Regression Discontinuity Designs

Ritsu Kitagawa

Waseda University

Seminar on Economics (Owan, H.): TA Session

Spring 2022

1. Sharp Regression Discontinuity
   - Identification
   - Estimation
   - Example
   - Diagnostics

2. Fuzzy Regression Discontinuity
   - Identification and Estimation

Regression Discontinuity Design (RDD)

- RDD: A fairly old idea (Thistlethwaite and Campbell, 1960)
- Recently experienced a renaissance because of the formal theoretical foundation given by the causal inference framework
- Applicable when treatment is assigned according to a *rule* based on another variable (called the forcing or running variable)
- Often useful for analysis in a "rule-based" world (administrative programs, elections, etc.)
- High internal validity: One of the few observational designs that reproduced an experimental benchmark (Cook and Wong 2008)
- Limited external validity: Effect is only identified for a small subpopulation

Sharp RDD: Basic Setup

- $D_i \in \{0, 1\}$: Treatment
- X_i: Forcing variable that perfectly determines the value of D_i with cutpoint c

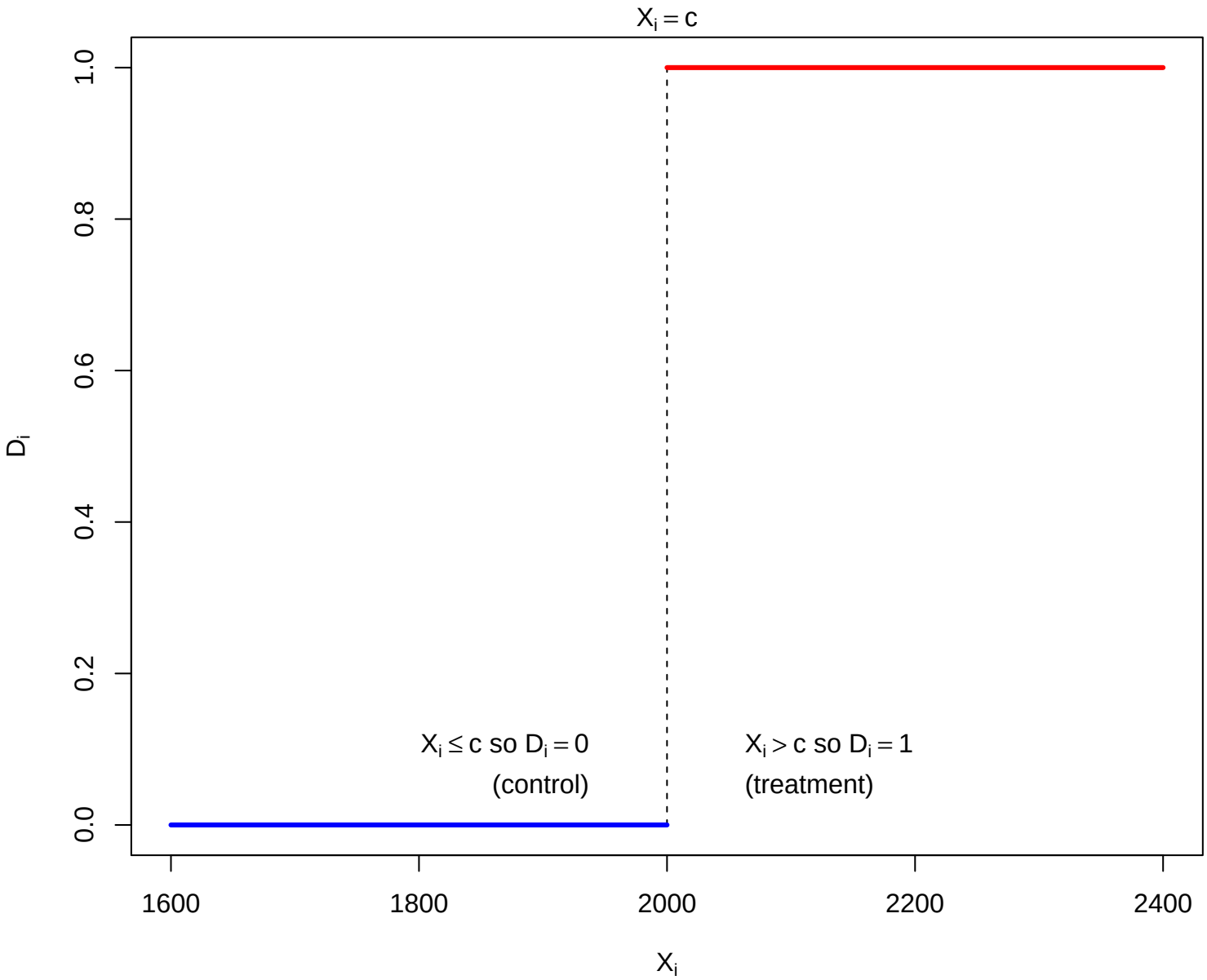
$$D_i = 1\{X_i > c\} \quad \text{or equivalently} \quad D_i = \begin{cases} 1 & \text{if } X_i > c \\ 0 & \text{if } X_i \leq c \end{cases}$$

- X_i may be correlated with $Y_i(0)$ and $Y_i(1)$, either directly or via other unobserved confounders
- Simply adjusting for X_i does not work because of lack of common support
- Basic intuition: Right at the cutpoint $X_i = c$, assignment to D_i may be as-if random

A Hypothetical Example: Effect of Scholarship

- Thistlethwaite and Campbell (1960) study the effects of college scholarships on later students' achievements
- Scholarships are given on the basis of whether or not a student's test score exceeds some threshold c
  - Treatment D_i is scholarship
  - Forcing variable X_i is SAT score with cutoff c
  - Outcome Y_i is subsequent earnings
  - $Y_i(0)$: potential earnings without the scholarship
  - $Y_i(1)$: potential earnings with the scholarship
- $Y_i(1)$ and $Y_i(0)$ are correlated with X_i: on average, students with higher SAT scores obtain higher earnings

Probability of Treatment in Sharp RDD



Identification of the Threshold Causal Effect

Key assumption: Continuity of average potential outcomes

$$\mathbb{E}[Y_i(d) \mid X_i = x] \text{ is continuous in } x \text{ around } X_i = c \text{ for } d = 0, 1$$

Causal estimand: Local ATE at the threshold

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

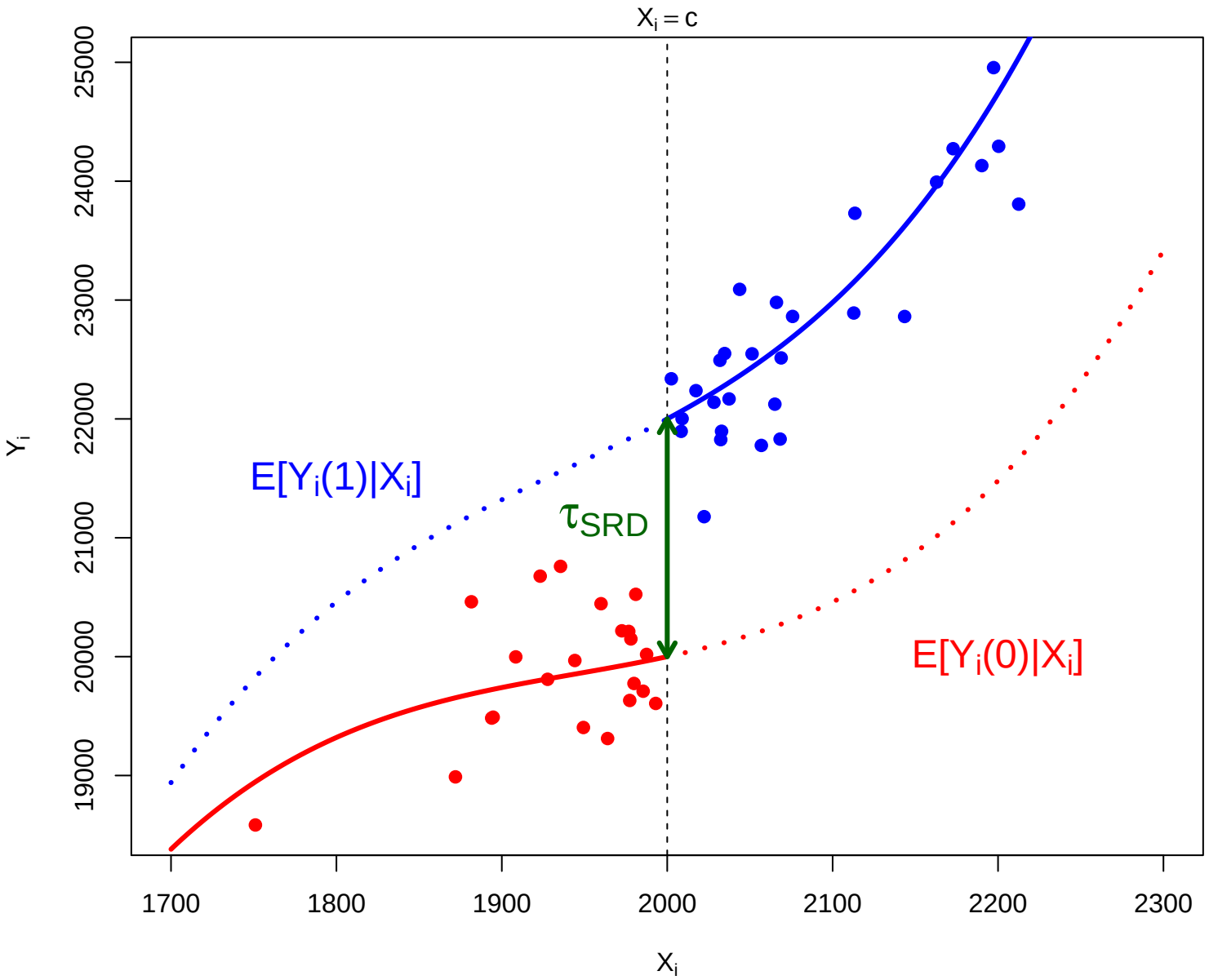
Identification result: If the continuity assumption holds, τ_{SRD} is nonparametrically identified as

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

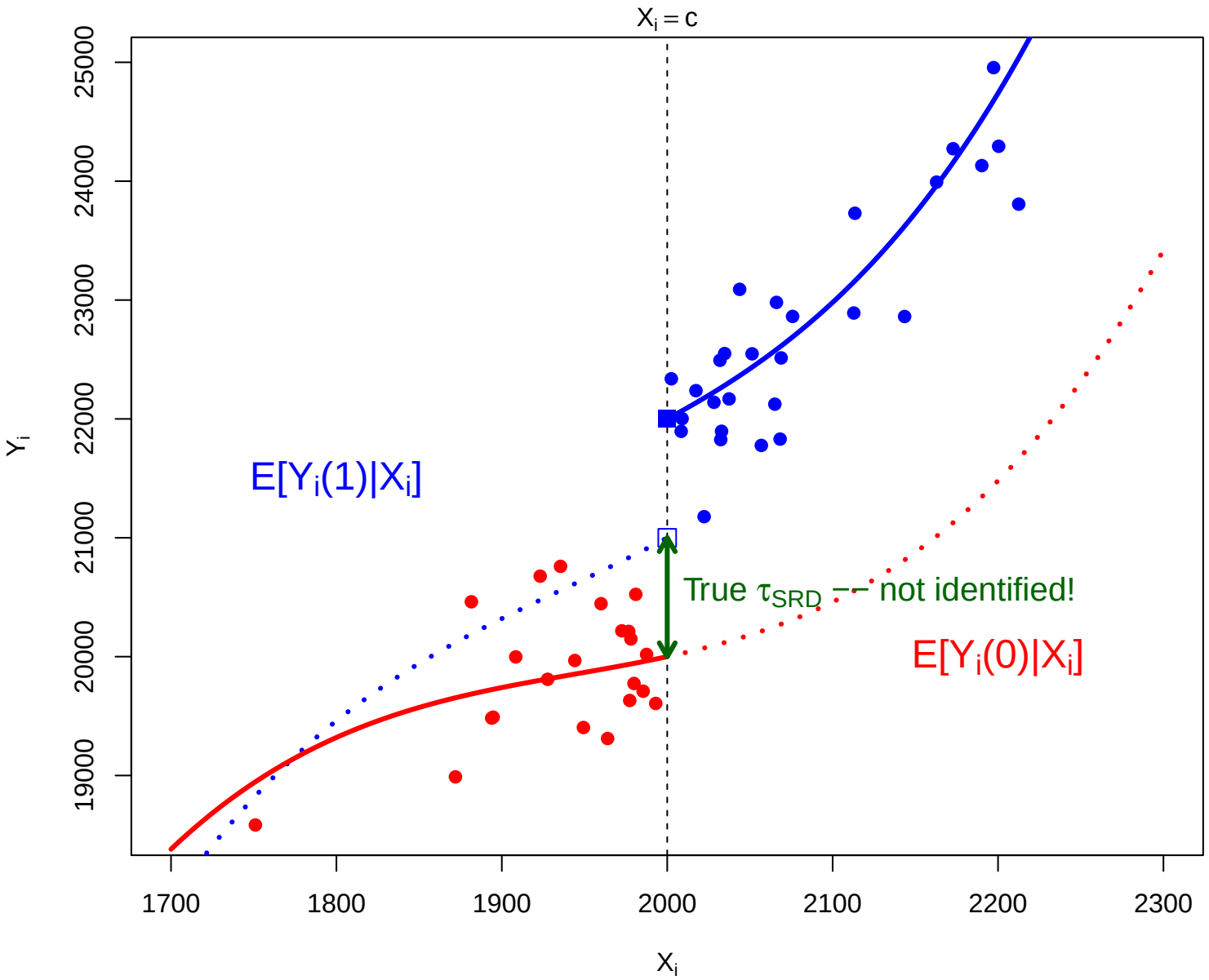
A "proof":

- D_i is wholly determined by X_i, so contional ignorability is trivially satisfied given X_i: $Y_i(1), Y_i(0) \perp\!\!\!\perp D_i \mid X_i$
- However, *there is no common support*, so conditioning on X_i in a usual way won't work.
- The continuity assumption allows us to do a tiny bit of extrapolation and compensate for the lack of common support at the threshold.

Graphical Illustration: Continuous $\mathbb{E}[Y_i(d) \mid X_i]$



Graphical Illustration: Discontinuous $\mathbb{E}[Y_i(d) \mid X_i]$



1. Sharp Regression Discontinuity
   - Identification
   - Estimation
   - Example
   - Diagnostics
2. Fuzzy Regression Discontinuity
   - Identification and Estimation

Estimation of the LATE at the Threshold

1. Trim the sample to a reasonable window around the threshold c (discontinuity sample)
   - $c - h \leq X_i \leq c + h$, were $h > 0$ determines the width of the window
2. Recode forcing variable to deviations from threshold: $\tilde{X}_i = X_i - c$
   - $\tilde{X}_i = 0$ if $X_i = c$
   - $\tilde{X}_i > 0$ if $X_i > c$ and thus $D_i = 1$
   - $\tilde{X}_i < 0$ if $X_i < c$ and thus $D_i = 0$
3. Decide on a model for $\mathbb{E}[Y_i|\tilde{X}_i]$:
   - linear, common slope for $\mathbb{E}[Y_i \mid \tilde{X}_i < 0]$ and $\mathbb{E}[Y_i|\tilde{X}_i > 0]$
   - linear, different slopes
   - non-linear
   - each model corresponds to a particular set of assumptions about the potential outcomes
   - always start with visual inspection (e.g. scatter plot with lowess) to check which model is plausible

Estimation with a Linear Model with a Common Slope

- Assumptions:
  1. $\mathbb{E}[Y_i(0)|X_i = x]$ is linear in x
  2. Treatment effect, τ, does not depend on X_i

  i.e.,

$$\mathbb{E}[Y_i(0)|X_i] = \alpha + \beta X_i \quad \text{and} \quad \mathbb{E}[Y_i(1) - Y_i(0)|X_i] = \tau$$

  which implies

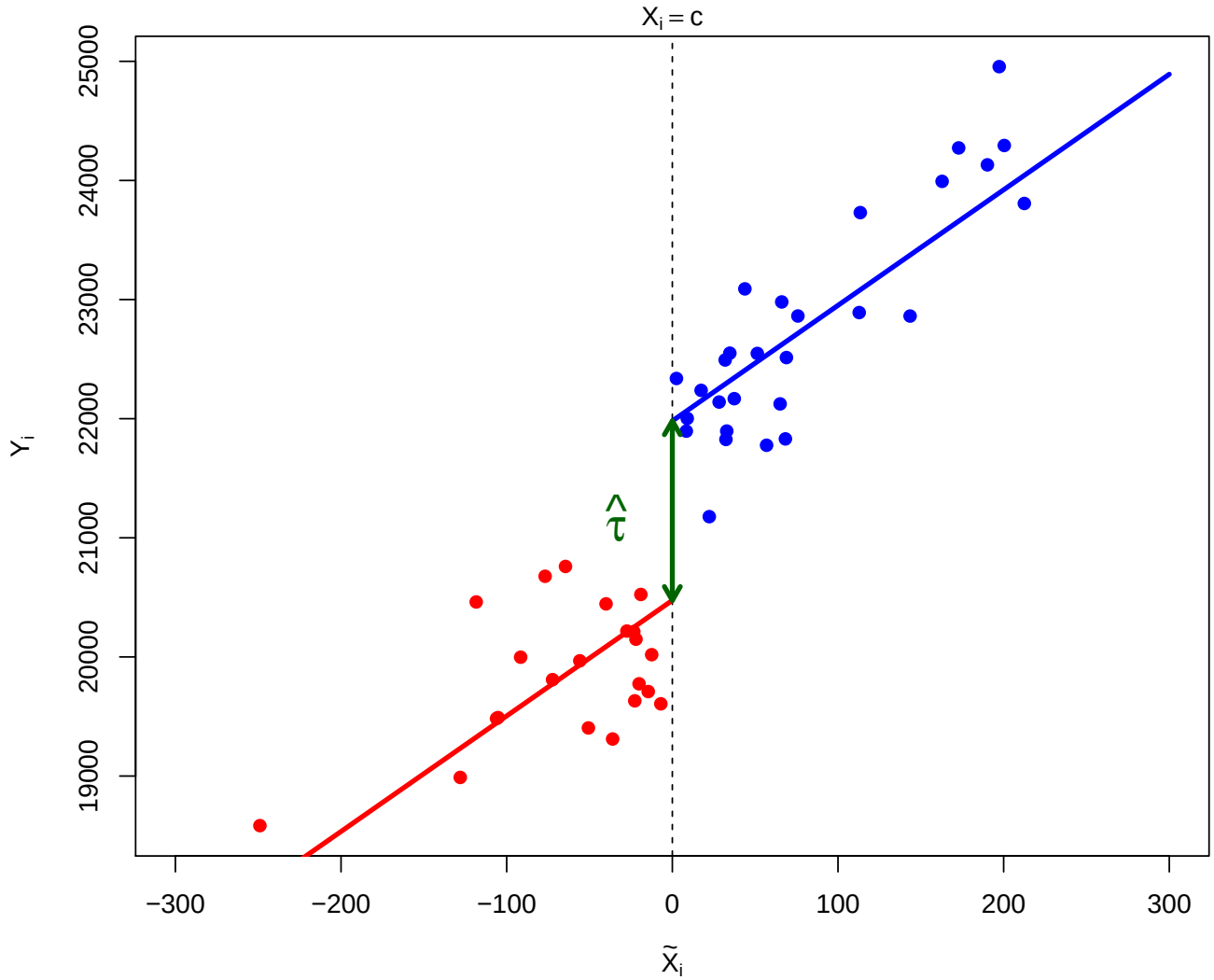
$$\mathbb{E}[Y_i(1)|X_i] = \tau + \mathbb{E}[Y_i(0)|X_i] = \tau + \alpha + \beta X_i$$

- Therefore, the model for the observed outcome should be:

$$\begin{aligned}
\mathbb{E}[Y_i|X_i, D_i] &= D_i \cdot \mathbb{E}[Y_i(1)|X_i] + (1 - D_i) \cdot \mathbb{E}[Y_i(0)|X_i] \\
&= \alpha + \tau D_i + \beta X_i \\
&= \tilde{\alpha} + \tau D_i + \beta \tilde{X}_i \quad (\text{where } \tilde{\alpha} = \alpha + \beta c)
\end{aligned}$$

- So we just regress the observed outcome (Y_i) on D_i and $\tilde{X}_i$

Estimation with a Linear Model with a Common Slope



Estimation with a Linear Model with a Different Slope

- Assumptions:
  1. $\mathbb{E}[Y_i(0)|X_i = x]$ and $\mathbb{E}[Y_i(1)|X_i = x]$ are both linear in x
  2. But we now allow treatment effect to vary with X_i

  i.e.,

$$\mathbb{E}[Y_i(0)|X_i] = \alpha_0 + \beta_0 X_i \quad \text{and} \quad \mathbb{E}[Y_i(1)|X_i] = \alpha_1 + \beta_1 X_i$$

  such that

$$\mathbb{E}[Y_i(1) - Y_i(0)|X_i] = (\alpha_1 - \alpha_0) + (\beta_1 - \beta_0)X_i$$

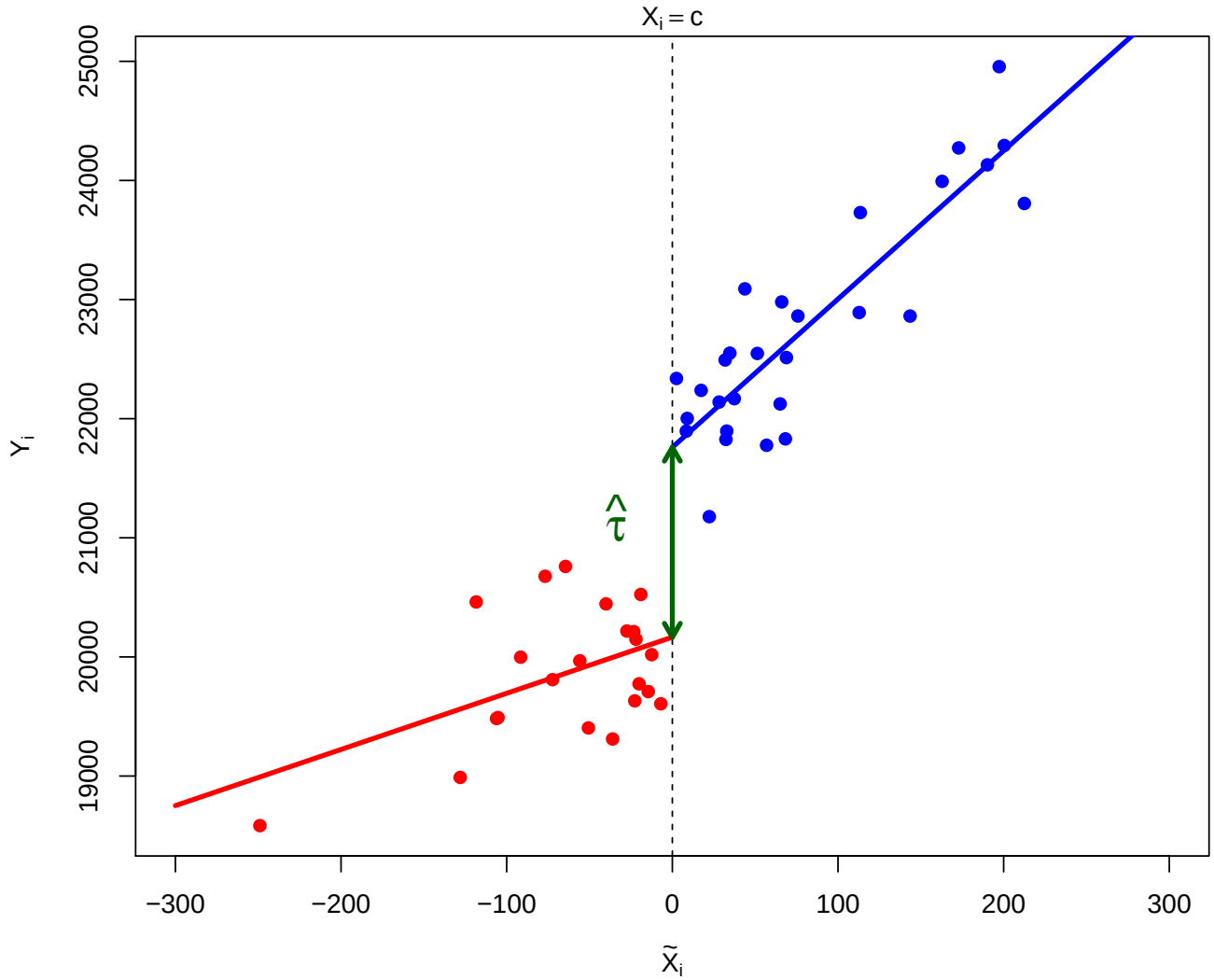
- The observed outcome model is therefore:

$$\begin{aligned}
\mathbb{E}[Y_i|X_i, D_i] &= D_i \cdot \mathbb{E}[Y_i(1)|X_i] + (1 - D_i) \cdot \mathbb{E}[Y_i(0)|X_i] \\
&= \alpha_0 + \beta_0 X_i + (\alpha_1 - \alpha_0)D_i + (\beta_1 - \beta_0)D_i X_i \\
&= (\alpha_0 + \beta_0 c) + \beta_0 \tilde{X}_i \\
&\quad + \{(\alpha_1 - \alpha_0) + (\beta_1 - \beta_0)c\} D_i + (\beta_1 - \beta_0)D_i \tilde{X}_i \\
&\equiv \tilde{\alpha} + \beta_0 \tilde{X}_i + \tau D_i + \tilde{\beta} D_i \tilde{X}_i
\end{aligned}$$

  Note that $\tau = \mathbb{E}[Y_i(1) - Y_i(0)|X_i = c]$, LATE at the threshold

- So, regress Y_i on $\tilde{X}_i$, D_i and the interaction $D_i \tilde{X}_i$

Estimation with a Linear Model with a Different Slope



Estimation with a Nonlinear Model

- Assumptions:
  1. $\mathbb{E}[Y_i(0)|X_i = x]$ and $\mathbb{E}[Y_i(1)|X_i = x]$ are now allowed to be non-linear in X_i, but must be correctly specified
  2. Treatment effect is allowed to vary across X_i
- Include quadratic, cubic, etc. terms in $\tilde{X}_i$ and their interactions with D_i in the equation
- The specification with quadratic terms:

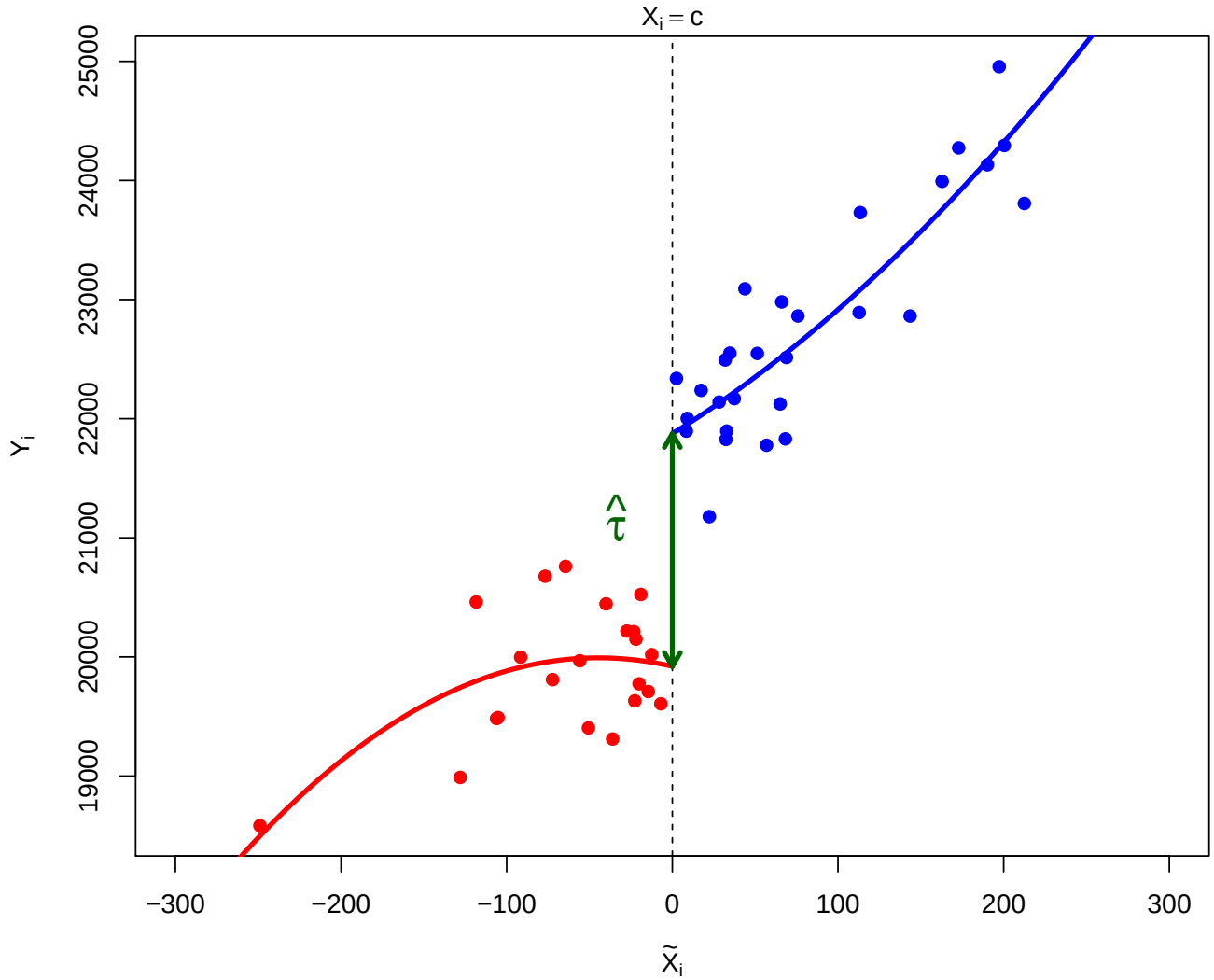
$$\mathbb{E}[Y_i|X_i, D_i] = \gamma_0 + \gamma_1\tilde{X}_i + \gamma_2\tilde{X}_i^2 + \tau D_i + \alpha_1\tilde{X}_i D_i + \alpha_2\tilde{X}_i^2 D_i$$

  The specification with cubic terms is

$$\begin{aligned}\mathbb{E}[Y_i|X_i, D_i] &= \gamma_0 + \gamma_1\tilde{X}_i + \gamma_2\tilde{X}_i^2 + \gamma_3\tilde{X}_i^3 \\ &\quad + \alpha_0 D_i + \tau\tilde{X}_i D_i + \alpha_2\tilde{X}_i^2 D_i + \alpha_3\tilde{X}_i^3 D_i\end{aligned}$$

- In both cases, the coefficient on D_i corresponds to the LATE at the threshold: $\tau = \mathbb{E}[Y_i(1) - Y_i(0)|X_i = c]$

Estimation with a Nonlinear Model



Model Selection and Choice of Bandwidth in RDD

- How should we pick the "right" model and bandwidth?
- No ex-ante correct answer usually, but several data-driven procedures are available

Model choice:

- A tradeoff between bias and variance
- Standard practice: Use different specs and show robustness
- Local linear regression with a kernel smoother is a popular choice

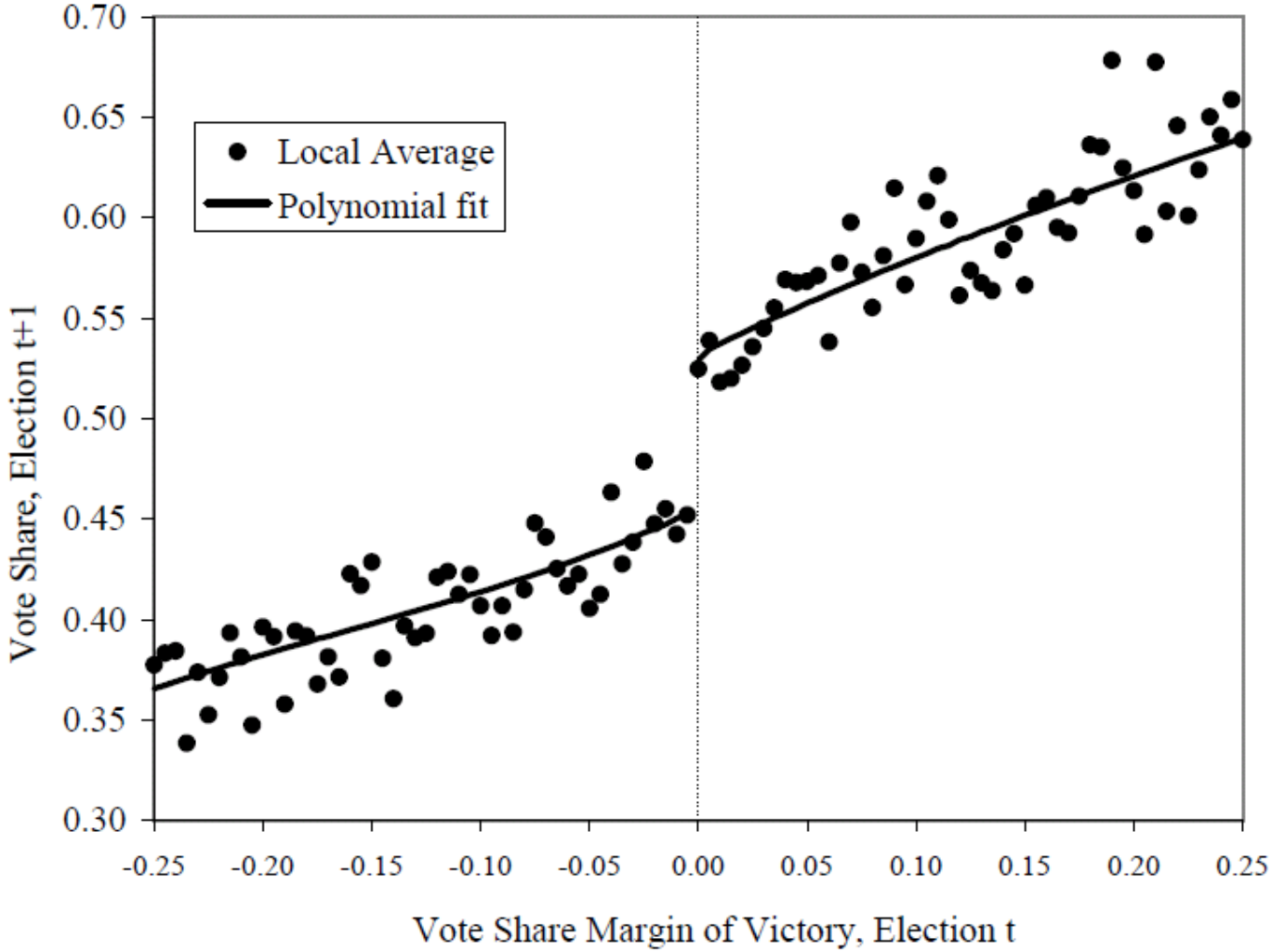
Bandwidth selection:

- Imbens-Kalyanaraman (IK) algorithm: Pick h that minimizes (a first-order approximation of) the MSE in $\hat{\tau}_{SRD}$
- Cross-validation: See Imbens and Lemieux (2008) for details

1. Sharp Regression Discontinuity
   - Identification
   - Estimation
   - Example
   - Diagnostics
2. Fuzzy Regression Discontinuity
   - Identification and Estimation

Example: Party Incumbency Advantage

Figure IVa: Democrat Party's Vote Share in Election t+1, by Margin of Victory in Election t: local averages and parametric fit



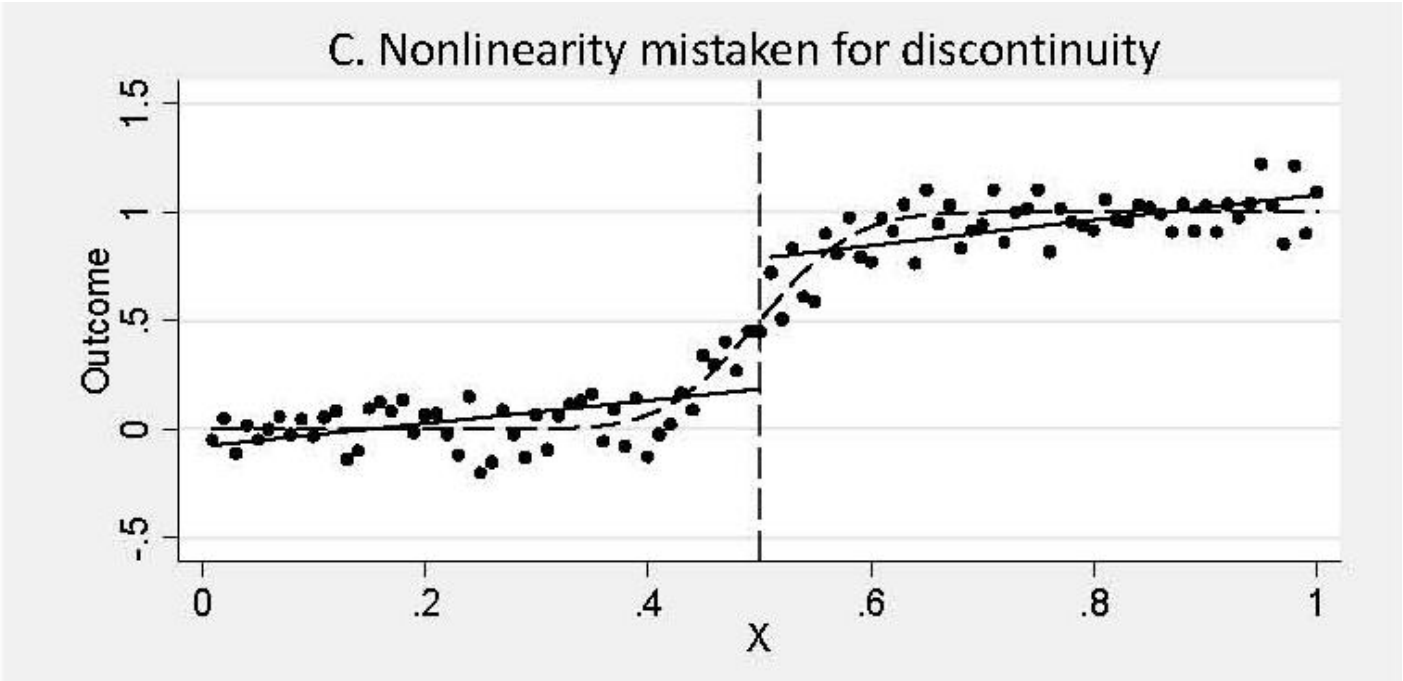
1. Sharp Regression Discontinuity
   - Identification
   - Estimation
   - Example
   - Diagnostics
2. Fuzzy Regression Discontinuity
   - Identification and Estimation

Falsification Checks

Diagnoze the robustness of your results via falsification checks:

1. Sensitivity: Are results sensitive to alternative specifications?
2. Balance checks: Does any covariate Z_i jump at the threshold?
3. Check if jumps occur at placebo thresholds c^*?
4. Sorting: Do units sort around the threshold?

Sensitivity to Specification



- RDD requires specification of the functional form and bandwidth
- Misspecification of either can lead to a spurious jump
- Take care not to confuse a nonlinear relation with a discontinuity!
- More flexibility (e.g. polynomials) reduces bias but increases variance
- Check sensitivity to size of bandwidth h

Balance Checks: Covariates as Placebo Outcomes

- Test for comparability of agents around the cutoff:
  - Visual tests: Plot $\mathbb{E}[Z_i|X_i, D_i]$ and look for jumps
  - Relation between covariates and treatment should be smooth around threshold
  - Use Z_i as a placebo outcome and see if there is inbalance:

$$\mathbb{E}[Z_i|X_i, D_i] = \beta_0 + \beta_1\tilde{X}_i + \tau_z D_i + \beta_3\tilde{X}_i D_i$$

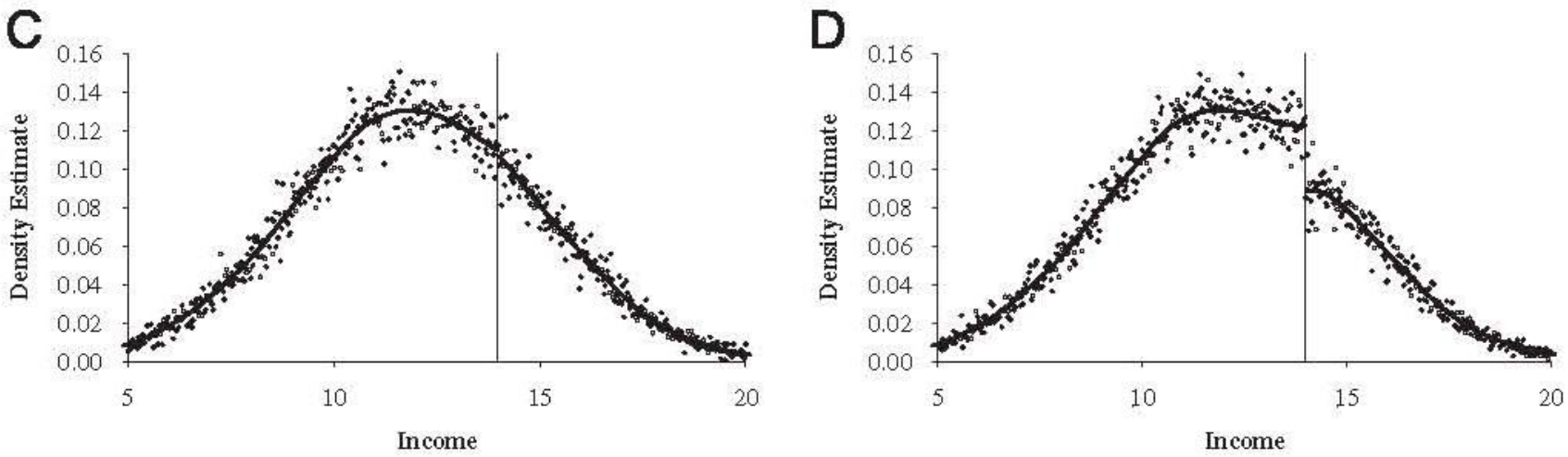
  - $\tau_z = 0$ if Z_i is balanced at the threshold
- Discontinuity in Z_i indicates possible evidence of discontinuous $\mathbb{E}[Y_i(d) \mid X_i = x]$, violating the key assumption
- Inbalance can be addressed by incorporating Z_i in the analysis:
  - Use Z_i as an additional covariate in the model
  - Alternatively, regress Y_i on Z_i and use the residuals in the model, instead of Y_i itself
- Balance checks address only observables, not unobservables

Sorting Around the Threshold

- Agents' behavior can invalidate the continuity assumption:
  - Agents may exercise control over their values of X_i to fall on the beneficial side of the threshold
  - Administrators may strategically choose what X_i to use or which threshold to use
  - Such sorting of agents invalidate the continuity assumption
- When this occurs, distribution of X_i will discontinuously change at the threshold
- Diagnostics:
  1. Visual inspection of Histograms (make sure no bin overlaps with the threshold!)
  2. Formal tests (e.g. McCrary 2008)
- A related problem: Other treatments assigned by the exact same X_i and c (e.g. geographic boundary)
  $\longrightarrow$ Effect needs to be interpreted as a composite treatment effect

Example: Job Training Program

- Beneficial job training program offered to agents with income $< c$
- Concern: People may withhold labor to lower their income just below the cutoff to gain access to the program



1. Sharp Regression Discontinuity
   - Identification
   - Estimation
   - Example
   - Diagnostics

2. Fuzzy Regression Discontinuity
   - Identification and Estimation

1. Sharp Regression Discontinuity
   - Identification
   - Estimation
   - Example
   - Diagnostics

2. Fuzzy Regression Discontinuity
   - Identification and Estimation

Generalization to Imperfect Thresholds

- Threshold may not perfectly determine treatment exposure, but it may still create a discontinuity in the *probability* of treatment
- Hypothetical example: SAT and scholarship
  - Scholarship may not be wholly determined by SAT scores X_i, but the rule specifies a threshold $X_i = c$ to become eligible
  - Students above c get a chance of receiving the scholarship, but may not actually get it
  - Students below c is excluded from consideration, perhaps except ones who are otherwise remarkable
- That is, an encouragement to receive treatment is discontinuously determined by X_i at the threshold c
- Fuzzy RDD can therefore be thought of an instrumental variable version of an RDD

Fuzzy RDD: Setup and Identification

- $Z_i \in \{0, 1\}$: Encouragement
- X_i: Forcing variable that perfectly determines the value of Z_i with cutpoint c

$$Z_i = 1\{X_i > c\} \quad \text{or equivalently} \quad Z_i = \begin{cases} 1 & \text{if } X_i > c \\ 0 & \text{if } X_i \leq c \end{cases}$$

Identification assumptions:

- Both $\mathbb{E}[D_i(z) \mid X_i = x]$ (potential treatment) and $\mathbb{E}[Y_i(z) \mid X_i = x]$ (potential outcome) are continuous in x around $X_i = c$ for $z = 0, 1$
- IV assumptions: Monotonicity, exclusion restriction, relevance of Z_i

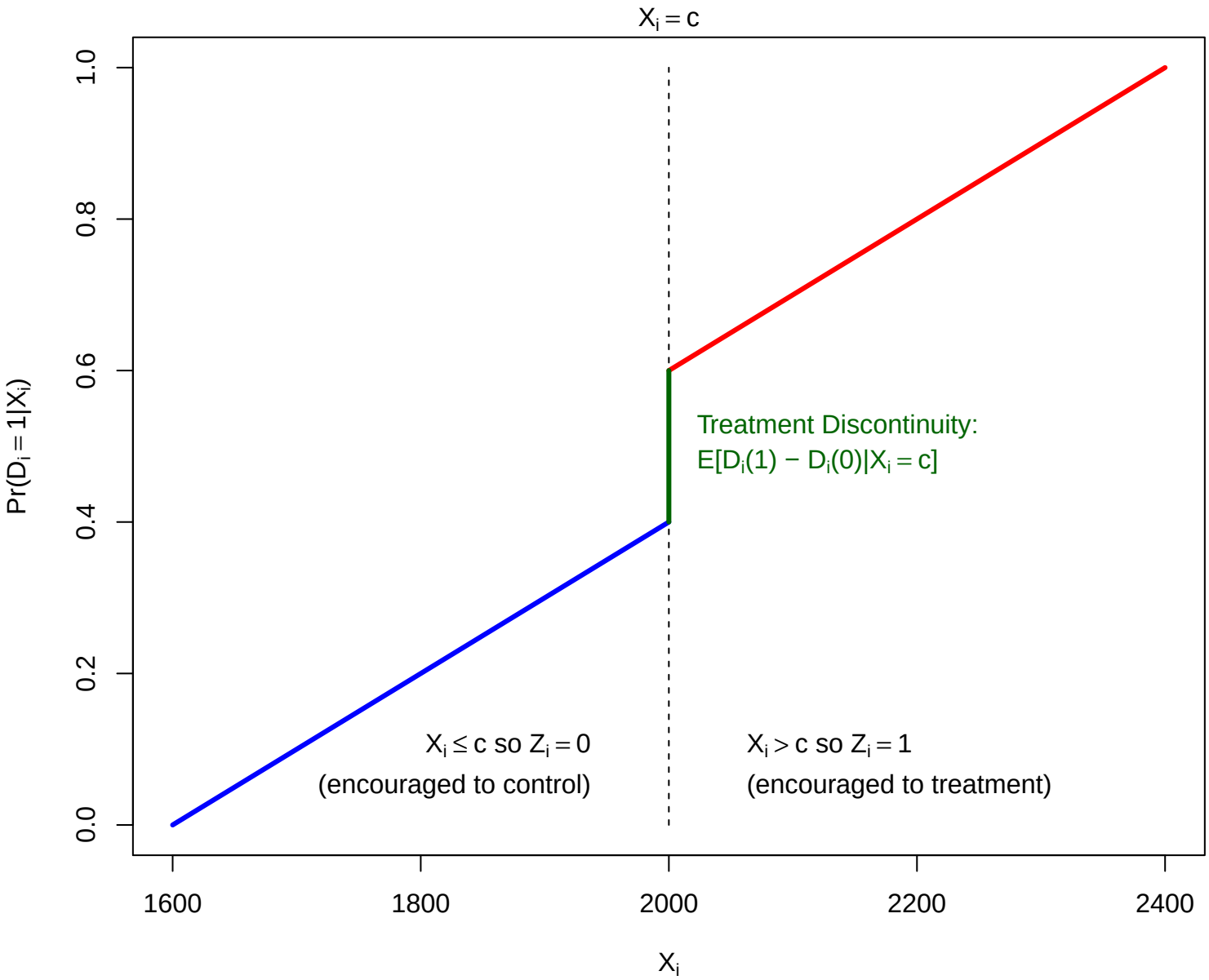
Causal estimand: Local ATE for compliers at the threshold

$$\tau_{FRD} = \mathbb{E}[Y_i(1) - Y_i(0) \mid \text{unit } i \text{ is a complier and } X_i = c]$$

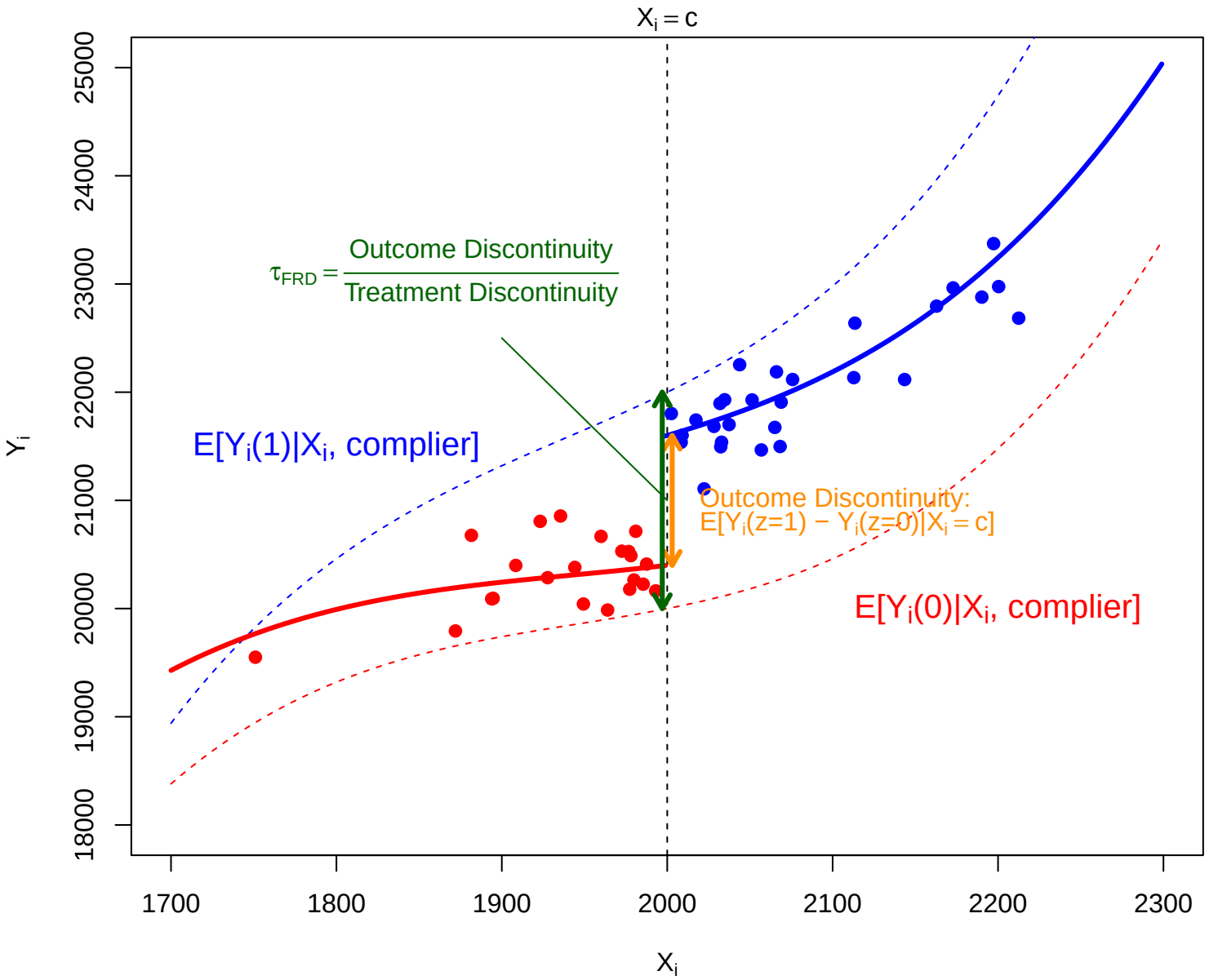
Identification result: Under the above assumptions, τ_{FRD} is identified as

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

Probability of Treatment in Fuzzy RDD



Fuzzy RDD: Discontinuity in $\mathbb{E}[Y|X]$



Fuzzy RDD: Estimation

1. Trim the sample to a reasonable window above and below the threshold c (discontinuity sample)
2. Code the encouragement indicator: $Z_i = 1\{X_i > c\}$
3. Recode the forcing variable to deviation from c: $\tilde{X}_i = X_i - c$
4. Estimate the outcome model using two-stage least squares:

$$Y_i = \beta_0 + \beta_1 \tilde{X}_i + \beta_2 Z_i \tilde{X}_i + \tau D_i + \varepsilon_i,$$

   where D_i is instrumented by Z_i
   - More flexible specifications can be used (e.g. polynomials of $\tilde{X}_i$)
5. Then $\hat{\tau}_{2SLS}$ consistently (but not unbiasedly) estimates τ_{FRD}

- In addition, it is also helpful to separately plot (and estimate) the outcome discontinuity and treatment discontinuity for interpretation
- Usual diagnostics can be applied to check plausibility of assumptions