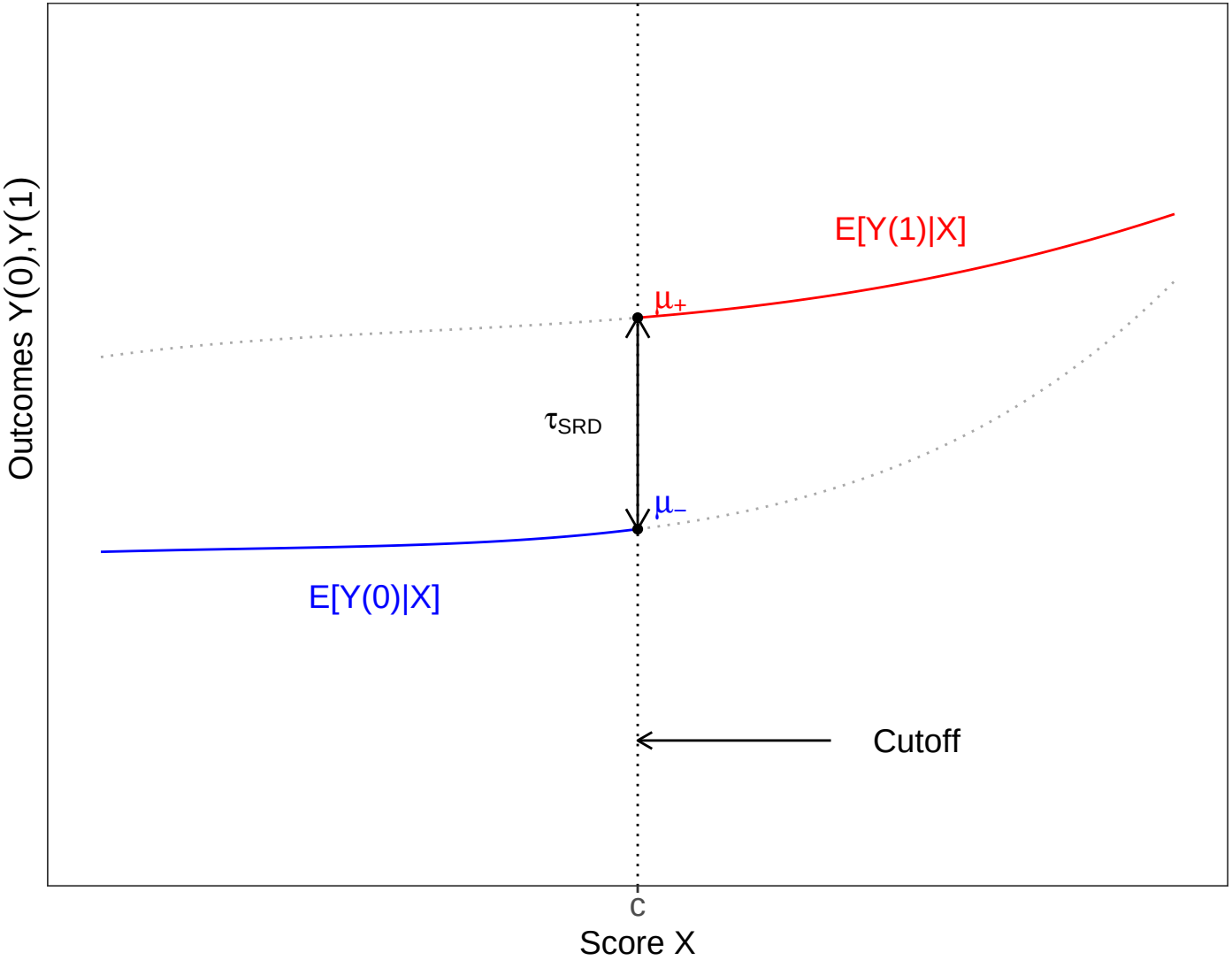
$$\tau_{\text{SRD}} = \mathbb{E}[Y_i(1) - Y_i(0)|X_i = c]$$



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



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



Sharp RD design: Summary

- Canonical Parameter:

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

- Perfect compliance:
  - every unit with score above c receives treatment.
  - every unit with score below c receives control.
- Not a "causal parameter" in the "proper" sense.
- Lee (2008) interpretation:

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

- Different interpretation under "local randomization".

Example: Incumbency Advantage in U.S. Senate

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

Example: Incumbency Advantage in U.S. Senate

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

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

  $T_i =$ whether party wins election at t .

  $X_i =$ margin of victory at $t \quad (c = 0)$.

  $Z_i =$ covariates (*demvoteshlag1*, *demvoteshlag2*, *dopen*, etc.).

- **Potential outcomes**:

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

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

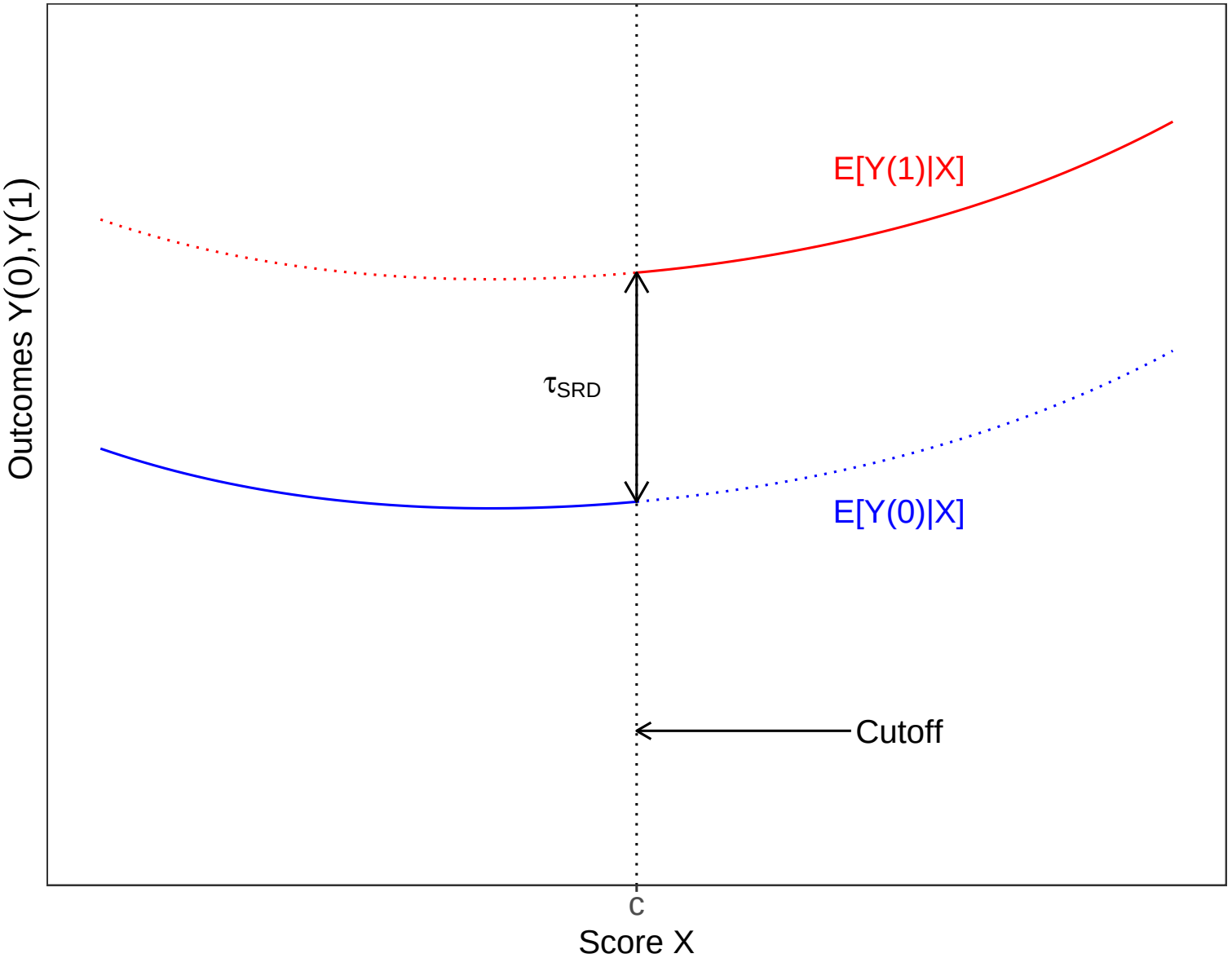
- **Causal Inference**:

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

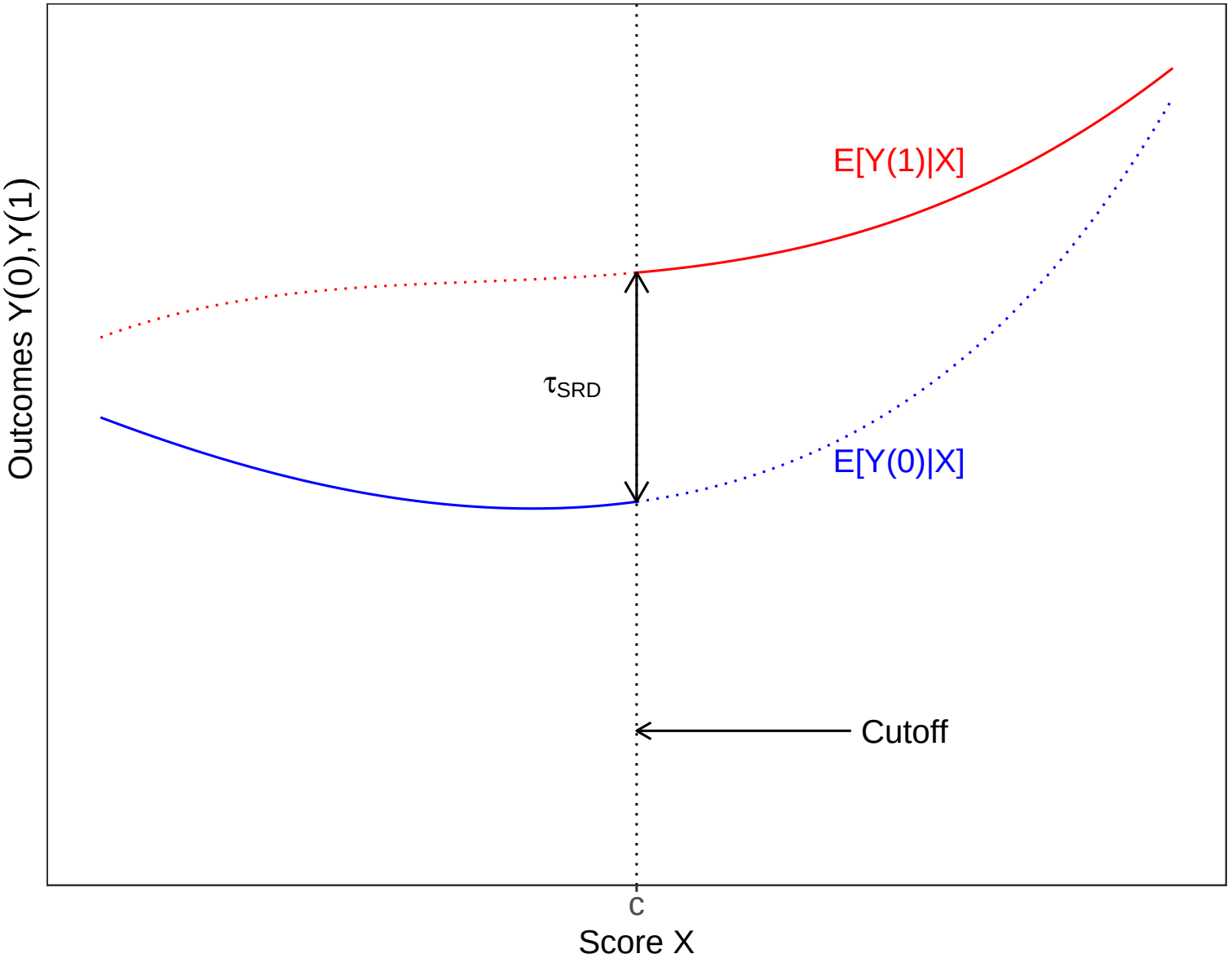
Local Nature of RD Effects

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

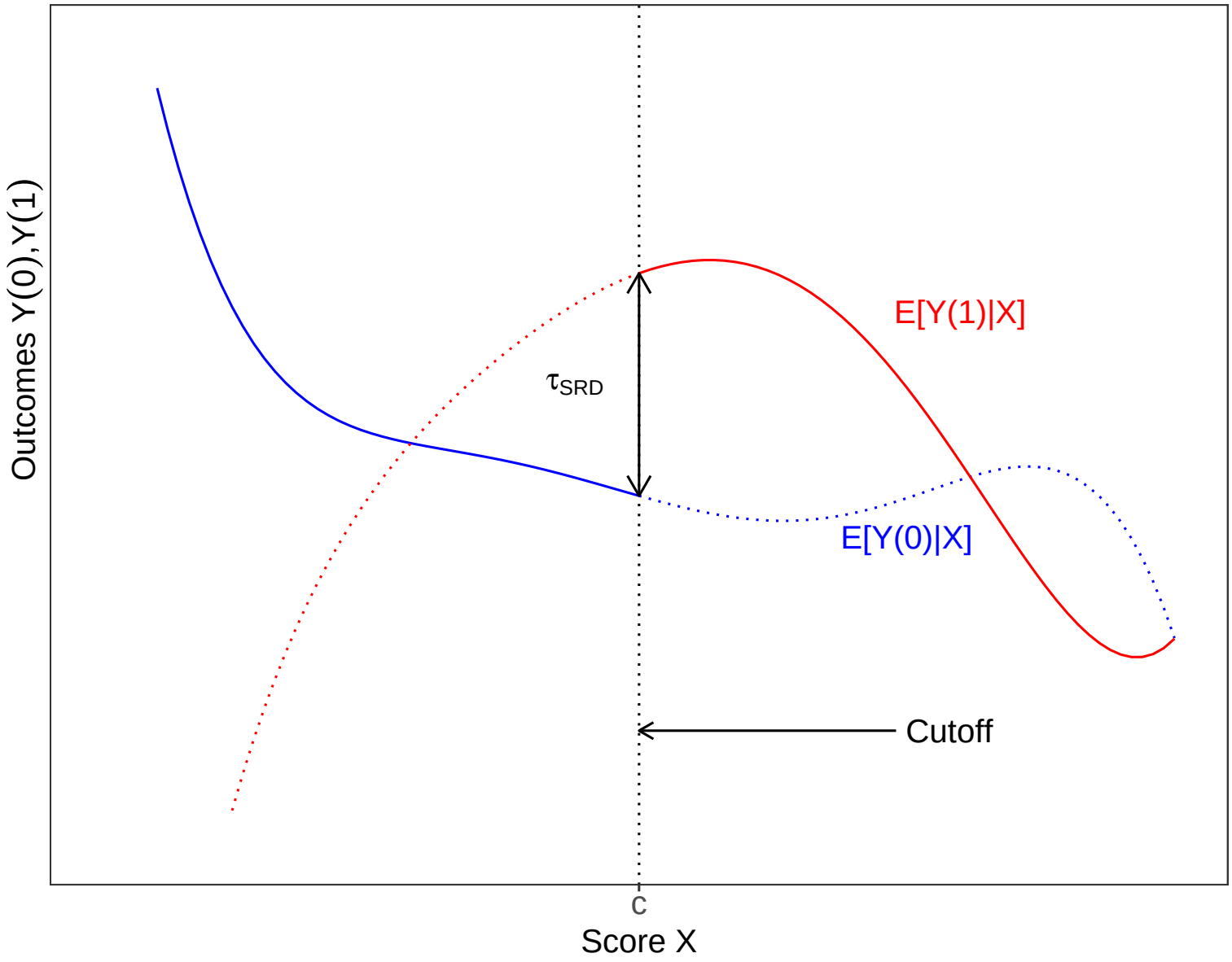
The RD Parameter: No Heterogeneity



The RD Parameter: Mild Heterogeneity



The RD Parameter: Wild Heterogeneity



Part 3:

Graphical illustration of RD models

RD Packages

https://rdpackages.github.io/

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

Graphical illustration of RD models

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

Graphical illustration of RD models

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

(i) Global fit: smooth approximation to the unknown regression functions

- 4th or 5th order polynomials, separately above and below the cutoff.

(ii) Local sample means:

- disjoint intervals (bins) of the score, calculating the mean of the outcome within each bin.

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

Graphical illustration of RD models

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

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

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

  $Y_i =$ child mortality 5 to 9 years old

  $T_i =$ whether county received Head Start assistance

  $X_i =$ 1960 poverty index $\quad (c = 59.1984)$

  $Z_i =$ see database.

- **Potential outcomes**:

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

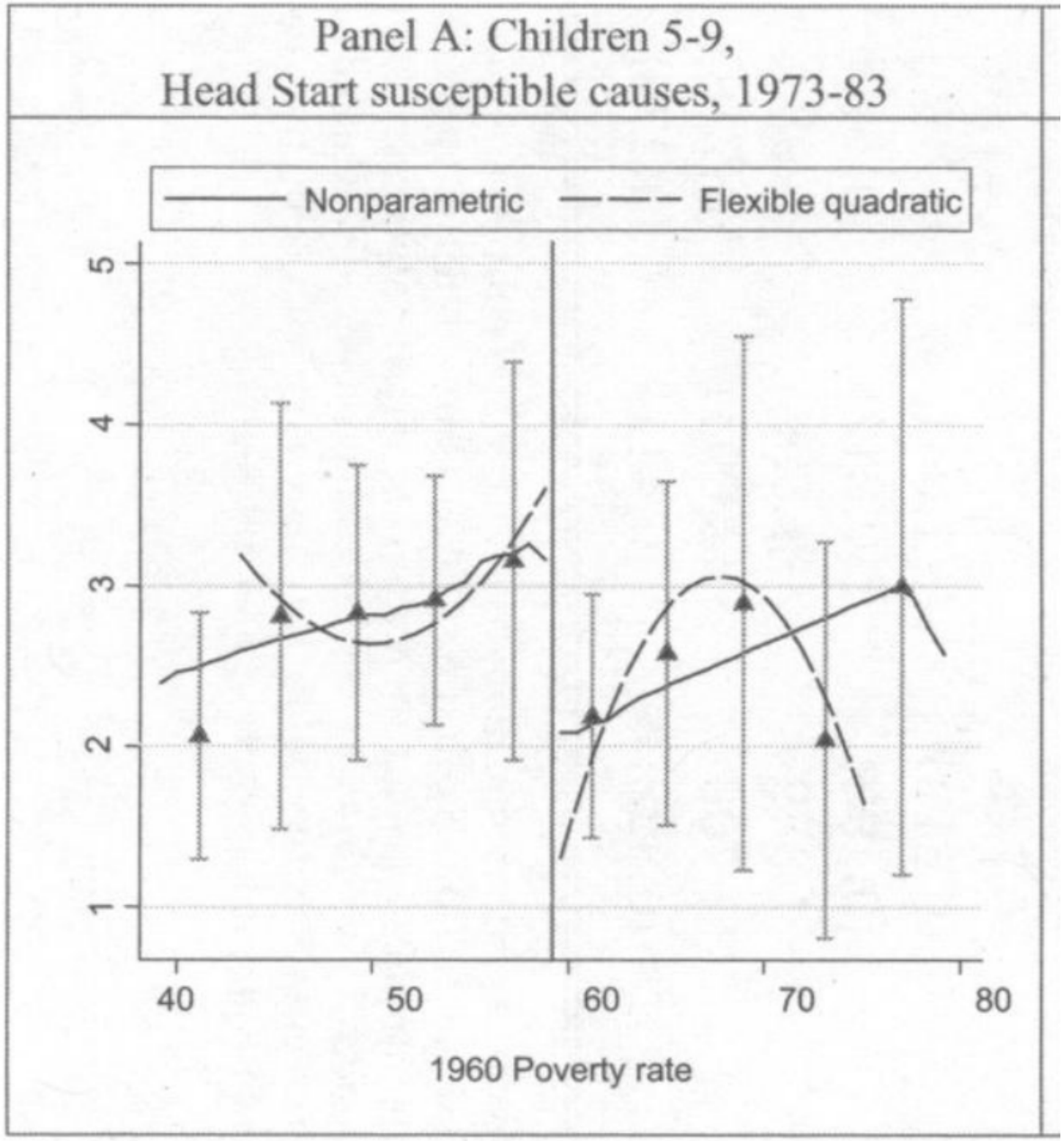
  $Y_i(1) =$ child mortality if **had received** Head Start

- **Causal Inference**:

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

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

Effect of Head Start Assistance on Child Mortality



Part 4:

RD Designs: Local Polynomial Analysis

RD software: https://rdpackages.github.io

- rdrobust: estimation, inference and graphical presentation using local polynomials, partitioning, and spacings estimators.

  rdrobust: RD inference (point estimation and CI; classic, BC, robust).

  rdbwselect: bandwidth or window selection (MSE, CE, etc.).

  rdplot: plots data with “optimal" bin length.

- rddensity: test continuity of density at cutoff using novel local polynomial estimation method. Main command: rddensity.
- rdlocrand: covariate balance, binomial tests, randomization inference methods for window selection & inference.

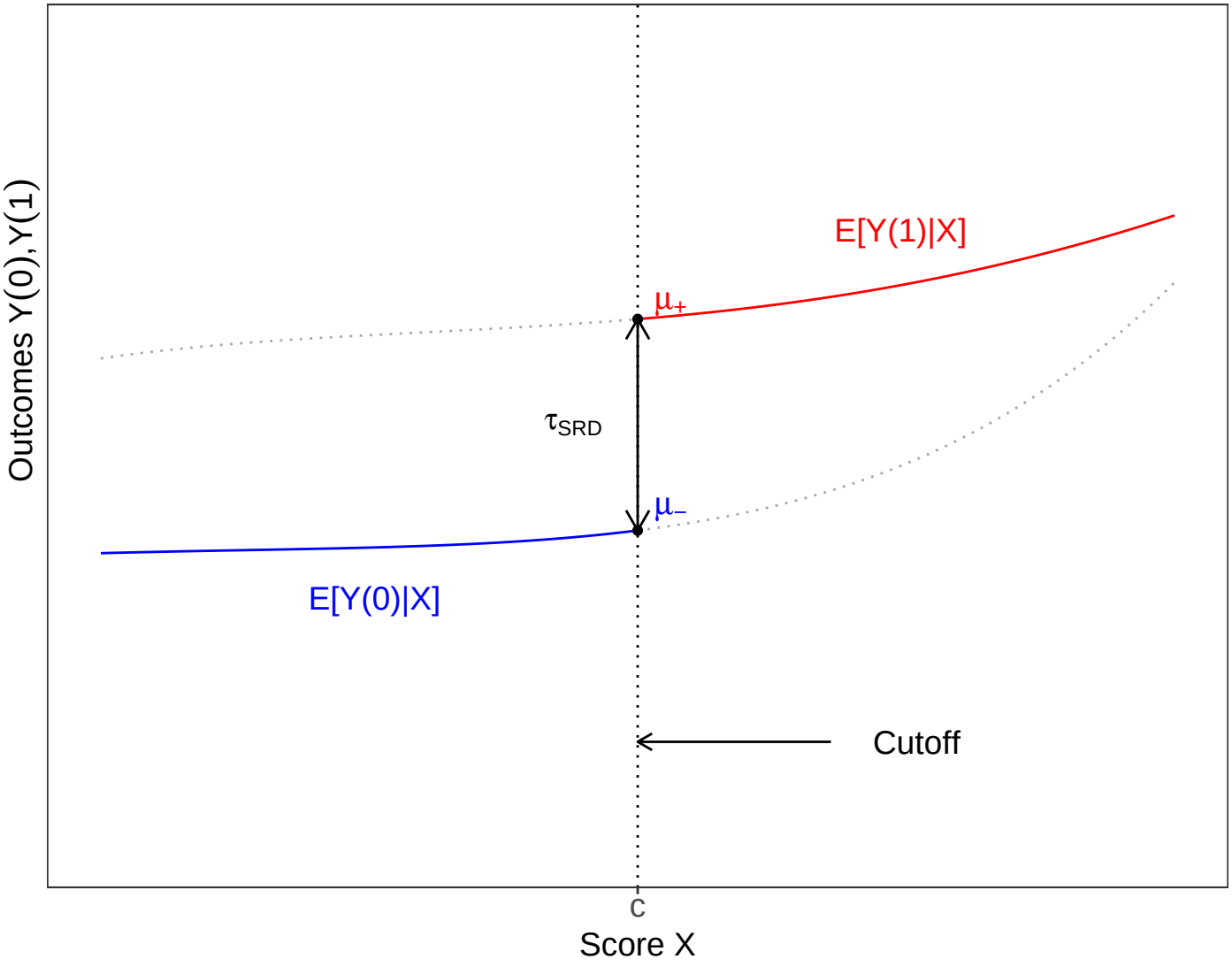
  rdrandinf: inference using randomization inference methods.

  rdwinselect: falsification testing and window selection.

  rdrbounds; rdsensitivity: Rosenbaum bounds and sensitivity analyses

- rdpower: power calculation and sample selection for local polynomial methods.
- rdmulti: RD plots, estimation, inference, and extrapolation with multiple cutoffs and multiple scores.

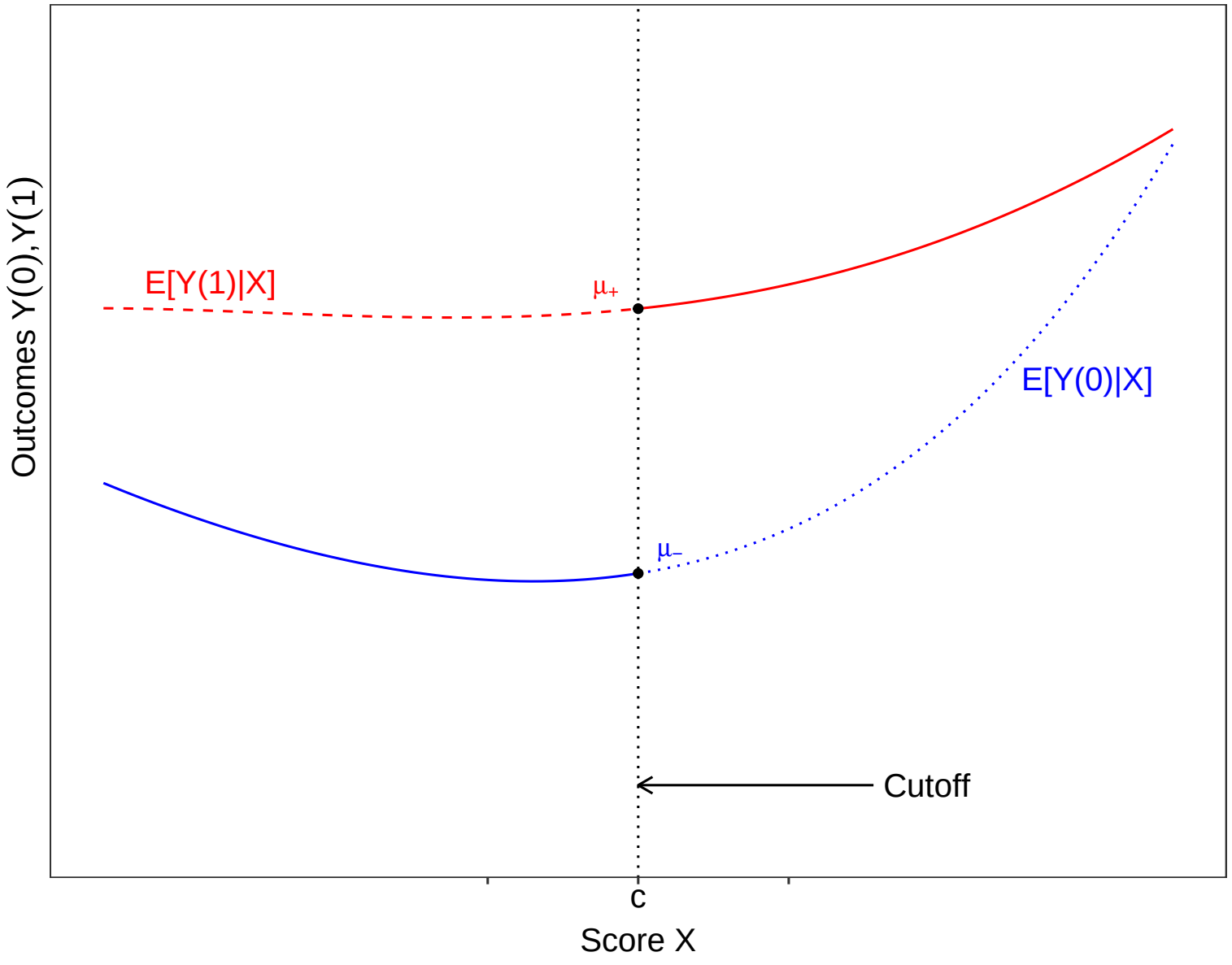
$$\tau = \underbrace{\mathbb{E}[Y_{1i} - Y_{0i} | X_i = x_0]}_{\text{Unobservable}} = \underbrace{\lim_{x \downarrow x_0} \mathbb{E}[Y_i | X_i = x]}_{\text{Observable}} - \underbrace{\lim_{x \uparrow x_0} \mathbb{E}[Y_i | X_i = x]}_{\text{Observable}}$$



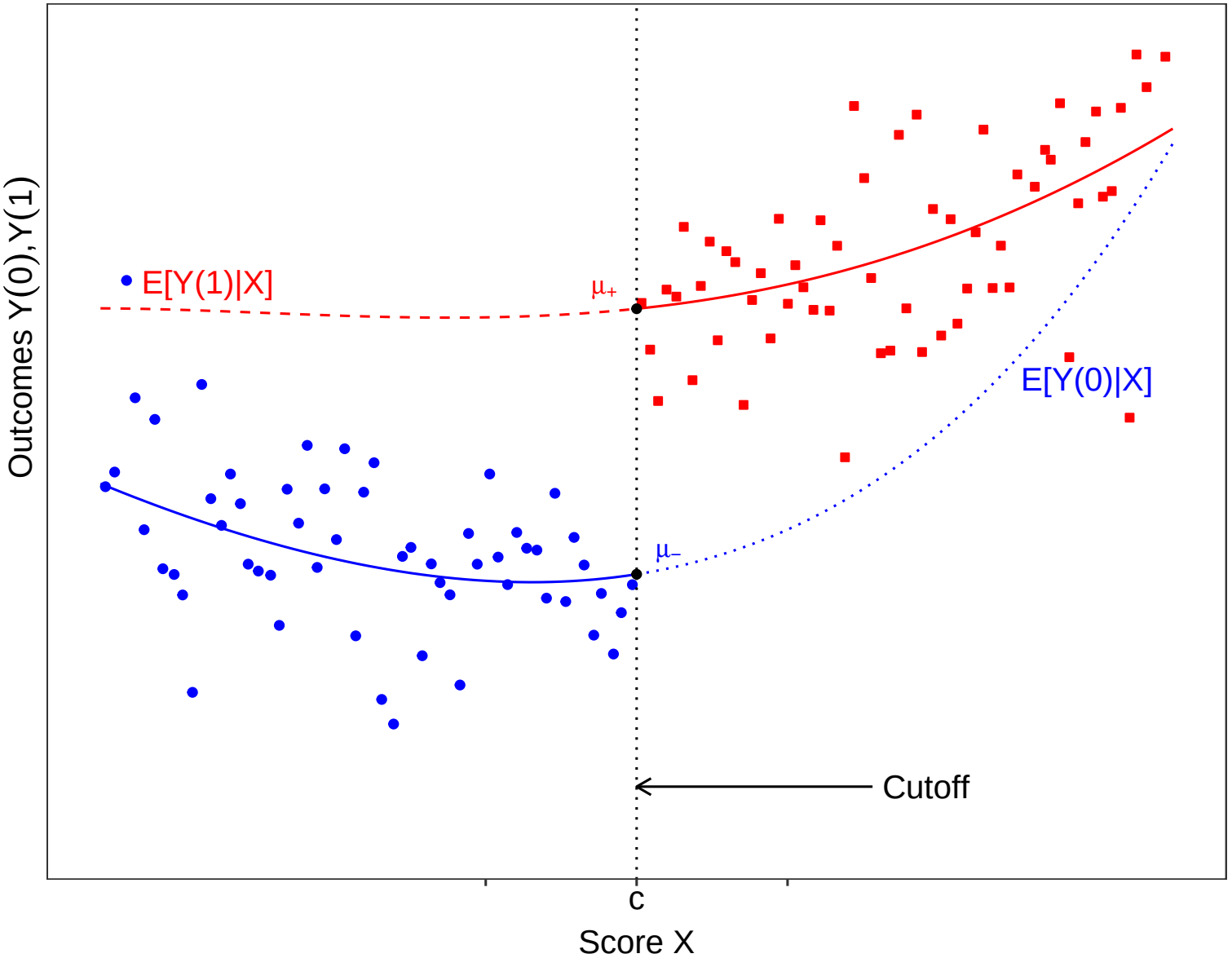
Standard RD: Treatment Effect Estimation

- $\mathbb{E}[Y_i|X_i = x]$ approximated in neighborhood of x_0 by polynomial function
- Local polynomial estimation:
  - Choose order of polynomial p
  - Choose bandwidth h to keep observations in $[x_0 - h, x_0 + h]$
  - Choose kernel function to weigh observations, $w_i = K(\frac{x_i - x_0}{h})$

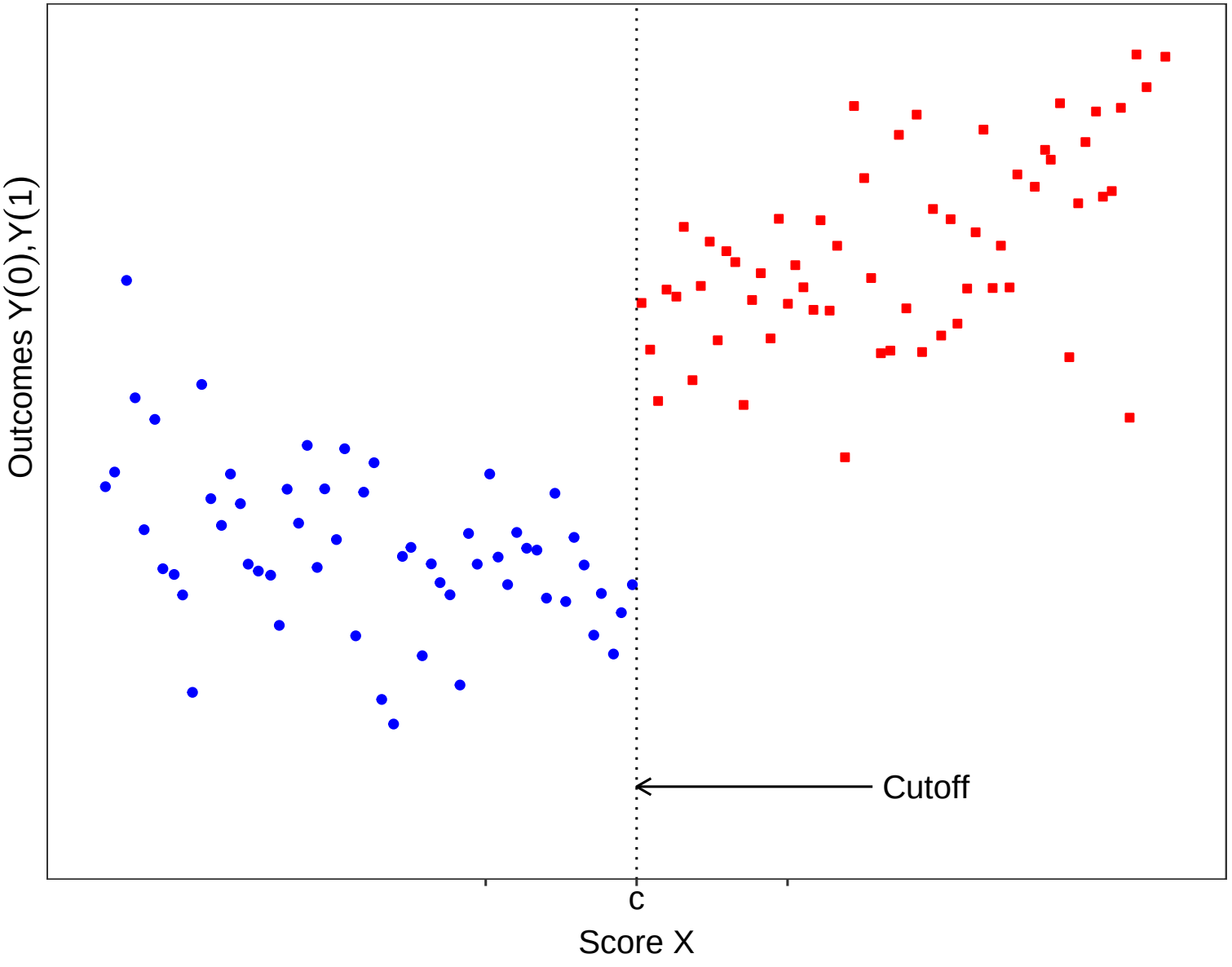
Local Polynomial Estimation



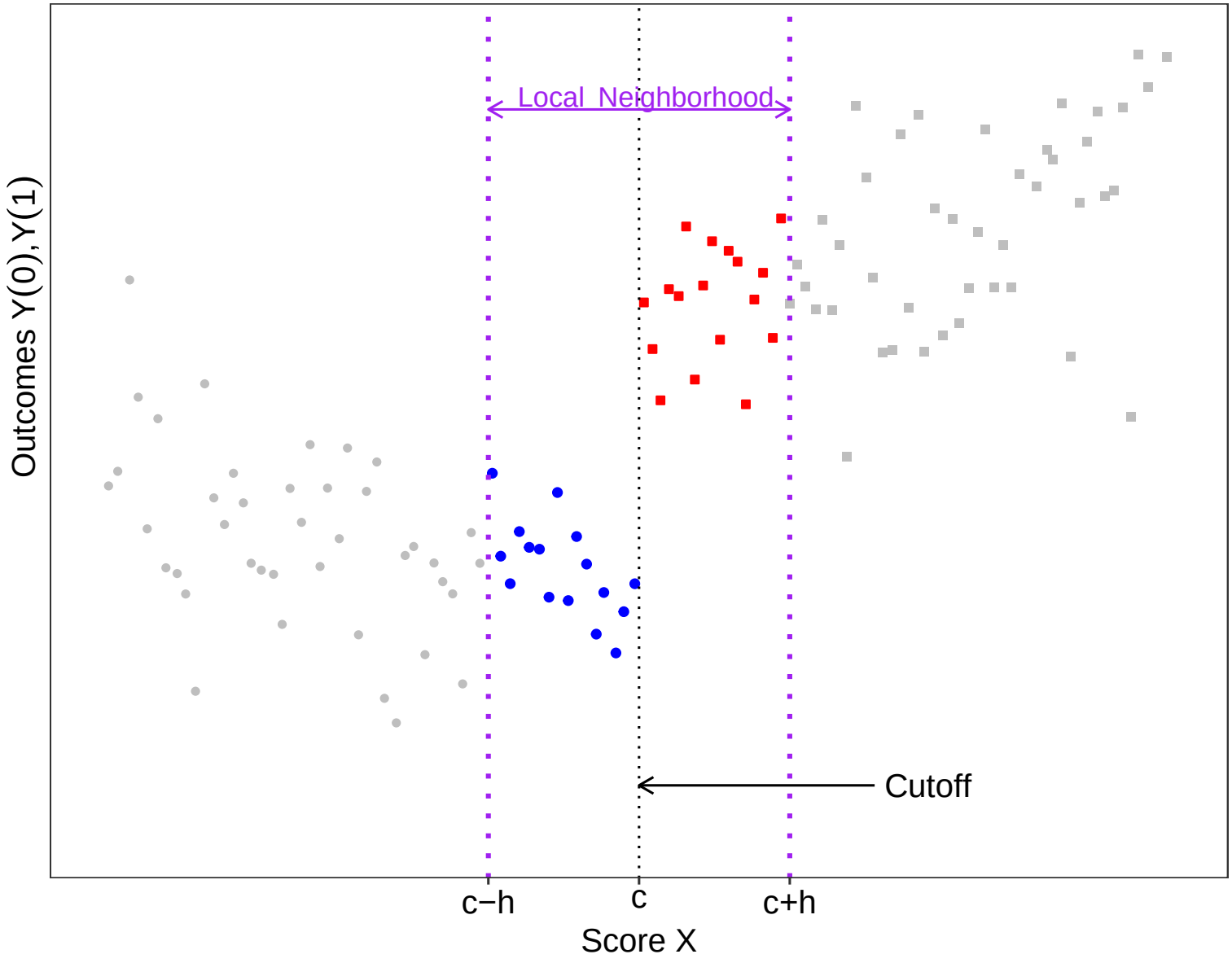
Local Polynomial Estimation



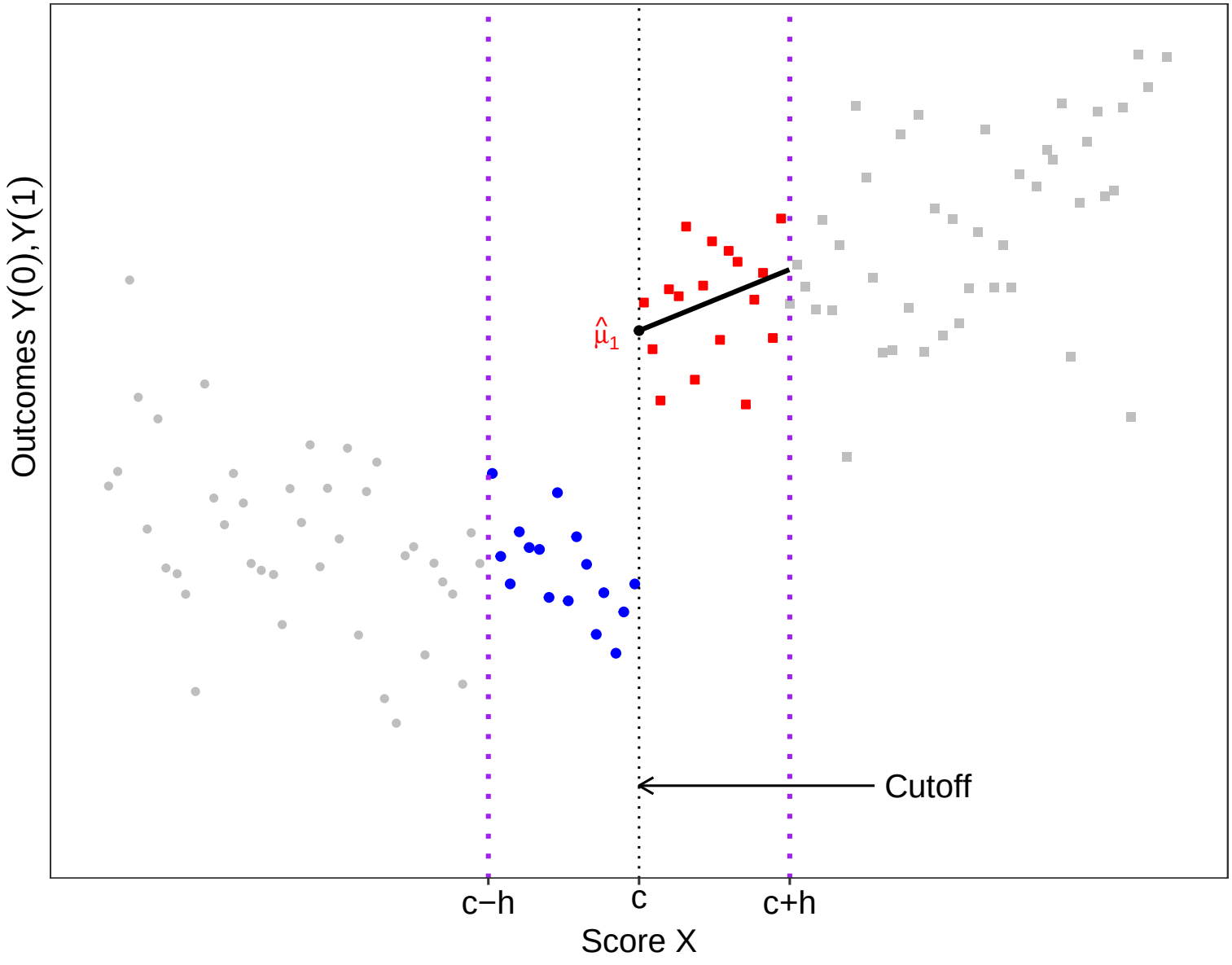
Local Polynomial Estimation



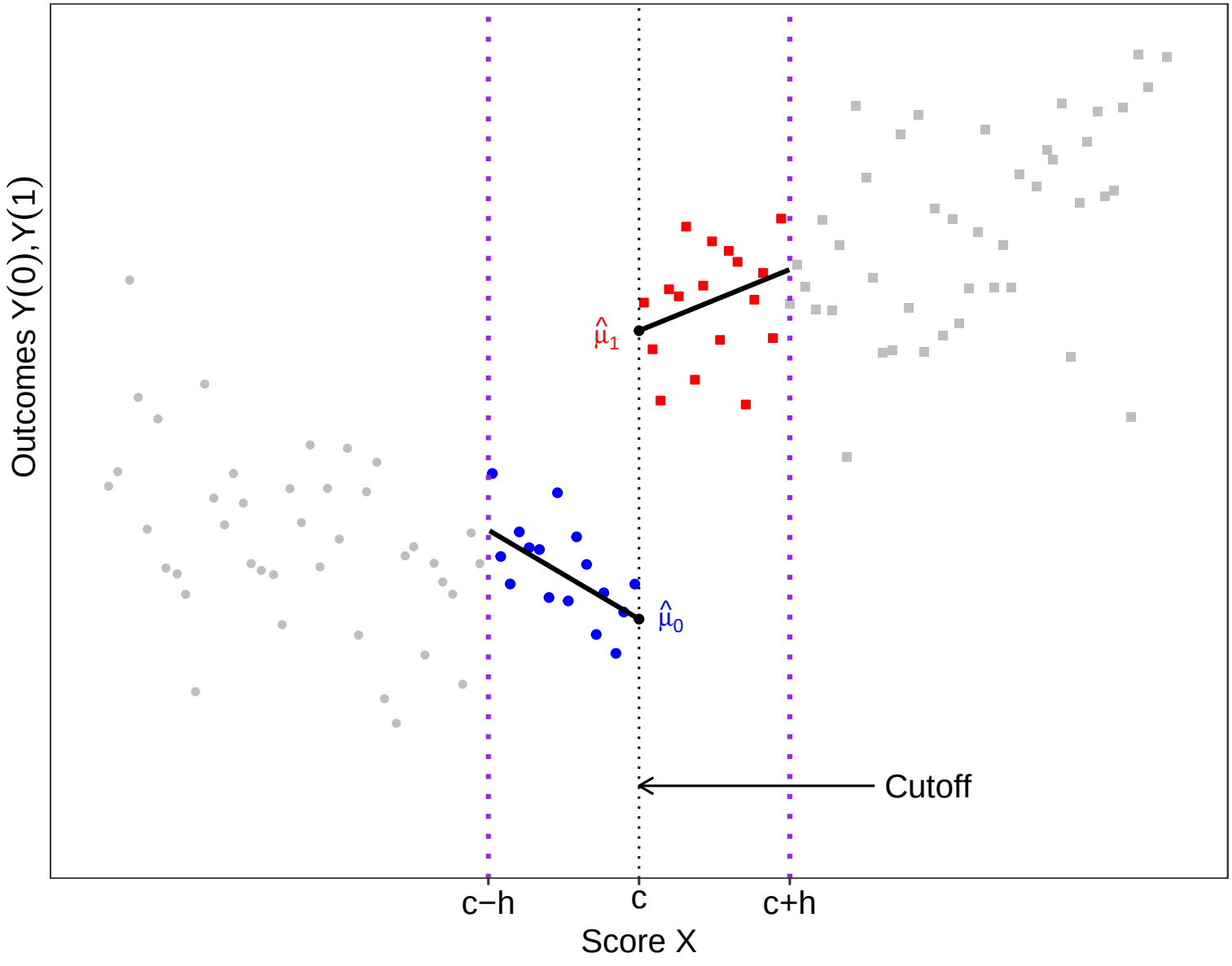
Local Polynomial Estimation



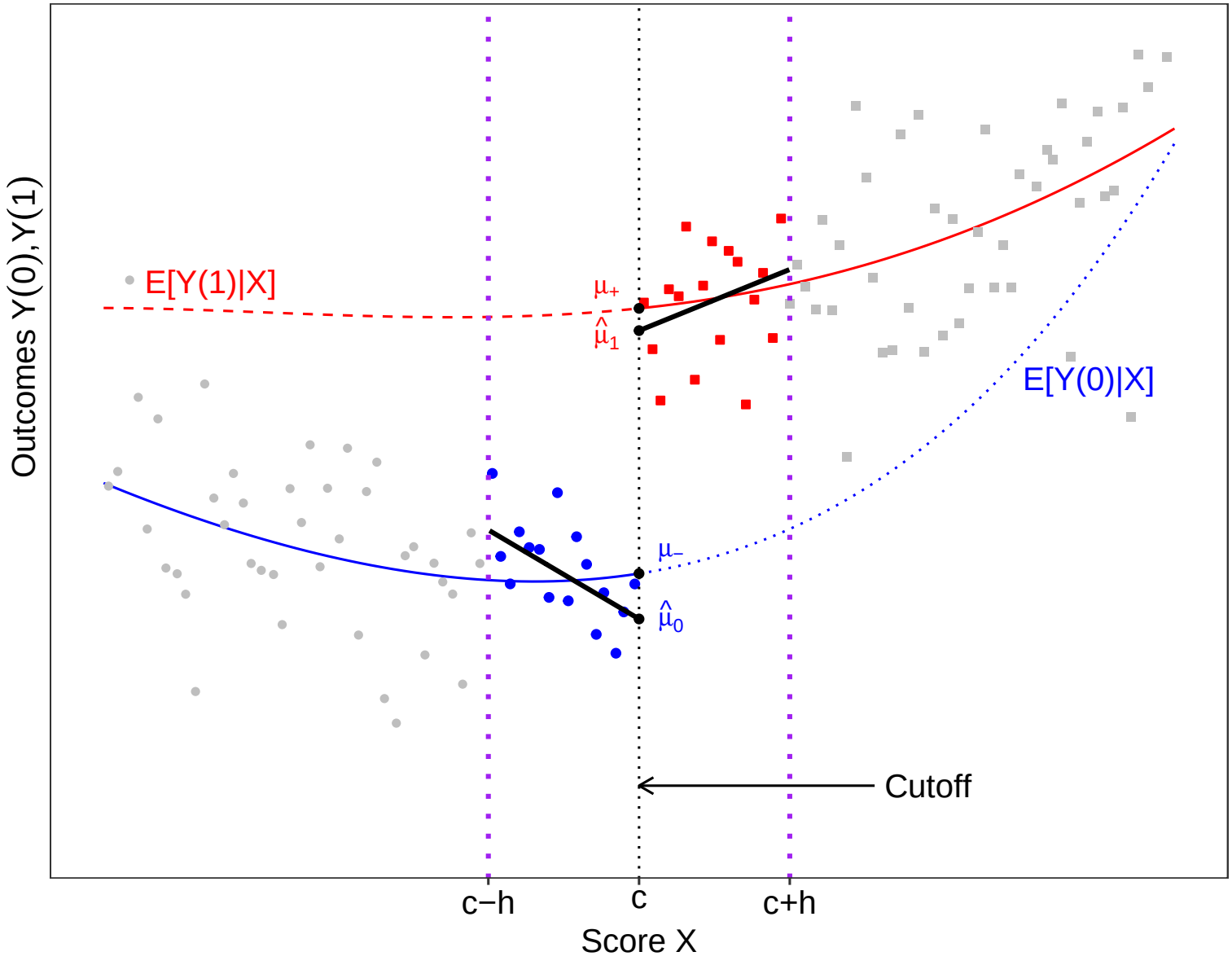
Local Polynomial Estimation



Local Polynomial Estimation



Local Polynomial Estimation



RD Local Polynomial Estimation and Inference

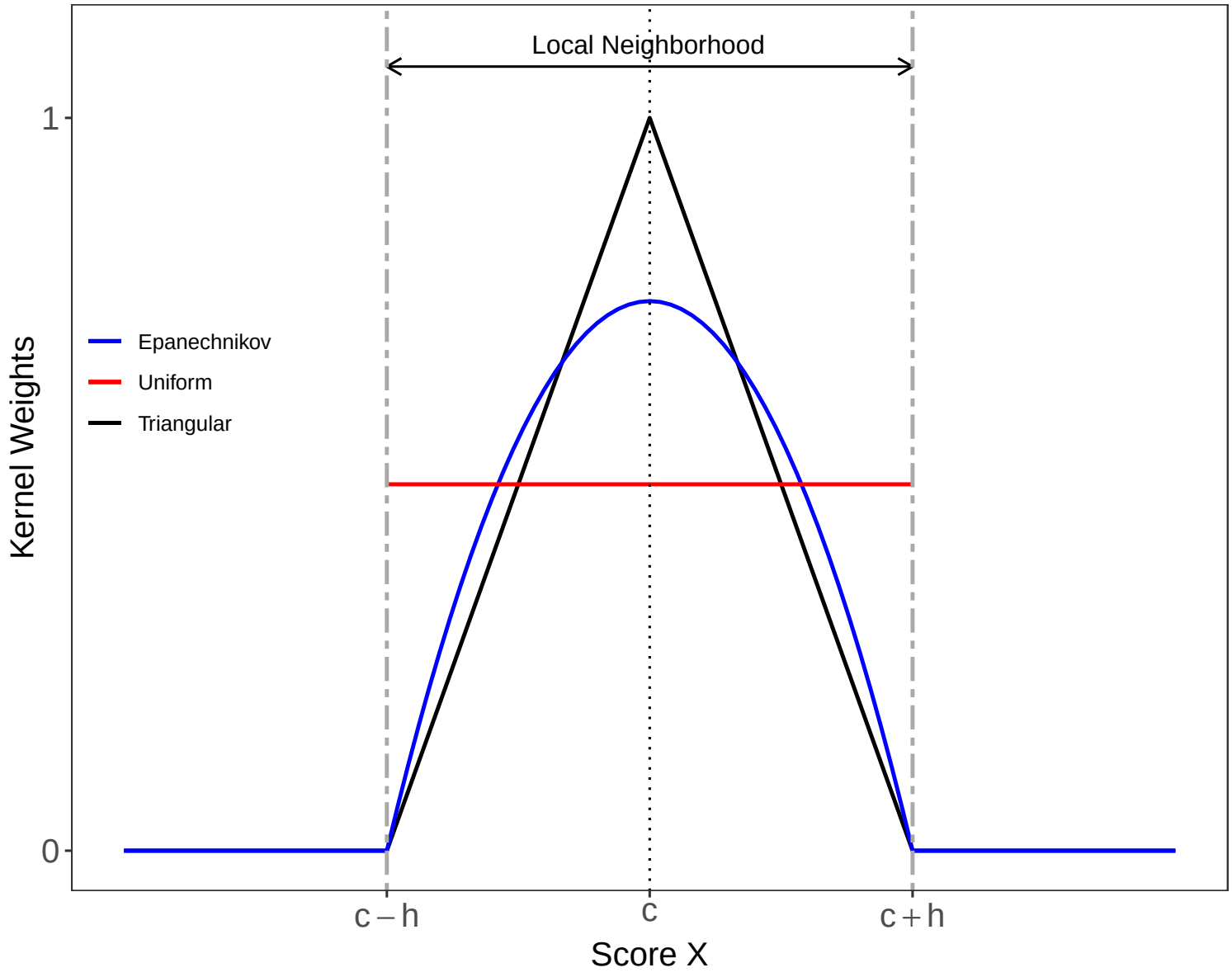
Choose low p and a kernel function $K(\cdot)$

Choose bandwidth h: MSE-optimal or CER-optimal

Construct point estimator $\hat{\tau}_n$ (optimal)

Given above steps, how do we make inferences about τ?

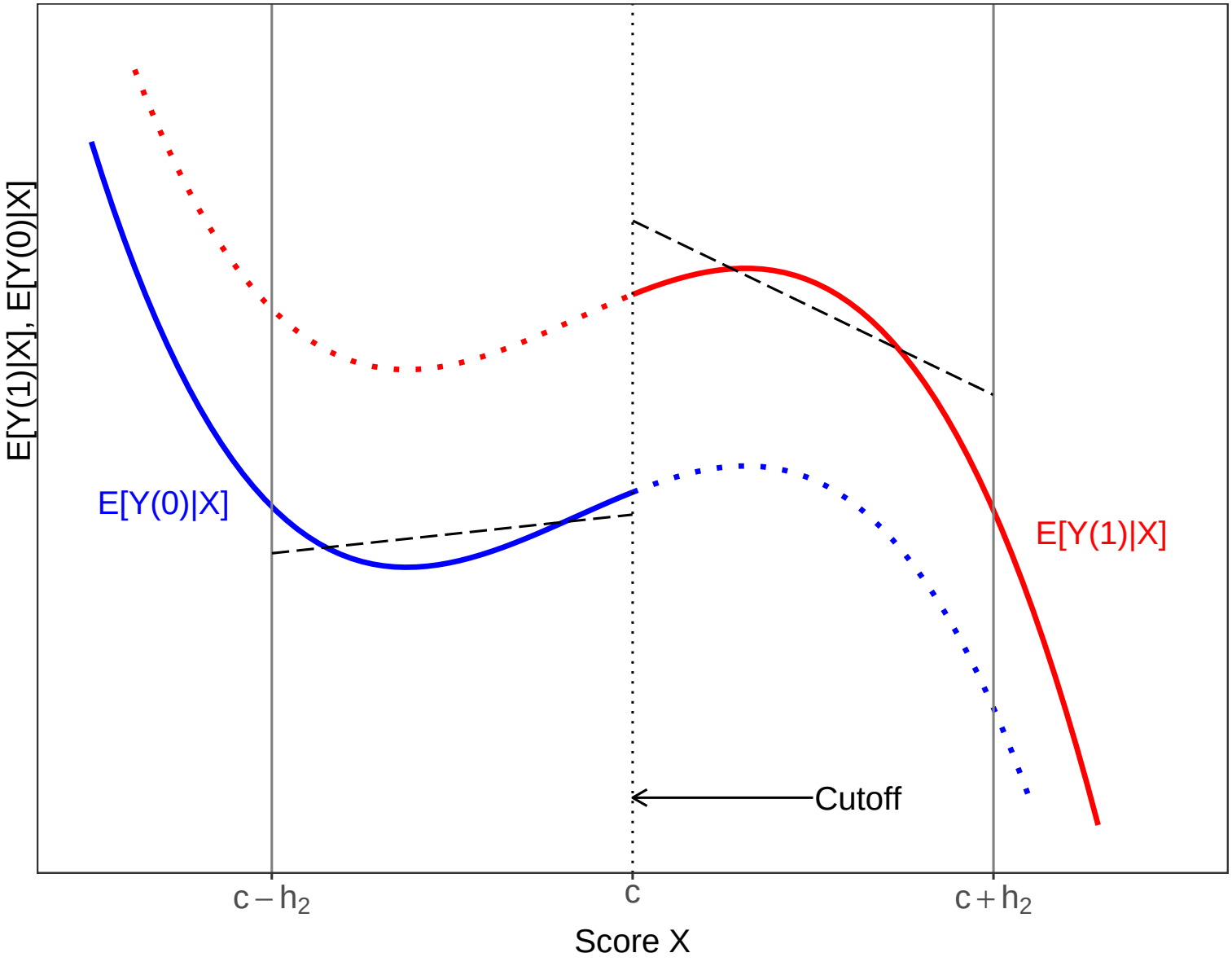
Choice of Kernel Weights



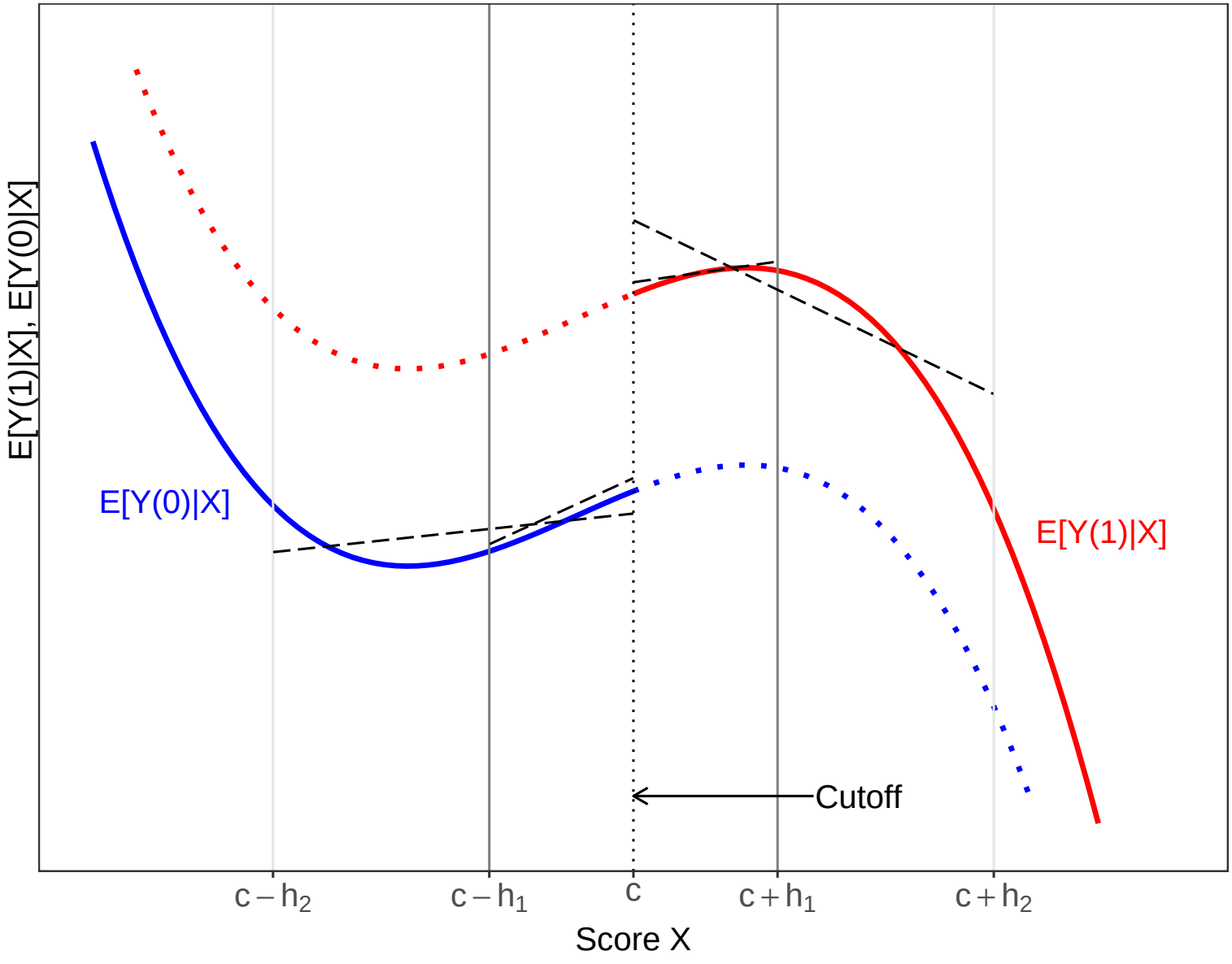
Choice of Polynomial Order (p)

- The higher p, the more flexible the approximation
- However, since approximation is local, p should be low to avoid overfitting
- Given p, approximation can be improved by focusing on a smaller neighborhood around the cutoff
- Standard practice is to choose $p = 1$ ("local linear")

Approximation for fixed $p = 1$



Approximation for fixed $p = 1$



Choice of Bandwidth

- Given p, find h to ensure optimal properties of the point estimator $\hat{\tau_{RD}}$
- MSE-optimal plug-in rule:

$$MSE(\hat{\tau_{RD}}) = Bias^2 + Variance \approx h^{2(p+1)}\mathcal{B}^2 + \frac{1}{nh}\mathcal{V}$$

$$h_{\texttt{MSE}} = C_{\texttt{MSE}}^{1/(2p+3)} \cdot n^{-1/(2p+3)} \qquad C_{\texttt{MSE}} = C(K) \cdot \frac{\mathsf{Var}(\hat{\tau}_{\texttt{SRD}})}{\mathsf{Bias}(\hat{\tau}_{\texttt{SRD}})^2}$$

- **Key idea**: trade-off bias and variance of point estimator $\hat{\tau}$

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

- Kernel function gives higher weight to observations close to cutoff.
- In the context of MSE-optimal h, triangular kernel is optimal

Choice of Bandwidth

- Coverage Error Rate (CER) optimal plug-in rule,

$$h_{\texttt{CER}} = n^{-\frac{p}{(3+p)(3+2p)}} \times h_{\texttt{MSE}}$$

- Key idea: choose optimal bandwidth rate to minimize coverage error of the RBC confidence intervals.

Conventional Local Polynomial Point Estimation

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

$$-h_n \leq X_i < c : \qquad Y_i = \alpha_0 + (X_i - c) \cdot \beta_0 + \varepsilon_{0,i}$$

$$c \leq X_i \leq h_n : \qquad Y_i = \alpha_1 + (X_i - c) \cdot \beta_1 + \varepsilon_{1,i}$$

- RD effect: $\hat{\tau}_n = \hat{\alpha}_1 - \hat{\alpha}_0$

- When choosing MSE-optimal h, this point estimator $\hat{\tau}_n$ is optimal (also consistent)

Conventional Local Polynomial RD Inference

- RD effect: $\hat{\tau}_n = \hat{\alpha}_1 - \hat{\alpha}_0$
- Once $\hat{\tau}_n$ is estimated with optimal h, we might be tempted to use conventional (OLS) inference
- Construct usual t-statistic. For $\mathsf{H}_0 : \tau = 0$,

$$\mathsf{T} = \frac{\hat{\tau}_n}{\sqrt{\mathsf{V}_n}} = \frac{\hat{\alpha}_1 - \hat{\alpha}_0}{\sqrt{\mathsf{V}_{1,n} + \mathsf{V}_{0,n}}} \to_{\mathrm{d}} \mathcal{N}(0,1)$$

- 95% Confidence interval:

$$\mathsf{CI} = \left[\hat{\tau}_n \ \pm \ 1.96 \cdot \sqrt{\mathsf{V}_n} \right]$$

Conventional Local Polynomial RD Inference

- However, with conditions on $h_n \to 0$, the distributional approximation

$$\mathsf{T} = \frac{\hat{\tau}_n}{\sqrt{\mathsf{V}_n}} \to_{\mathrm{d}} \mathcal{N}(\mathsf{B}_n, 1) \quad \neq \mathcal{N}(0,1)$$

  - Bias B_n in RD point estimator captures "curvature" of regression functions
  - In particular, the bias B_n occurs when the MSE-optimal bandwidth is used

- Conventional approach $\to$ assume bias negligible or undersmoothing

$$\mathsf{T} = \frac{\hat{\tau}_n}{\sqrt{\mathsf{V}_n}} \to_{\mathrm{d}} \mathcal{N}(0,1) \quad \Big| \quad \mathsf{CI} = \left[\hat{\tau}_n \pm 1.96 \cdot \sqrt{\mathsf{V}_n} \right]$$

$\Longrightarrow$ Not clear guidance & power loss!

- Bias-correction approach

$$\mathsf{T}^{\mathrm{bc}} = \frac{\hat{\tau}_n - \mathsf{B}_n}{\sqrt{\mathsf{V}_n}} \to_{\mathrm{d}} \mathcal{N}(0,1) \quad \Big| \quad \mathsf{CI}^{\mathrm{bc}} = \left[\left(\hat{\tau}_n - \hat{\mathsf{B}}_n \right) \pm 1.96 \cdot \sqrt{\mathsf{V}_n} \right]$$

$\Longrightarrow$ Poor finite sample properties!

Robust Local Polynomial Inference

- Key observation: $\hat{\mathsf{B}}_n$ is constructed to estimate leading bias

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

- **Our robust approach** $\to$ *Non-standard Asymptotics*

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

- Robust Bias-Correction Approach:

$$\mathsf{T}^{\mathtt{rbc}} = \frac{\hat{\tau}_n - \hat{\mathsf{B}}_n}{\sqrt{\mathsf{V}_n + \mathsf{W}_n}} \to_d \mathcal{N}(0,1)$$

$$\mathsf{CI}^{\mathtt{rbc}} = \left[\left(\hat{\tau}_n - \hat{\mathsf{B}}_n\right) \pm 1.96 \cdot \sqrt{\mathsf{V}_n + \mathsf{W}_n} \right]$$

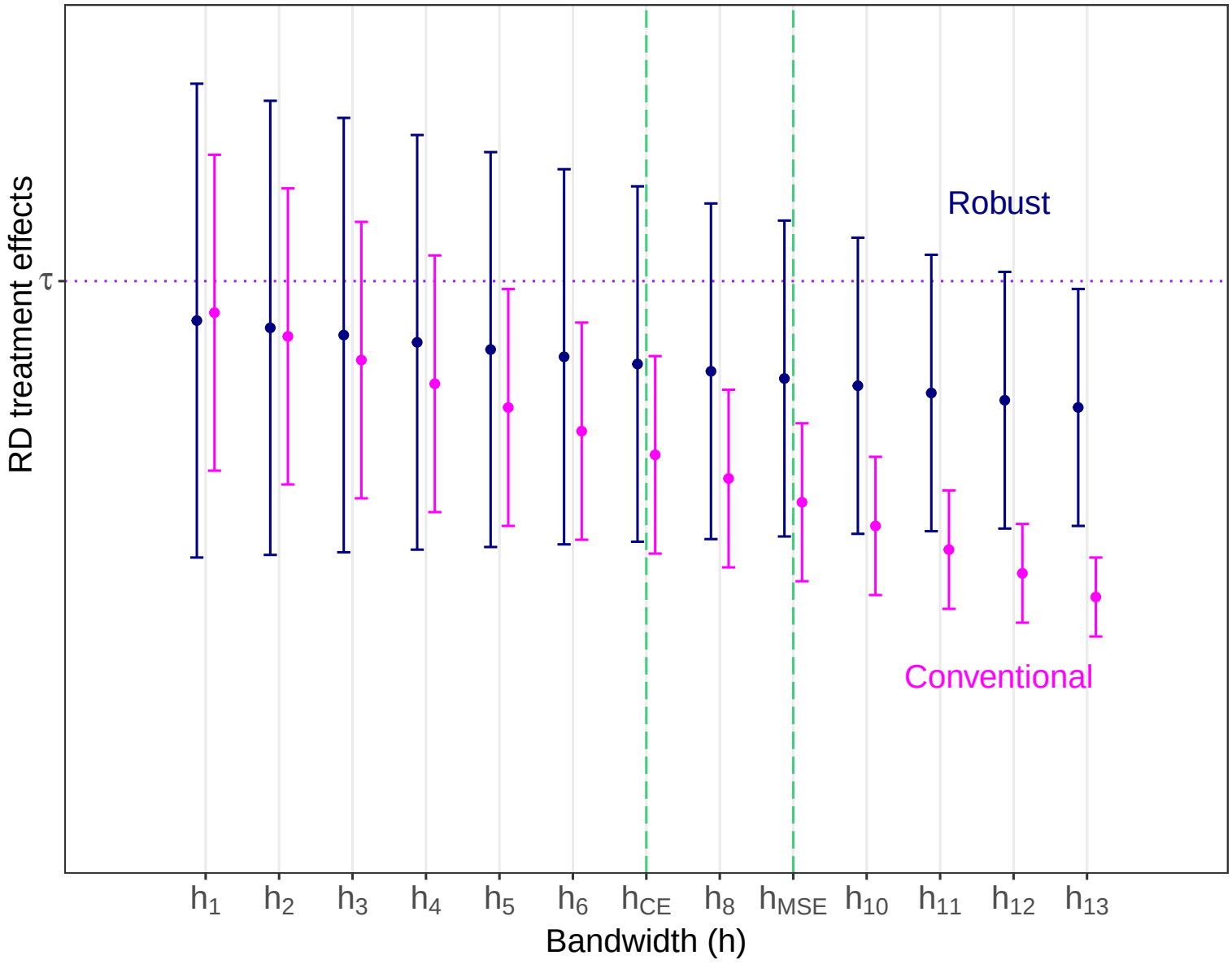
Robust Local Polynomial Inference

- Highlights conceptual distinction between estimation and inference
- Conventional procedure to derive confidence intervals:
  - Derive asymptotic normal distribution of t statistic $\mathsf{T} = (\hat{\beta} - \beta)/\mathsf{std.err}$
  - Build confidence intervals as the dual of T: $\hat{\beta} \pm 1.96 \times \mathsf{std.err}$
- General underlying idea
  - Choose statistic and obtain (asymptotic) distribution
  - Build confidence intervals as collection of all hypotheses not rejected by it
  - Point estimator (if it exists) need not be the center of the confidence interval
- RD robust CI
  - Alternative statistic $(\hat{\tau}_n - \tau - \hat{\mathsf{B}}_n)/\sqrt{\mathsf{V}_n + \mathsf{W}_n}$
  - $\mathsf{CI}^{\mathtt{rbc}} = \left[\left(\hat{\tau}_n - \hat{\mathsf{B}}_n \right) \pm 1.96 \cdot \sqrt{\mathsf{V}_n + \mathsf{W}_n} \right]$
  - Not centered at point estimator $\hat{\tau}_n$, and rescaled (different variance estimator)

Table: Local Polynomial Confidence Intervals

	Centered at	Standard Error
Conventional: $\texttt{CI}_{\texttt{us}}$	$\hat{\tau}_{\texttt{SRD}}$	$\sqrt{\hat{\mathscr{V}}}$
Bias Corrected: $\texttt{CI}_{\texttt{bc}}$	$\hat{\tau}_{\texttt{SRD}} - \hat{\mathscr{B}}$	$\sqrt{\hat{\mathscr{V}}}$
Robust Bias Corrected: $\texttt{CI}_{\texttt{rbc}}$	$\hat{\tau}_{\texttt{SRD}} - \hat{\mathscr{B}}$	$\sqrt{\hat{\mathscr{V}}_{\texttt{bc}}}$

Confidence Intervals for Different Bandwidths



Part 5:

RD Local Randomization Methods

Overview

- Local randomization and randomization inference 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

Recap: Continuity-based Approach

- Assume regression functions are continuous to obtain

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

  - Approximates regression function and relies on continuity assumptions.
  - *Requires*: choosing weights, bandwidth and polynomial order.

- Alternative: local randomization approach

Analogies with experiments

- Lee (2008): RD design can be as credible as a randomized experiment for units very near cutoff

- Imagine that score depends on each unit's unobservables characteristics and choices

- If the two following conditions hold:

  there is a random chance element to score that unit receives

  probability of this random "error" doesn't change abruptly at cutoff

- Then the RD design can be seen as an experiment:

  units barely above the cutoff as-if randomly assigned to treatment

  units barely above the cutoff as-if randomly assigned to control

- This fails if individuals have ability to exactly control their score

Analogies with experiments

- Consider an RD Design where:

  treatment is assigned based on score exceeding cutoff

  units lack ability to manipulate score (continuity holds)

- Crucial distinction:

  Experiment $\rightarrow$ no need to make assumptions about shape of the average potential outcomes

  RD design $\rightarrow$ inferences depend crucially on assumptions regarding functional form of regression functions

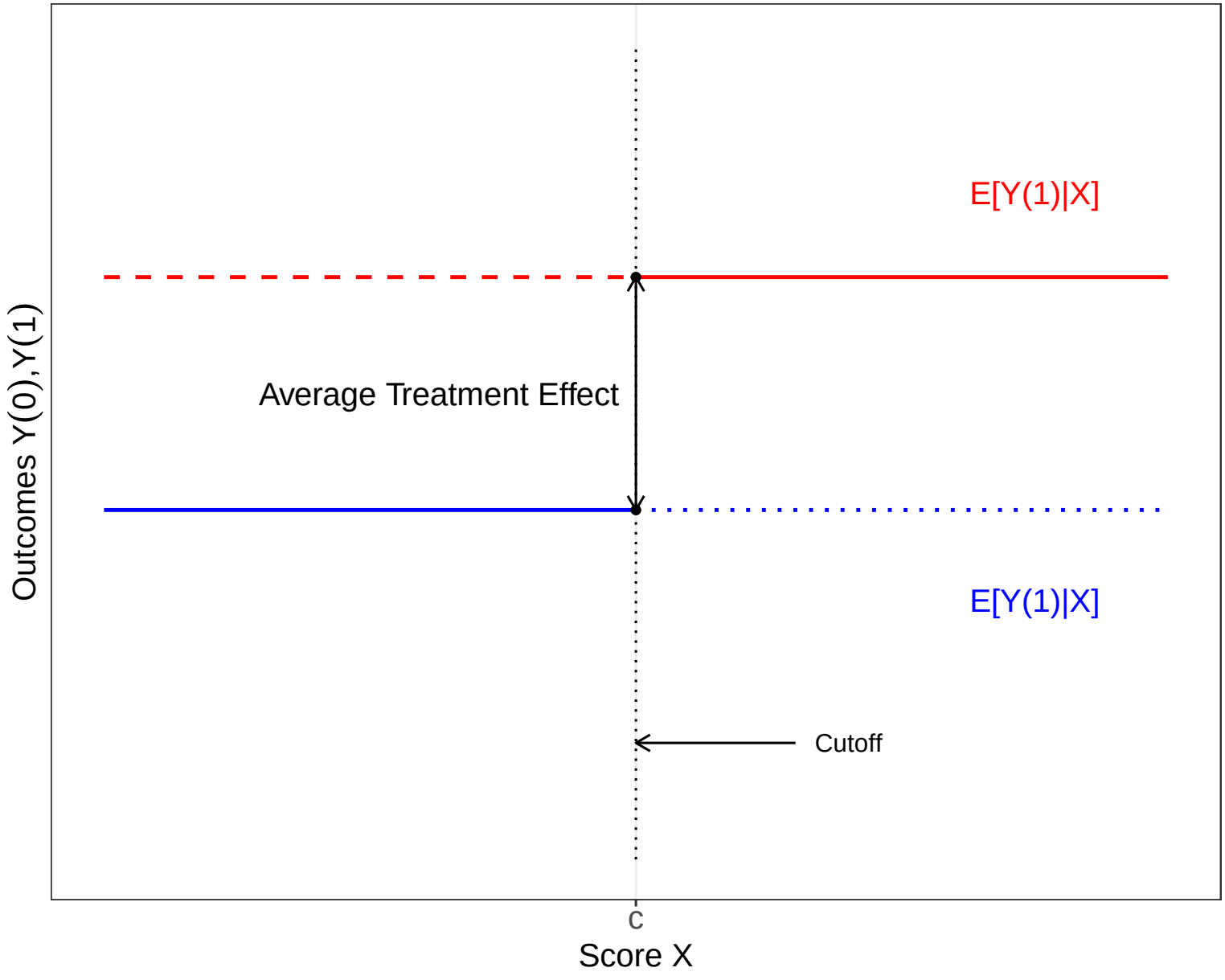
- Any experiment can be recast as an RD design where

  score is a uniform random variable

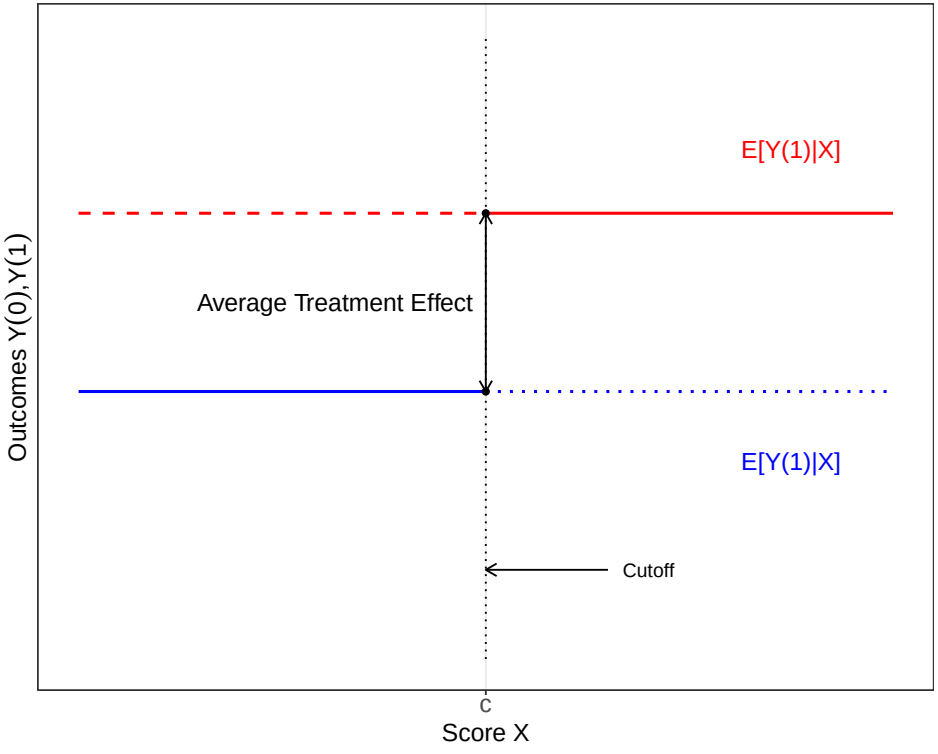
  cutoff chosen to ensure a given probability of treatment

  Ex: each student assigned uniform random number between 0 and 100, scolarship given to students whose score is above 50

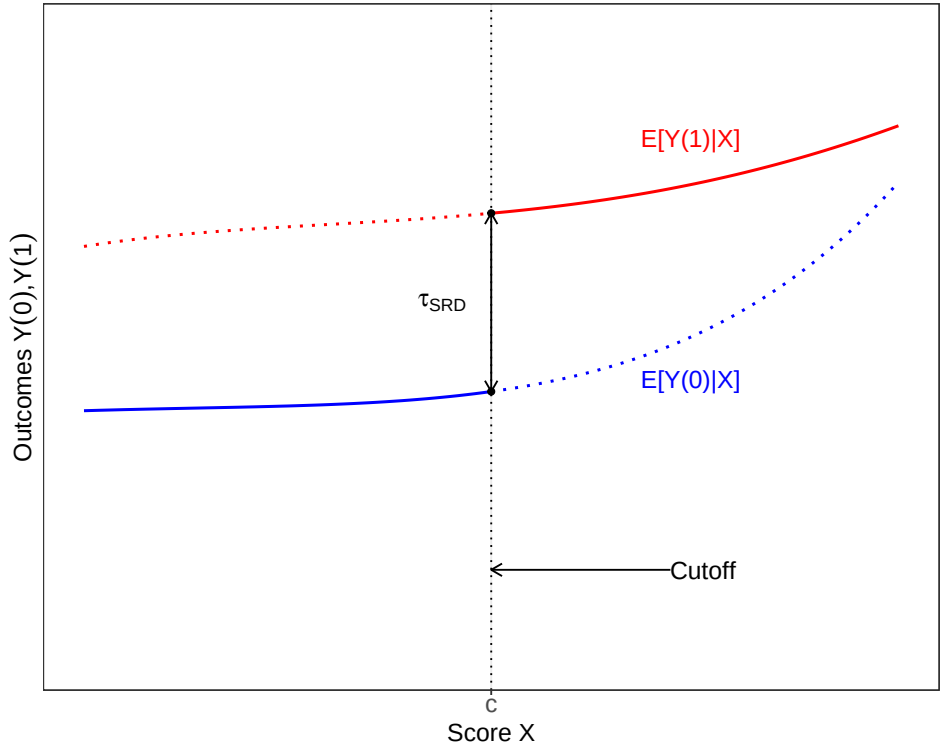
Randomized Experiment



Experiment versus RD Design

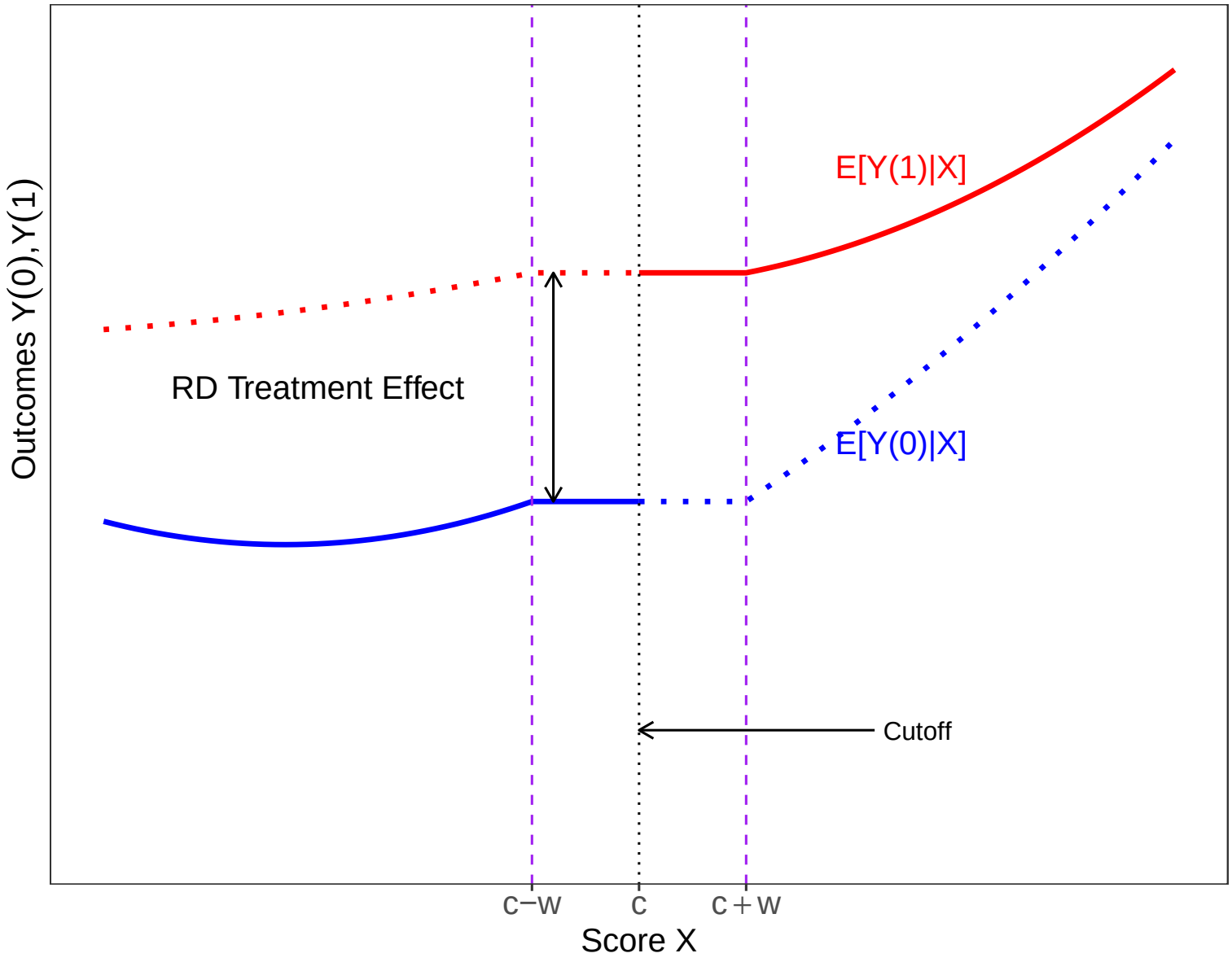


(a) Randomized Experiment



(b) RD Design

If as-if random interpretation is true: Local Randomization RD



Local Randomization Approach to RD Design

- Gives an alternative that can be used as a robustness check.
- **Key assumption**: exists window $W = [-w, w]$ around cutoff $(-w < c < w)$ where (assuming random potential outcomes)

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

- Thus, inside W_0 subjects are as-if randomly assigned to either side of cutoff
  - The distribution of running variable same for all units inside W_0
  - Potential outcomes in W_0 depend on running variable only through threshold indicators within W_0
- Stronger than Continuity-Based Approach$\Rightarrow$ Relevant population functions are not only continuous at x_0, but also completely unaffected by the running variable in W_0

Local Randomization Approach to RD Design

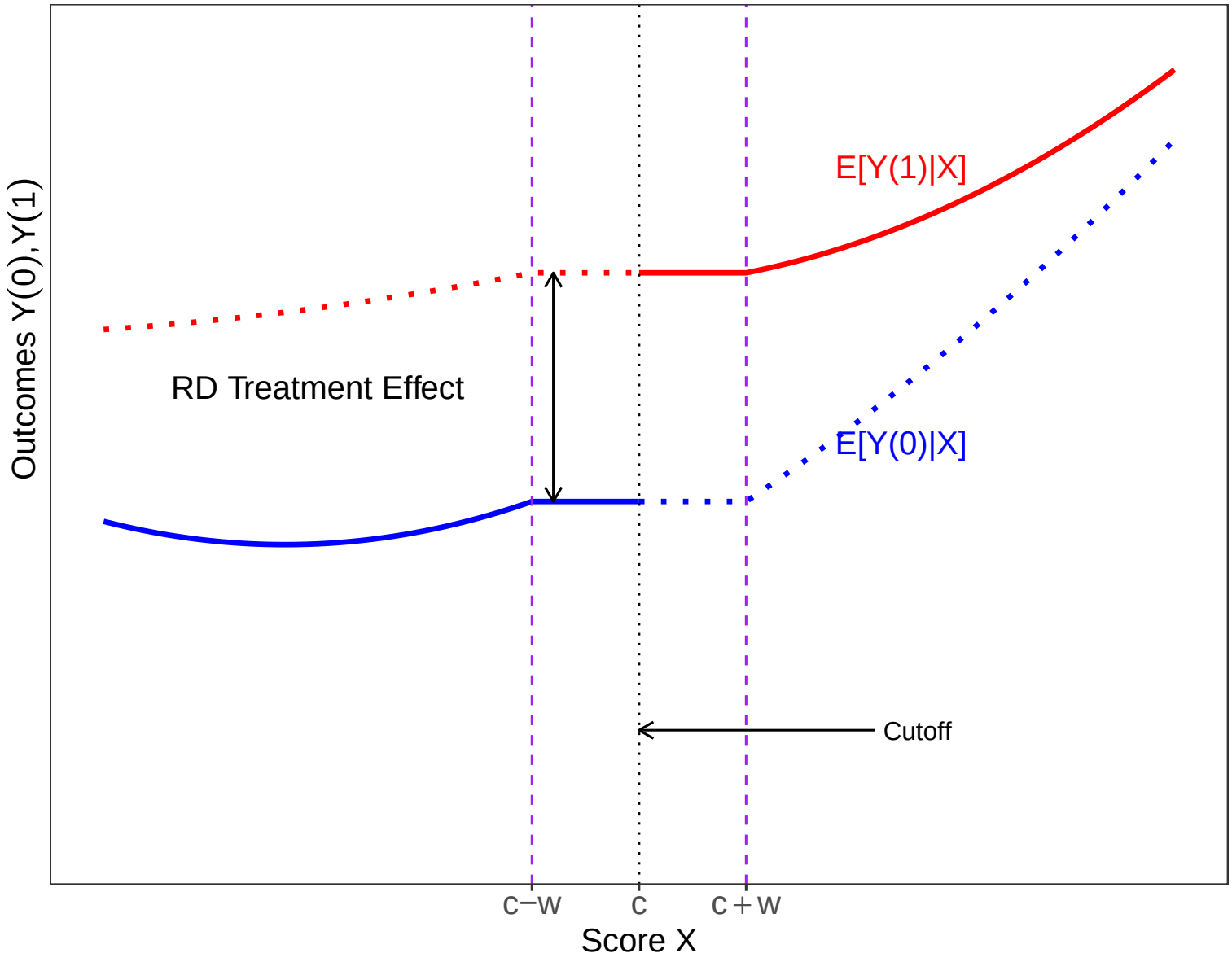
Under Fisherian framework, with $W_0 = [c - w, c + w]$, local randomization assumption is:

- The distribution of the running variable in the window W_0, $F_{X_i|X_i \in W_0}(x)$, is known, is the same for all units, and does not depend on the potential outcomes: $F_{X_i|X_i \in W_0}(x) = F(x)$

- Inside W_0, the potential outcomes depend on the running variable solely through the treatment indicator $T_i = \mathbb{1}(X_i \geq c)$ but not directly: $Y_i(X_i, T_i) = Y_i(T_i)$ for all i such that $X_i \in W_0$

Under both conditions, inside W_0, placement above/below cutoff is unrelated to potential outcomes and potential outcomes unrelated to score

If as-if random interpretation is true: Local Randomization RD



Local Randomization Approach to RD Design

- In window W_0, subjects randomly assigned to either side of cutoff:
  - Window W_0
  - Assignment mechanism
- If assignment mechanism and W_0 are *known*, RD becomes an experiment in W_0
- If few units inside W_0, adopt a Fisherian setup: potential outcomes are fixed, only randomness is in the assignment of subjects

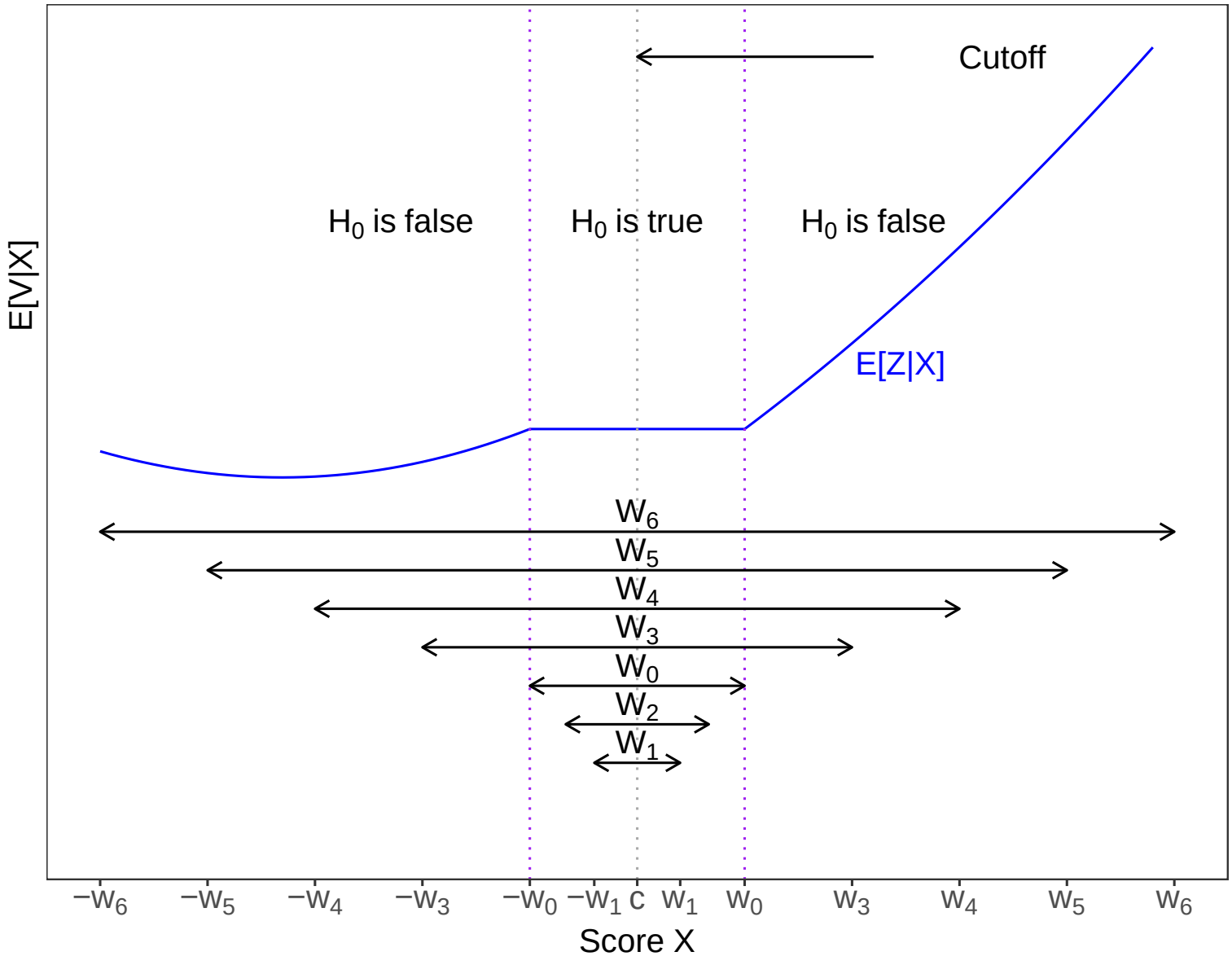
Local Randomization Approach Using Fisherian Methods

- Approach has two steps:
  - Step 1: Choose window around cutoff where randomization holds
  - Step 2: Apply randomization inference tools, given a hypothesized treatment assignment, within W_0

Step 1: Choose the window W_0

- How to choose window?
  - Use balance tests on pre-determined/exogenous covariates.
  - Very intuitive, easy to implement.

Window Selector Based on Covariate Balance in Locally Random RD



Step 2: Use Randomization Inference Tools within W_0

- Under this framework, we can treat observations within the window W_0 as if generated by a randomized experiment
- One possible randomization mechanism:
  - T_i is Bernoulli with parameter π: for all for all vectors $\mathbf{t}$ in Ω_{W_0}, $\Pr\left(\mathbf{T}_{W_0} = \mathbf{t}\right) = \pi^{\mathbf{t}'\mathbf{1}}\left(1-\pi\right)^{(\mathbf{1}-\mathbf{t})'\mathbf{1}}$
  - Since π is unknown, we estimate it $\hat{\pi} = \frac{\mathbf{T}'_{W_0}\mathbf{1}}{n_{W_0}}$
- Another possible randomization mechanism:
  - Fix number of treated units within the window at m_{W_0}, which leads to $\Pr\left(\mathbf{T}_{W_0} = \mathbf{t}\right) = \frac{1}{\binom{n_{W_0}}{m_{W_0}}}$ for all $\mathbf{t} \in \Omega_{W_0}$

Step 2: Use Randomization Inference Tools within W_0

- Given local random assumption, can test sharp null hypothesis of no treatment effect for any i
- Under this hypothesis, observed outcomes are fixed regardless of realization of $\mathbf{T}_{W_0}$: $y_i(\mathbf{t}) = y_i$ for all i within W_0 and for all $\mathbf{t} \in \Omega_{W_0}$
- Thus, the distribution of any test statistic $Q(\mathbf{T}_{W_0}, \mathbf{y}_{W_0})$ is known, since it depends only on the known distribution of $\mathbf{T}_{W_0}$
- One-sided significance level:

$$\Pr\left(Q(\mathbf{T}_{W_0}, \mathbf{y}_{W_0}) \geq Q(\mathbf{t}_{W_0}, \mathbf{y}_{W_0})\right) = \sum_{\mathbf{t} \in \dot{}_W} \mathbf{1}\left(Q(\mathbf{t}, \mathbf{y}_{W_0}) \geq Q(\mathbf{t}_{W_0}, \mathbf{y}_{W_0})\right) \Pr\left(\mathbf{T}_{W_0} = \mathbf{t}\right)$$

- Different test statistics may be used

Empirical Illustration 1: 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 $t \quad (c = 0)$.

  $Z_i =$ covariates (*demvoteshlag1*, *demvoteshlag2*, *dopen*, etc.).

- **Potential outcomes**:

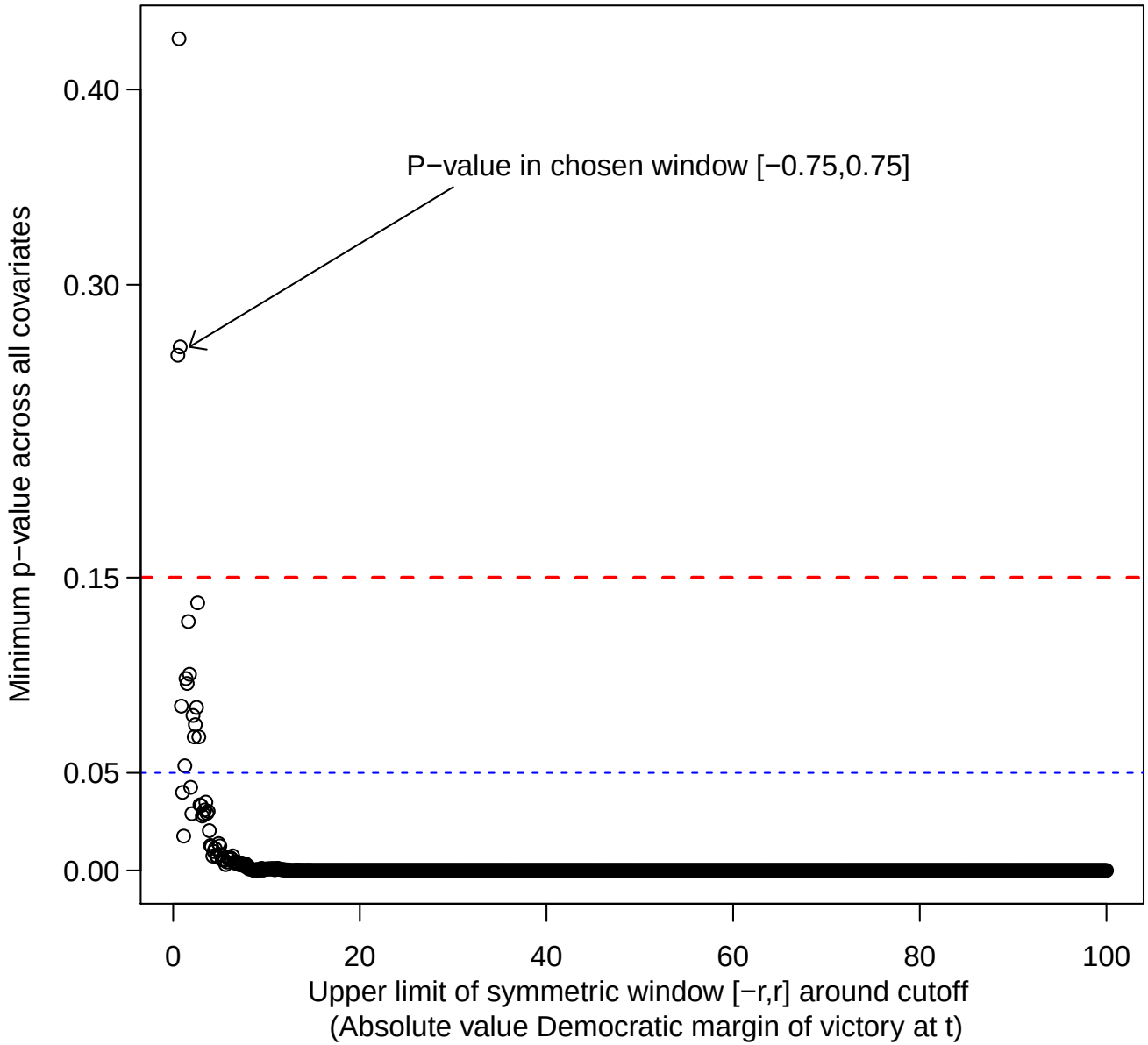
  $Y_i(0) =$ election outcome at $t+1$ if **had not been** incumbent.

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

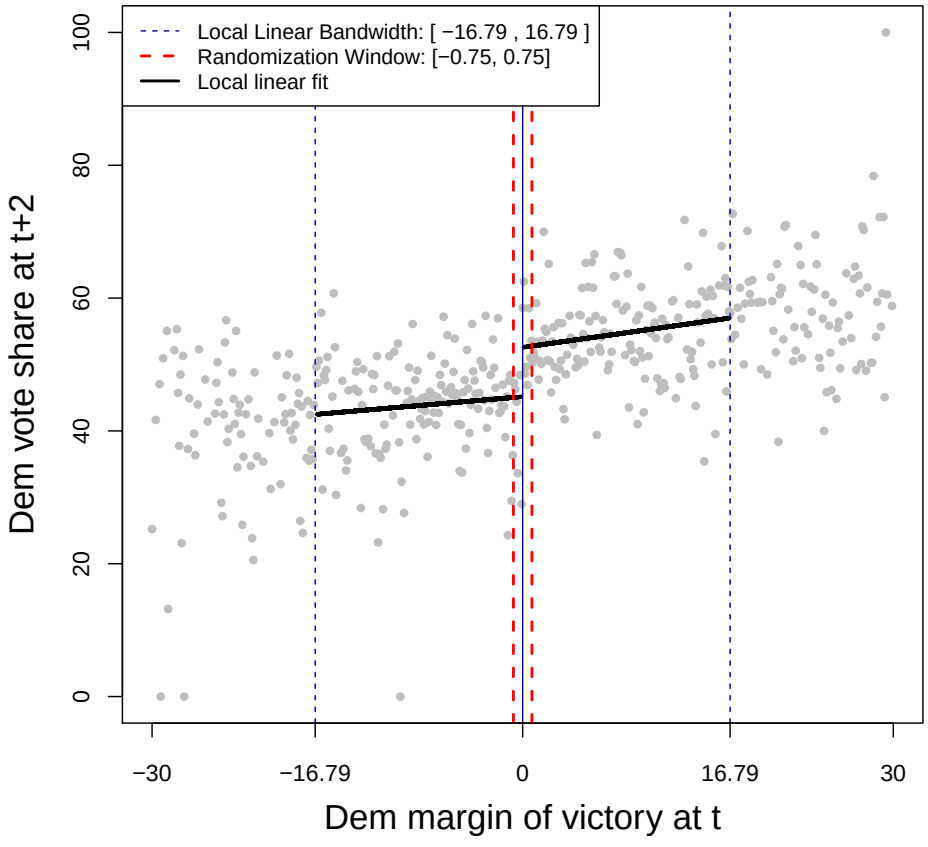
- **Causal Inference**:

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

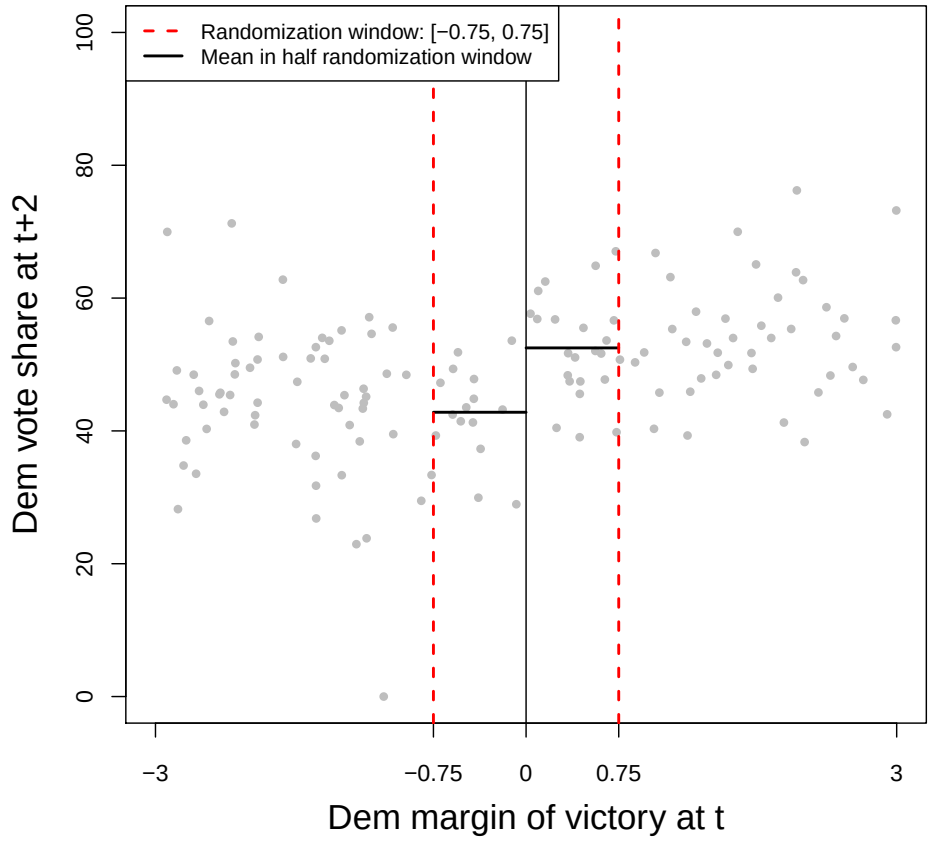
Window Selection Based on Covariates, CFT



Continuity-Based vs Local Randomization Analysis, CFT



(a) Continuity-Based Analysis

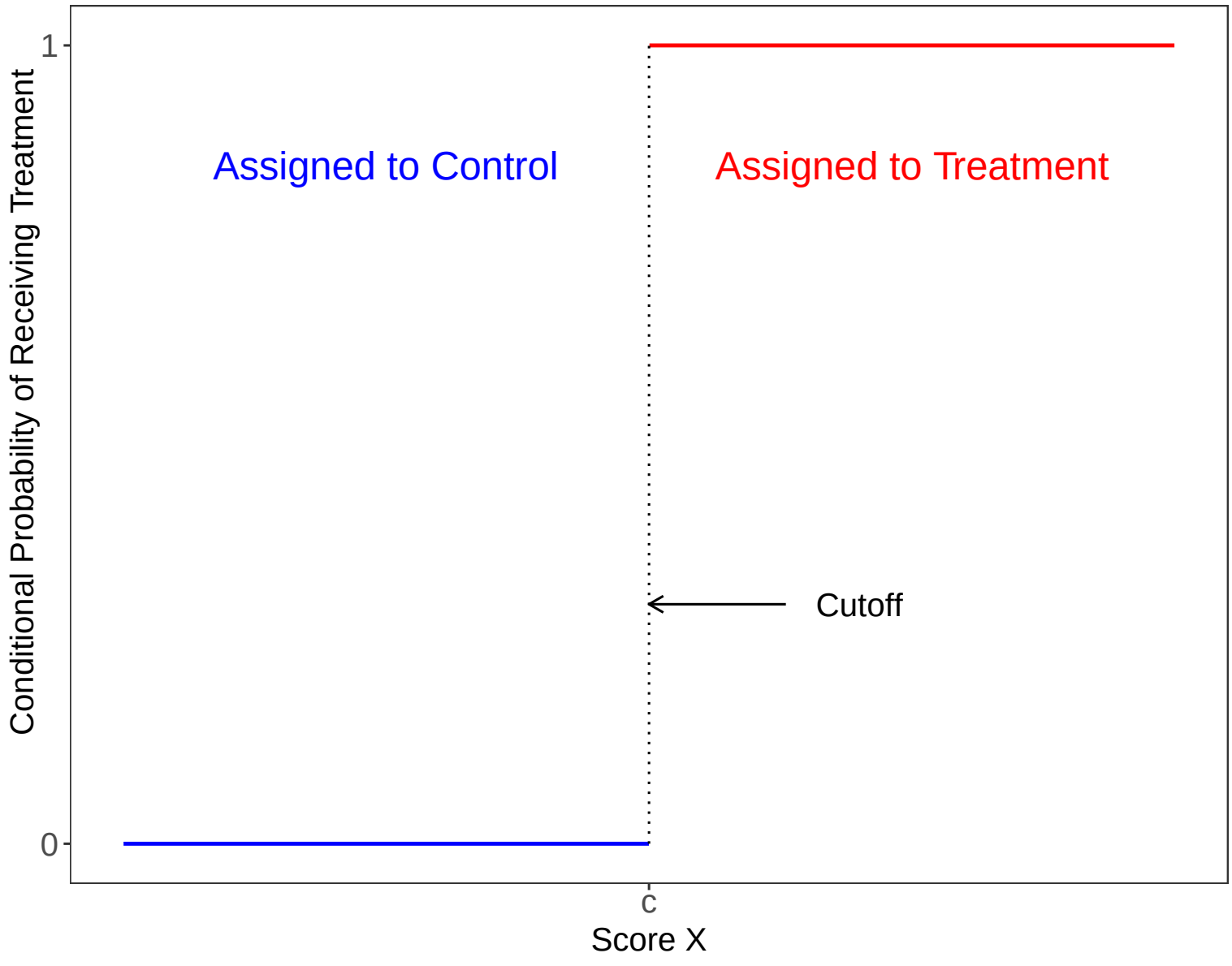


(b) Local Randomization Analysis

Part 5:

Fuzzy RD Designs

Treatment Assignment in (Sharp) RD Design



Sharp Regression Discontinuity Design

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

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

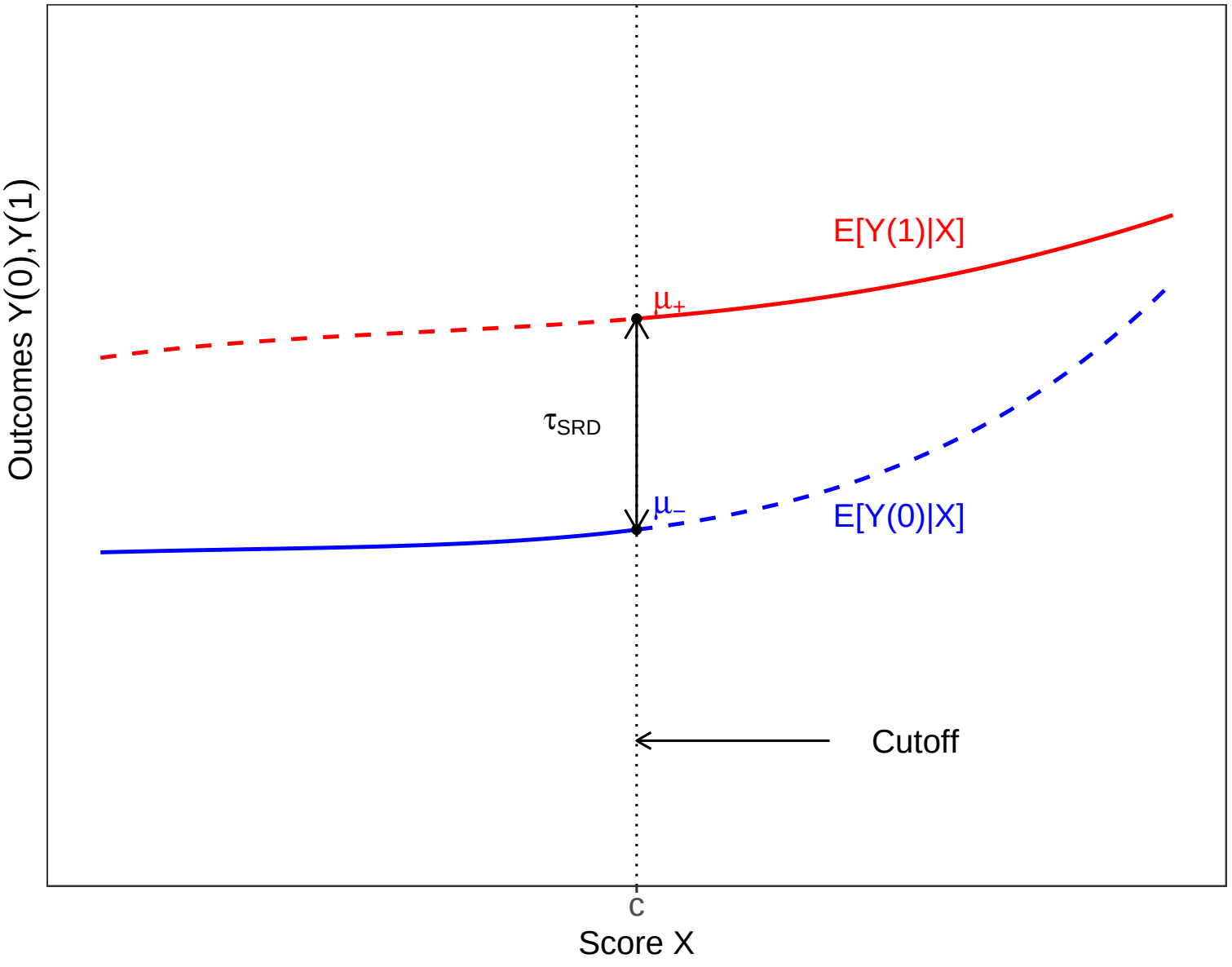
  $Y_i(0)$: outcome that would be observed if i received control

- The *observed* outcome is

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

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

RD Treatment Effect in Sharp RD Design



Fuzzy RD Design

- Imperfect compliance:
  - Probability of treatment changes at c, but not necessarily from 0 to 1
  - Some units with score above c may decide not to take up treatment
  - Example: voting eligibility at 18

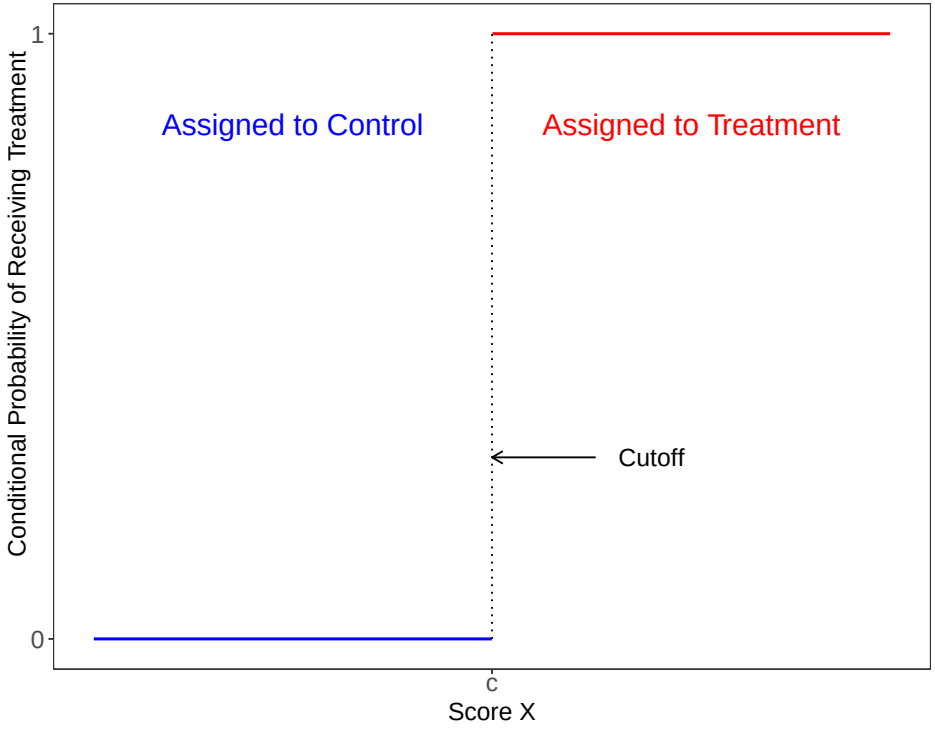
- T_i is treatment assigned, D_i is treatment taken

- Now for some units $T_i \neq D_i$

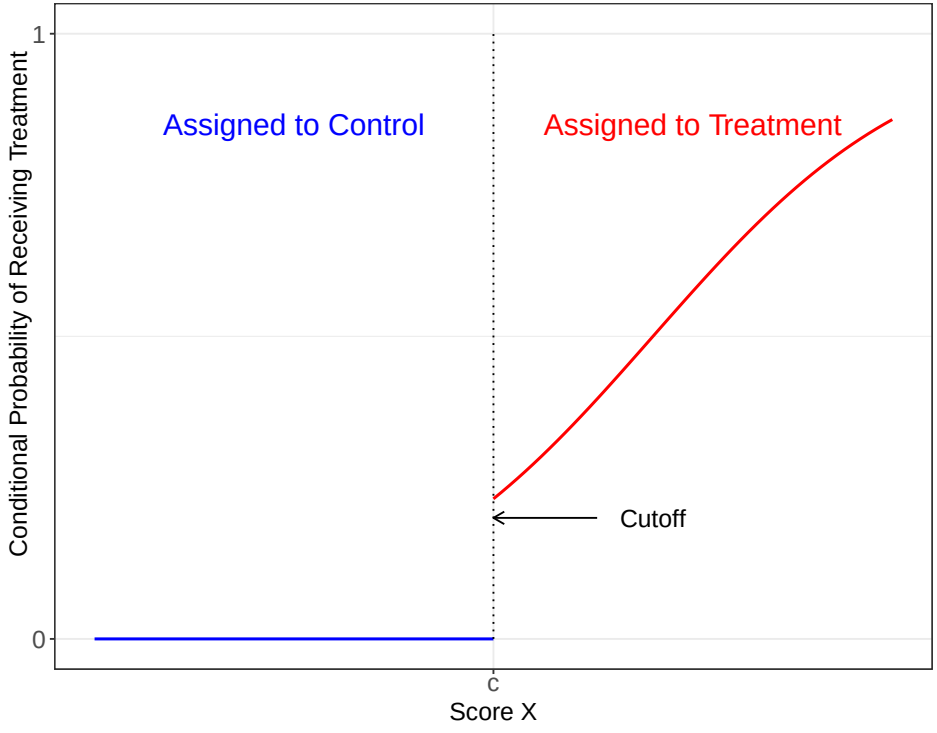
- Treatment taken has two potential values, $D_i(1)$ and $D_i(0)$, and observed treatment taken is $D_i = T_i \cdot D_i(1) + (1 - T_i) \cdot D_i(0)$

- Four potential outcomes instead of two:
  $Y_i(1, D_i(1)) = D_i(1)Y_i(1,1) + (1 - D_i(1))Y_i(1,0)$
  $Y_i(0, D_i(0)) = D_i(0)Y_i(0,1) + (1 - D_i(0))Y_i(0,0).$

Conditional Probability of Receiving Treatment
Sharp vs. Fuzzy RD Designs



(a) Sharp RD



(b) Fuzzy RD (One-Sided)

Fuzzy RD Design

- Interest in both the effect of being assigned to treatment (i.e., the effect of T) and the effect of actually receiving treatment (i.e., the effect of D)
- Since treatment assignment cannot be changed, compliance with the assignment is always perfect. Thus, analysis of the effect of T follows a Sharp RD design
- In contrast, the study of the effect of D requires modifications and additional assumptions

Fuzzy RD Design: Continuity-based parameters

- The Sharp RD estimator of the effect of T_i on Y_i consistently estimates the quantity

$$\begin{aligned}\tau_Y &:= \lim_{x \downarrow c} \mathbb{E}[Y_i | X_i = x] - \lim_{x \uparrow c} \mathbb{E}[Y_i | X_i = x] \\ &= \lim_{x \downarrow c} \mathbb{E}[Y_i(1, D_i(1)) | X_i = x] - \lim_{x \uparrow c} \mathbb{E}[Y_i(0, D_i(0)) | X_i = x]\end{aligned}$$

  where the equality follows from the more general definition of the observed outcome as $Y_i = T_i Y_i(1, D_i(1)) + (1 - T_i) Y_i(0, D_i(0))$, and thus requires no special assumptions.

Fuzzy RD Design: Intention-to-treat

- Assuming continuity of $\mathbb{E}[Y_i(1, D_i(1))|X_i = x]$ and $\mathbb{E}[Y_i(0, D_i(0))|X_i = x]$, seen as functions of x, at the cutoff c, we have

$$\tau_Y = \tau_{\texttt{ITT}} \quad , \quad \tau_{\texttt{ITT}} := \mathbb{E}[Y_i(1, D_i(1)) - Y_i(0, D_i(0))|X_i = c],$$

  and thus estimated jump in the average observed outcome at the cutoff recovers the average effect of T on Y at c.
- $\tau_{\texttt{ITT}}$ is usually called average “intention-to-treat" effect, and it captures effect (at the cutoff) of being assigned to treatment
- This parameter is different from Sharp RD parameter $\tau_{\texttt{SRD}}$ under perfect compliance,

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

Fuzzy RD Design: Intention-to-treat

- Perfect compliance is a particular case where
  - $\mathbb{P}[D_i(0) = 0|X_i = x] = 1$ for $x < c$ and $\mathbb{P}[D_i(1) = 1|X_i = x] = 1$ for $x \geq c$
  - $D_i = T_i = \mathbb{1}(X_i \geq c)$
  - $Y_i(1,1) := Y_i(1)$ and $Y_i(0,0) := Y_i(0)$

- Thus, when compliance is perfect, the RD ITT effect of the treatment assignment on the outcome is equivalent to the Sharp RD effect of the treatment received:

$$\tau_{\texttt{ITT}} = \mathbb{E}[Y_i(1) - Y_i(0)|X_i = c]$$

- But when some units are non-compliers, $\tau_{\texttt{ITT}}$ captures the effect of the treatment assignment, which will be in general different from the effect of actually receiving the treatment

Fuzzy RD Design: First Stage

- Fuzzy analysis includes study of how the RD assignment rule affects the probability of receiving the treatment.
- Treating D_i as the outcome, a Sharp RD strategy estimates

$$\tau_{\mathrm{D}} := \lim_{x \downarrow c} \mathbb{E}[D_i | X_i = x] - \lim_{x \uparrow c} \mathbb{E}[D_i | X_i = x]$$

- Since D_i is binary, τ_{D} captures the difference in the probability of receiving the treatment between units assigned to treatment vs. assigned to control, at the cutoff.

Fuzzy RD Design: First Stage

- Assuming continuity at c of $\mathbb{E}[D_i(1)|X_i = x]$ and $\mathbb{E}[D_i(0)|X_i = x]$, seen as functions of x, we have

$$\tau_{\mathtt{D}} = \tau_{\mathtt{FS}} \quad , \quad \tau_{\mathtt{FS}} := \mathbb{E}[D_i(1) - D_i(0)|X_i = c]$$

  and thus can interpret $\tau_{\mathtt{D}}$ as the causal effect of T_i on D_i.

- $\tau_{\mathtt{FS}}$ captures the effect of assigning the treatment on receiving the treatment for units with scores near or at the cutoff, usually called "first-stage" effect.

Fuzzy RD Design: Estimation of FS and ITT effects

- Since both $\tau_{\texttt{FS}}$ and $\tau_{\texttt{ITT}}$ are Sharp RD parameters, analysis follows standard continuity-based Sharp RD methods, using X_i as running variable, $T_i = \mathbb{1}(X_i \geq c)$ as treatment of interest, and D_i and Y_i as outcomes:

$$\hat{\tau}_{\texttt{ITT}} = \lim_{x \downarrow c} \widehat{\mathbb{E}}[Y_i | X_i = x] - \lim_{x \uparrow c} \widehat{\mathbb{E}}[Y_i | X_i = x]$$
$$\hat{\tau}_{\texttt{FS}} = \lim_{x \downarrow c} \widehat{\mathbb{E}}[D_i | X_i = x] - \lim_{x \uparrow c} \widehat{\mathbb{E}}[D_i | X_i = x],$$

  with bandwidth selection and inference methods as discussed before.

Fuzzy RD Design: Effect of Actual Treatment

- When interest is on the effect of the treatment received, it is common to focus on

$$\tau_{\texttt{FRD}} := \frac{\tau_{\texttt{Y}}}{\tau_{\texttt{D}}}$$

- We call $\tau_{\texttt{FRD}}$ the "fuzzy RD parameter."

- (Under the augmented continuity conditions for ITT effects, $\tau_{\texttt{FRD}} = \frac{\tau_{\texttt{ITT}}}{\tau_{\texttt{FS}}}$. This interpretation of the Fuzzy RD parameter as ratio of two ITT effects is analogous to result in IV literature. Below we do not assume that these conditions hold.)

Fuzzy RD Design: Effect of Actual Treatment

Explore conditions under which $\tau_{\texttt{FRD}}$ can be directly interpreted as the average treatment effect of the treatment for some subpopulations.

- **Non-zero first stage**: $\tau_{\texttt{FS}}$ must be nonzero—ideally, well-separated from zero: Moving above/below the cutoff must induce some units to actually take the treatment.
- **Exclusion Restriction**: the treatment assignment must affect the potential outcomes and potential treatments only via the treatment received, but not directly: $\mathbb{E}[Y_i(T_i, 0)|X_i = x]$ and $\mathbb{E}[Y_i(T_i, 1)|X_i = x]$ must be continuous in x at c.
- **Compliance Restriction**: many possibilities, including
  - Local independence: potential outcomes independent of potential treatments near the cutoff (Hahn, Todd, and vanderKlaauw, 2001).
  - Monotonicity: there are no units who receive the opposite treatment to the one they are assigned near the cutoff (i.e., no "defiers").

Fuzzy RD Design

Interpretation of $\tau_{\texttt{FRD}}$ will differ according to which assumptions we are willing to make. For example,

- Under **local independence**

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

- Under **monotonicity**

$$\tau_{\texttt{FRD}} = \mathbb{E}[Y_i(1,1) - Y_i(0,0) | X_i = c, \ \ i \ \texttt{is a complier}]$$

Important Issues for Implementation of Fuzzy RD analysis

- Falsification: density test and covariates effects should focus on intention-to-treat effects.
- Bandwidth Selection: two bandwidths if focus on ITT and FS effects, single bandwidth if focus on Fuzzy RD effect.
- Weak Assignment: Avoid analyzing Fuzzy RD effects when the RD assignment rule has weak effect on the adoption of the treatment.

Empirical Example

- Study by Londoño-Vélez, Rodríguez, and Sánchez (AEJ, 2020) on the effects Ser Pilo Paga (SPP), a governmental program in Colombia that funds full tuition to attend higher education institutions (HEIs).
- To be eligible, students must score in top 9 percent of scores in national high school exit exam ("SABER 11" score), and must come from a household with wealth index below a region-specific threshold ("SISBEN" wealth score).
- Focus on students who took the SABER 11 test in the fall of 2014.
- Transform this two-dimensional RD design into one-dimensional design: only students whose SABER 11 score is above cutoff.
- Score: difference between student's SISBEN wealth index and respective cutoff.
- Cutoff: normalized to zero.
- Treatment assignment (T): an indicator equal to one if score below zero. Treatment received (D): indicator equal to one if student received subsidy.

Part 7:

Falsification Analysis for RD designs

Falsification Methods

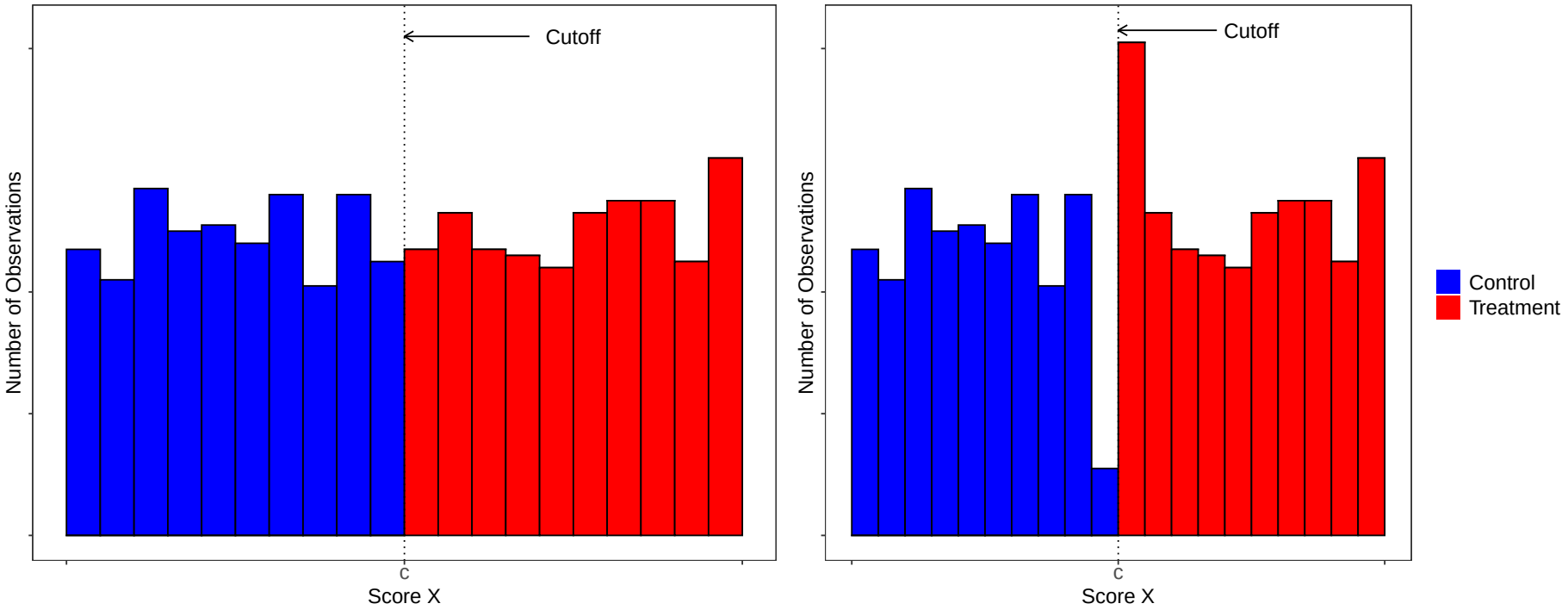
- RD rule of treatment assignment is not by itself enough to guarantee that continuity or local randomizations are met

- Qualitative information and quantitative falsification tests play crucial role
  - Qualitative information: were there mechanisms to appeal score? did people change their score?
  - Falsification: various statistical tests

Falsification Methods

- Density test of "sorting": is number of observations below the cutoff surprisingly different from number of observations above it?
- Treatment effect on
  - Predetermined covariates
  - Placebo outcomes
- Also: effect at different cutoffs, effect at different bandwidths, dougnout hole

Falsification Methods: Density Test



(a) No sorting

(b) Sorting

Falsification Methods: Tests on Predetermined Covariates and Placebo Outcomes

- Continuity-based falsification:
  - Test of continuity of density of the running variable
  - Local polynomial effects with optimal banwidth
  - Robust Inference
  - CRUCIAL: each covariate/placebo outcome must have its own optimal bandwidth
- Local randomization falsification:
  - Within chosen window, density test
  - Test that covariate and placebo outcome distributons are indistinguishable for treated and control
  - CRUCIAL: all tests are conducted within the same window for each covariate/placebo outcome