

Boundary Discontinuity Designs: Theory and Practice

Matias D. Cattaneo
Princeton University

Rocio Titiunik
Princeton University

Ruiqi (Rae) Yu
Princeton University

July 2025

Talk based on:

- “Identification, Estimation, and Inference for Boundary Average Treatment Effects”
- “Estimation and Inference in Boundary Discontinuity Designs: Distance-Based Methods”
- “Estimation and Inference in Boundary Discontinuity Designs: Location-Based Methods”
- “rd2d: Causal Inference in Boundary Discontinuity Designs”

Outline

1. Introduction
2. Boundary Average Treatment Effects
3. Distance-Based Methods
4. Location-Based Methods
5. Empirical Application
6. Conclusion

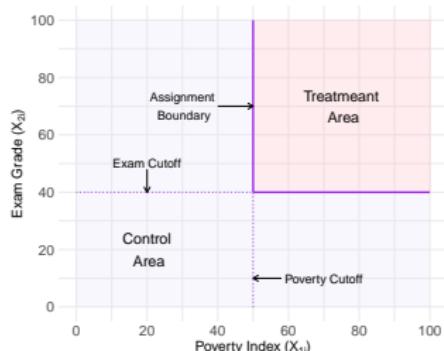
Introduction

Boundary Discontinuity Designs are used in causal inference and policy evaluation.

- ▶ Multi-dimensional Regression Discontinuity (RD) designs.
 - ▶ Multi-score RD designs / Geographic RD designs.
- ▶ Three modern approaches for analysis in practice.
 - ▶ Local regression using pooled data near the boundary.
 - ▶ Local regression using univariate distance to boundary point.
 - ▶ Local regression using bivariate location relative to boundary point.
- ▶ Today: foundational, thorough study of Boundary Discontinuity Designs.
 - ▶ *Methodology*: guidance on valid and invalid current practices, and more.
 - ▶ *Theory*: novel strong approximation approach for uniform inference, and more.
 - ▶ *Practice*: new R software (`rd2d` package).

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

Motivation: Basic Setup



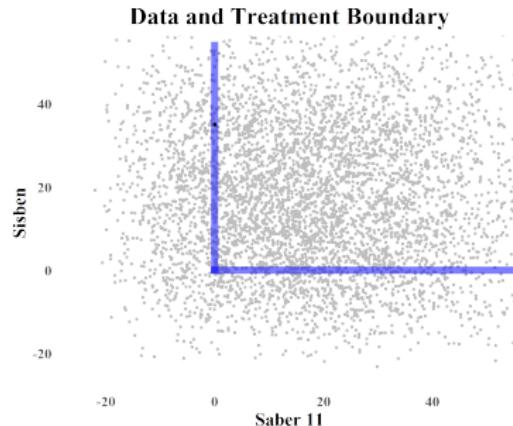
(a) Two-Score RD Design.



(b) Geographic RD Design.

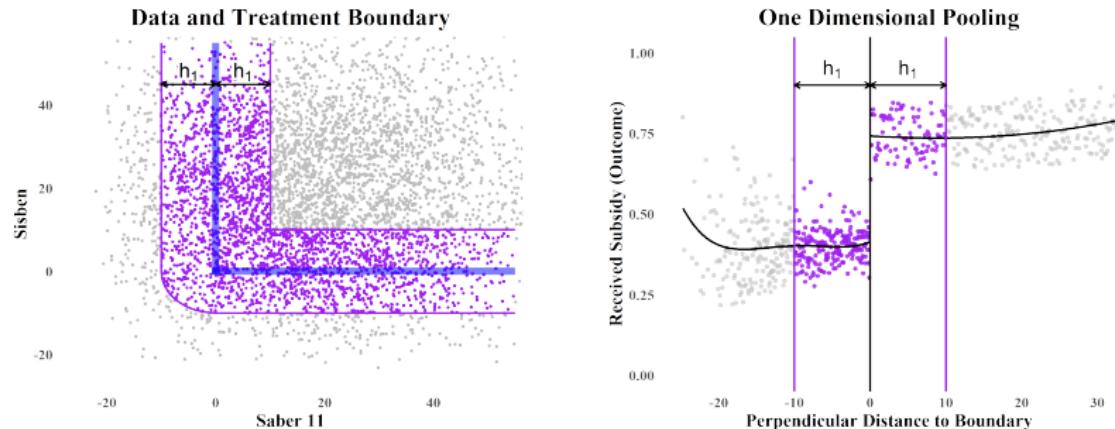
- ▶ \mathcal{B} = assignment boundary; \mathcal{A}_0 = control assignment; \mathcal{A}_1 = treatment assignment.
- ▶ \mathbf{X}_i = bivariate score; $Y_i = (1 - T_i) \cdot Y_i(0) + T_i \cdot Y_i(1)$; $T_i = \mathbf{1}(\mathbf{X}_i \in \mathcal{A}_1)$.
- ▶ $D_i(\mathbf{b}) = (2 \cdot T_i - 1) d(\mathbf{X}_i, \mathbf{b})$ distance to point $\mathbf{b} \in \mathcal{B}$; e.g., $d(\mathbf{x}, \mathbf{b}) = \|\mathbf{x} - \mathbf{b}\|$.
- ▶ $D_i = (2 \cdot T_i - 1) d(\mathbf{X}_i, \mathcal{B})$ distance to boundary \mathcal{B} ; $d(\mathbf{x}, \mathcal{B}) = \inf_{\mathbf{b} \in \mathcal{B}} d(\mathbf{x}, \mathbf{b})$.

Motivation: *Ser Pilo Paga* (SPP) Colombian policy program



- ▶ High-school graduates $i = 1, 2, \dots, n$ offered cash transfer to attend college ($T_i = 1$).
- ▶ $\mathbf{X}_i = (\text{SABER11}_i, \text{SISBEN}_i)^\top$; SABER11_i = exam score; SISBEN_i = wealth index.
- ▶ $\mathcal{B} = \{\text{SABER11} \geq 0 \text{ and } \text{SISBEN} = 0\} \cup \{\text{SABER11} = 0 \text{ and } \text{SISBEN} \geq 0\}$.
- ▶ $Y_i = 1$ if first year of college completed, $= 0$ otherwise.

Motivation: Boundary Average Treatment Effect



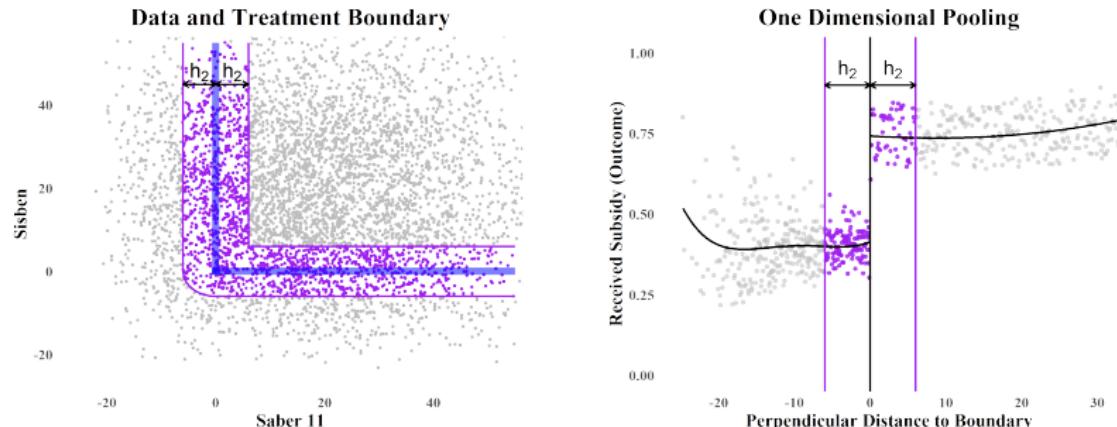
- Basic approach:

$$(\hat{\zeta}, \hat{\tau}) = \arg \min_{\zeta, \tau} \sum_{i=1}^n (Y_i - \zeta - T_i \tau)^2 \mathbb{1}(|D_i| \leq h),$$

$$T_i = \mathbb{1}(\mathbf{X}_i \in \mathcal{A}_1) = \mathbb{1}(D_i \geq 0)$$

$$D_i = (2 \cdot T_i - 1)d(\mathbf{X}_i, \mathcal{B}), \quad d(\mathbf{x}, \mathcal{B}) = \inf_{\mathbf{b} \in \mathcal{B}} d(\mathbf{x}, \mathbf{b})$$

Motivation: Boundary Average Treatment Effect



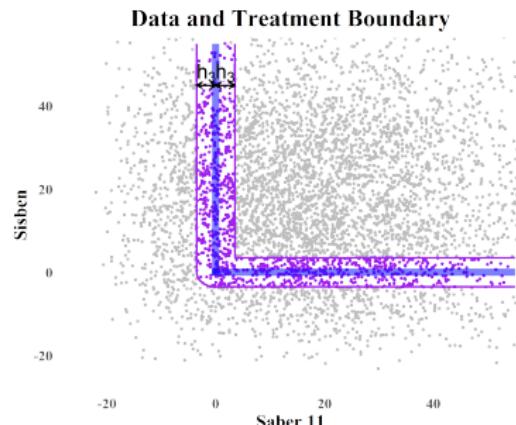
- Basic approach + boundary-segment FE + poly-expansion:

$$(\widehat{\boldsymbol{\zeta}}, \widehat{\boldsymbol{\tau}}, \widehat{\boldsymbol{\beta}}) = \arg \min_{\boldsymbol{\zeta}, \boldsymbol{\tau}, \boldsymbol{\beta}} \sum_{i=1}^n (Y_i - \boldsymbol{\nu}(S_i)^\top \boldsymbol{\zeta} - T_i \boldsymbol{\tau} - \mathbf{q}_p(D_i, T_i)^\top \boldsymbol{\beta})^2 \mathbb{1}(|D_i| \leq h),$$

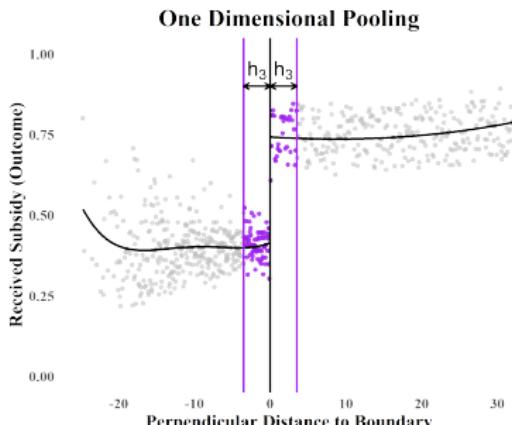
$$\boldsymbol{\nu}_L(S_i) = [\mathbb{1}(S_i = 1), \dots, \mathbb{1}(S_i = L)]^\top, \quad S_i = \arg \min_{1 \leq \ell \leq L} d(\mathbf{X}_i, \mathcal{B}_\ell), \quad \mathcal{B} = \sqcup_{1 \leq \ell \leq L} \mathcal{B}_\ell$$

$$\mathbf{q}_p(u, t) = [(u, u^2, \dots, u^p)^\top, t(u, u^2, \dots, u^p)^\top]^\top$$

Motivation: Boundary Average Treatment Effect



Saber 11

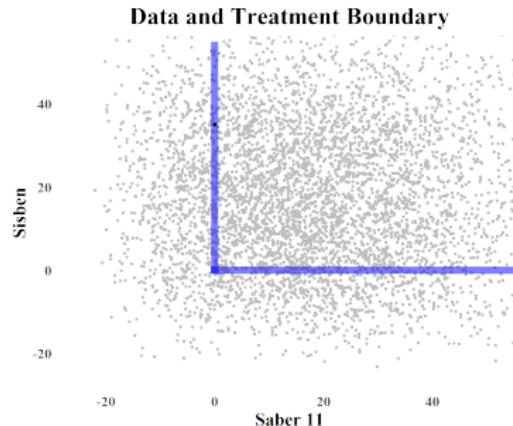


- ▶ Identification/Interpretation:

$$\hat{\tau} \rightarrow_{\mathbb{P}} \tau = \frac{\int_{\mathcal{B}} \tau(\mathbf{b}) f(\mathbf{b}) d\mathbf{b}}{\int_{\mathcal{B}} f(\mathbf{b}) d\mathbf{b}}, \quad \tau(\mathbf{b}) = \mathbb{E}[Y_i(1) - Y_i(0) | \mathbf{X}_i = \mathbf{b}].$$

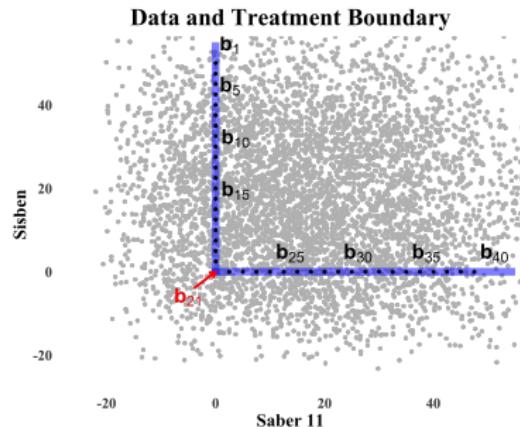
- ▶ How to define the above integral/convergence appropriately?
- ▶ What conditions on \mathcal{B} and $f(\cdot)$ are needed/sufficient?
- ▶ Estimation and inference results to help practice.

Motivation: *Ser Pilo Paga* (SPP) Colombian policy program



- ▶ High-school graduates $i = 1, 2, \dots, n$ offered cash transfer to attend college ($T_i = 1$).
- ▶ $\mathbf{X}_i = (\text{SABER11}_i, \text{SISBEN}_i)^\top$; SABER11_i = exam score; SISBEN_i = wealth index.
- ▶ $\mathcal{B} = \{\text{SABER11} \geq 0 \text{ and } \text{SISBEN} = 0\} \cup \{\text{SABER11} = 0 \text{ and } \text{SISBEN} \geq 0\}$.
- ▶ $Y_i = 1$ if first year of college completed, $= 0$ otherwise.

Motivation: Distance-Based vs. Location-Based



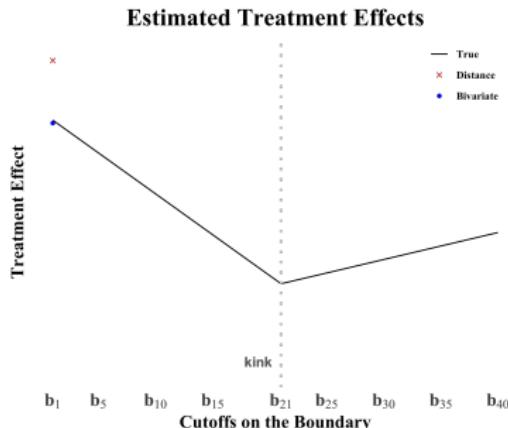
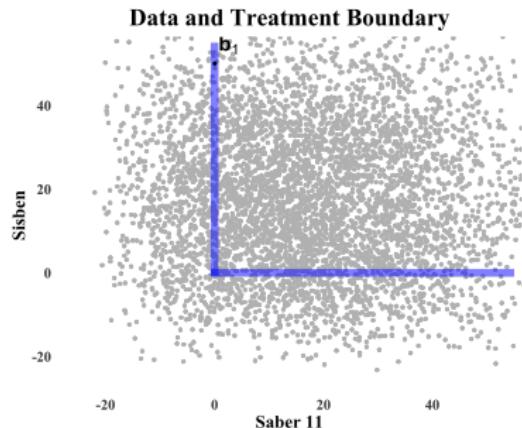
- ▶ Average treatment effect curve along the boundary:

$$\tau(\mathbf{x}) = \mathbb{E}[Y_i(1) - Y_i(0) | \mathbf{X}_i = \mathbf{x}], \quad \mathbf{x} \in \mathcal{B}.$$

- ▶ Estimation and Inference Approaches:

- ▶ Local regression based on univariate distance to point on boundary.
- ▶ Local regression based on bivariate location relative to point on boundary.

Motivation: Distance-Based vs. Location-Based



- Distance-based Estimator:

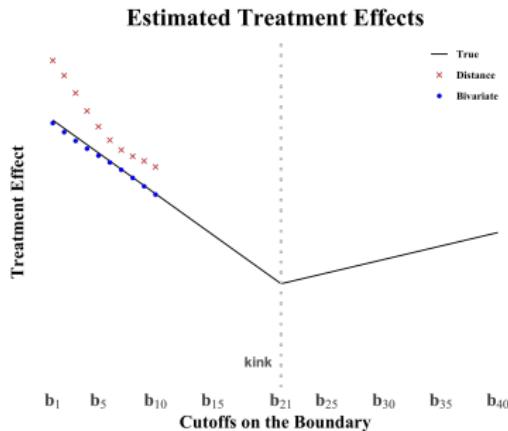
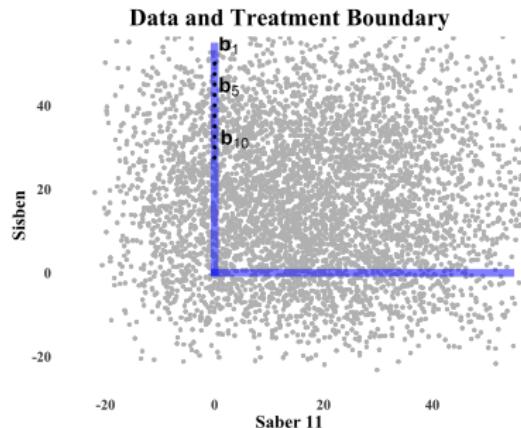
$$(\hat{\zeta}, \hat{\tau}_{\text{dis}}(\mathbf{b}_j), \hat{\beta}) = \arg \min_{\zeta, \tau, \beta} \sum_{i=1}^n (Y_i - \zeta - T_i \tau - \mathbf{q}_p(D_i(\mathbf{b}_j), T_i)^\top \beta)^2 k\left(\frac{D_i(\mathbf{b}_j)}{h}\right),$$

- Location-based Estimator:

$$(\hat{\zeta}, \hat{\tau}(\mathbf{b}_j), \hat{\beta}) = \arg \min_{\zeta, \tau, \beta} \sum_{i=1}^n (Y_i - \zeta - T_i \tau - \mathbf{q}_p(\mathbf{X}_i - \mathbf{b}_j, T_i)^\top \beta)^2 K\left(\frac{\mathbf{X}_i - \mathbf{b}_j}{h}\right),$$

- Evaluation points along \mathcal{B} : $\{\mathbf{b}_1\}$.

Motivation: Distance-Based vs. Location-Based



- Distance-based Estimator:

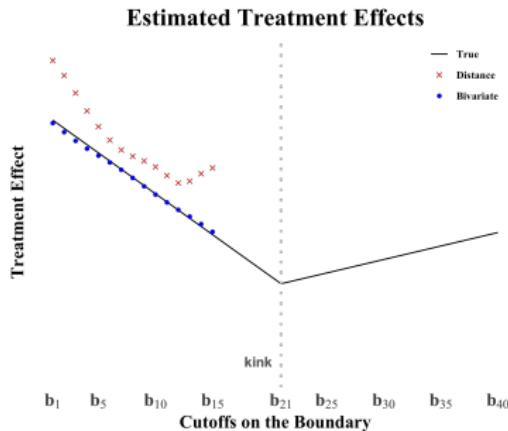
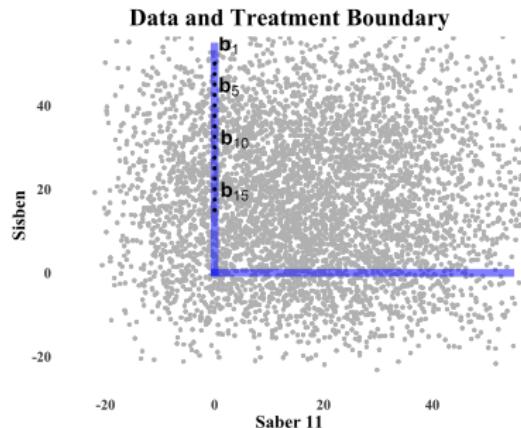
$$(\hat{\zeta}, \hat{\tau}_{\text{dis}}(\mathbf{b}_j), \hat{\beta}) = \arg \min_{\zeta, \tau, \beta} \sum_{i=1}^n (Y_i - \zeta - T_i \tau - \mathbf{q}_p(D_i(\mathbf{b}_j), T_i)^\top \beta)^2 k\left(\frac{D_i(\mathbf{b}_j)}{h}\right),$$

- Location-based Estimator:

$$(\hat{\zeta}, \hat{\tau}(\mathbf{b}_j), \hat{\beta}) = \arg \min_{\zeta, \tau, \beta} \sum_{i=1}^n (Y_i - \zeta - T_i \tau - \mathbf{q}_p(\mathbf{X}_i - \mathbf{b}_j, T_i)^\top \beta)^2 K\left(\frac{\mathbf{X}_i - \mathbf{b}_j}{h}\right),$$

- Evaluation points along \mathcal{B} : $\{\mathbf{b}_1, \dots, \mathbf{b}_{10}\}$.

Motivation: Distance-Based vs. Location-Based



- Distance-based Estimator:

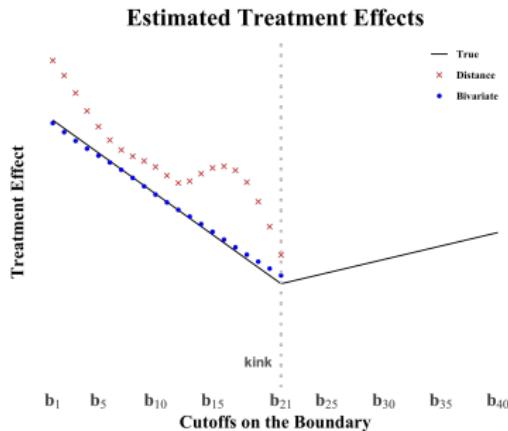
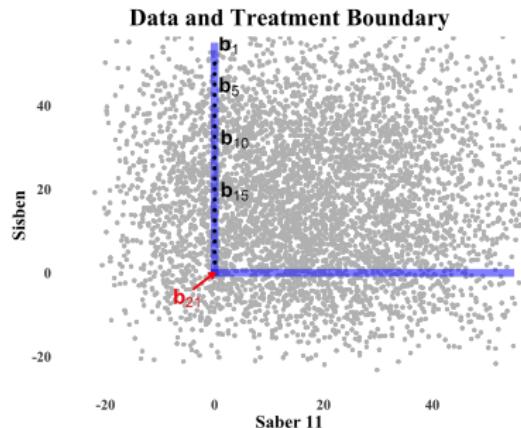
$$(\hat{\zeta}, \hat{\tau}_{\text{dis}}(\mathbf{b}_j), \hat{\beta}) = \arg \min_{\zeta, \tau, \beta} \sum_{i=1}^n (Y_i - \zeta - T_i \tau - \mathbf{q}_p(D_i(\mathbf{b}_j), T_i)^\top \beta)^2 k\left(\frac{D_i(\mathbf{b}_j)}{h}\right),$$

- Location-based Estimator:

$$(\hat{\zeta}, \hat{\tau}(\mathbf{b}_j), \hat{\beta}) = \arg \min_{\zeta, \tau, \beta} \sum_{i=1}^n (Y_i - \zeta - T_i \tau - \mathbf{q}_p(\mathbf{X}_i - \mathbf{b}_j, T_i)^\top \beta)^2 K\left(\frac{\mathbf{X}_i - \mathbf{b}_j}{h}\right),$$

- Evaluation points along \mathcal{B} : $\{\mathbf{b}_1, \dots, \mathbf{b}_{10}, \dots, \mathbf{b}_{15}\}$.

Motivation: Distance-Based vs. Location-Based



- Distance-based Estimator:

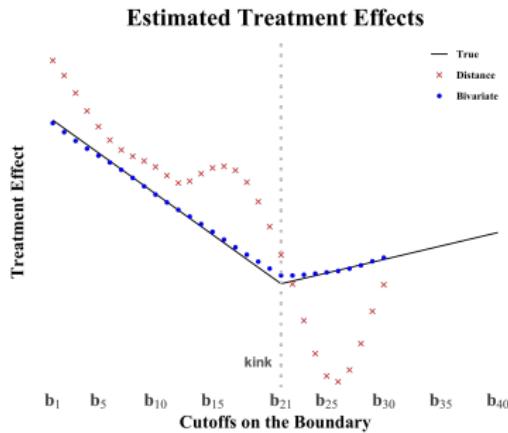
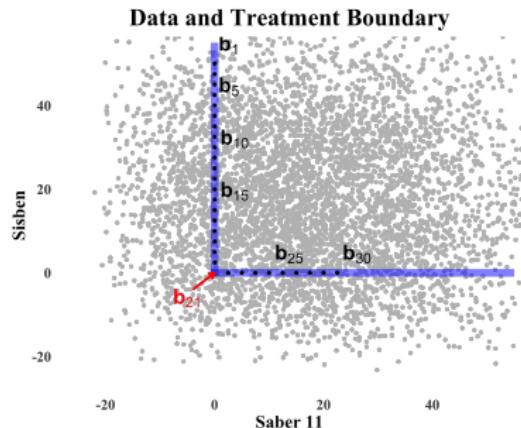
$$(\hat{\zeta}, \hat{\tau}_{\text{dis}}(\mathbf{b}_j), \hat{\beta}) = \arg \min_{\zeta, \tau, \beta} \sum_{i=1}^n (Y_i - \zeta - T_i \tau - \mathbf{q}_p(D_i(\mathbf{b}_j), T_i)^\top \beta)^2 k\left(\frac{D_i(\mathbf{b}_j)}{h}\right),$$

- Location-based Estimator:

$$(\hat{\zeta}, \hat{\tau}(\mathbf{b}_j), \hat{\beta}) = \arg \min_{\zeta, \tau, \beta} \sum_{i=1}^n (Y_i - \zeta - T_i \tau - \mathbf{q}_p(\mathbf{X}_i - \mathbf{b}_j, T_i)^\top \beta)^2 K\left(\frac{\mathbf{X}_i - \mathbf{b}_j}{h}\right),$$

- Evaluation points along \mathcal{B} : $\{\mathbf{b}_1, \dots, \mathbf{b}_{10}, \dots, \mathbf{b}_{15}, \dots, \mathbf{b}_{21}\}$.

Motivation: Distance-Based vs. Location-Based



- Distance-based Estimator:

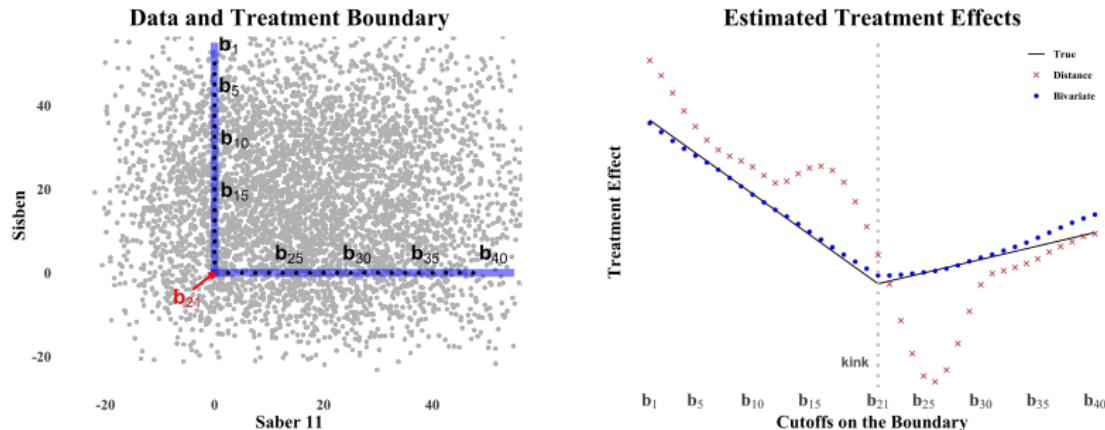
$$(\hat{\zeta}, \hat{\tau}_{\text{dis}}(\mathbf{b}_j), \hat{\beta}) = \arg \min_{\zeta, \tau, \beta} \sum_{i=1}^n (Y_i - \zeta - T_i \tau - \mathbf{q}_p(D_i(\mathbf{b}_j), T_i)^\top \beta)^2 k\left(\frac{D_i(\mathbf{b}_j)}{h}\right),$$

- Location-based Estimator:

$$(\hat{\zeta}, \hat{\tau}(\mathbf{b}_j), \hat{\beta}) = \arg \min_{\zeta, \tau, \beta} \sum_{i=1}^n (Y_i - \zeta - T_i \tau - \mathbf{q}_p(\mathbf{X}_i - \mathbf{b}_j, T_i)^\top \beta)^2 K\left(\frac{\mathbf{X}_i - \mathbf{b}_j}{h}\right),$$

- Evaluation points along \mathcal{B} : $\{\mathbf{b}_1, \dots, \mathbf{b}_{10}, \dots, \mathbf{b}_{15}, \dots, \mathbf{b}_{21}, \dots, \mathbf{b}_{30}\}$.

Motivation: Distance-Based vs. Location-Based



► Distance-based Estimator:

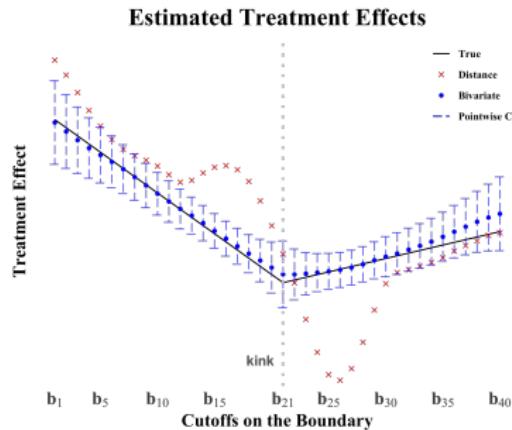
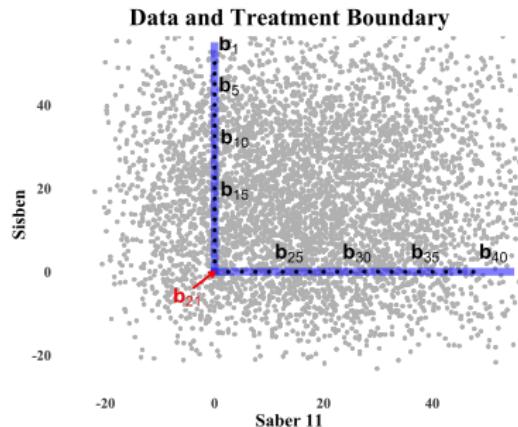
$$(\hat{\zeta}, \hat{\tau}_{\text{dis}}(\mathbf{b}_j), \hat{\beta}) = \arg \min_{\zeta, \tau, \beta} \sum_{i=1}^n (Y_i - \zeta - T_i \tau - \mathbf{q}_p(D_i(\mathbf{b}_j), T_i)^\top \beta)^2 k\left(\frac{D_i(\mathbf{b}_j)}{h}\right),$$

► Location-based Estimator:

$$(\hat{\zeta}, \hat{\tau}(\mathbf{b}_j), \hat{\beta}) = \arg \min_{\zeta, \tau, \beta} \sum_{i=1}^n (Y_i - \zeta - T_i \tau - \mathbf{q}_p(\mathbf{X}_i - \mathbf{b}_j, T_i)^\top \beta)^2 K\left(\frac{\mathbf{X}_i - \mathbf{b}_j}{h}\right),$$

► Evaluation points along \mathcal{B} : $\{\mathbf{b}_1, \dots, \mathbf{b}_{10}, \dots, \mathbf{b}_{15}, \dots, \mathbf{b}_{21}, \dots, \mathbf{b}_{30}, \dots, \mathbf{b}_{40}\}$.

Motivation: Distance-Based vs. Location-Based

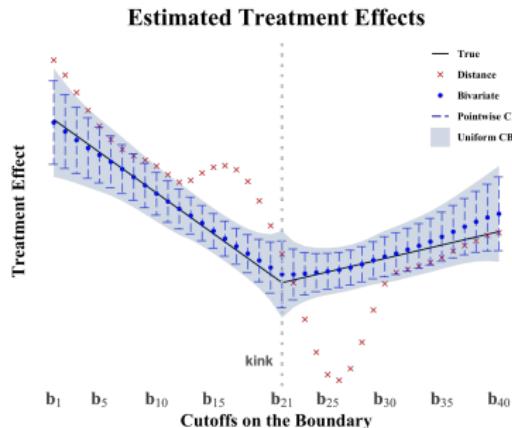
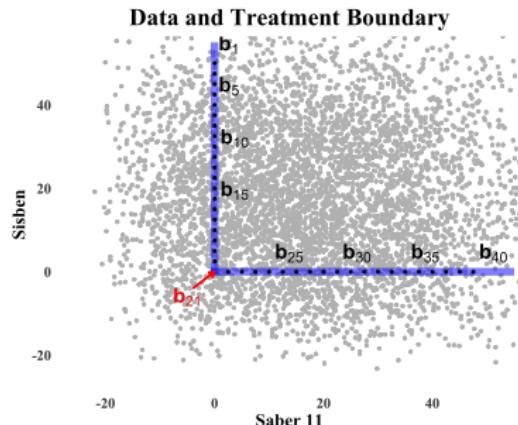


- ▶ Estimators: $\hat{\tau}_{\text{dis}}(\mathbf{x})$ and $\hat{\tau}(\mathbf{x})$, for each $\mathbf{x} \in \{b_1, \dots, b_{40}\}$.
- ▶ Uncertainty Quantification: Confidence Intervals. For each $\mathbf{x} \in \{b_1, \dots, b_{40}\}$,

$$\widehat{\mathbb{I}}(\mathbf{x}; \alpha) = \left[\hat{\tau}(\mathbf{x}) - \varphi_\alpha \sqrt{\widehat{\Omega}_{\mathbf{x}}}, \hat{\tau}(\mathbf{x}) + \varphi_\alpha \sqrt{\widehat{\Omega}_{\mathbf{x}}} \right].$$

- ▶ $\varphi_\alpha = \Phi^{-1}(1 - \alpha/2)$, where $\Phi(x)$ be the standard Gaussian CDF.
- ▶ $\varphi_{0.95} \approx 1.96$.

Motivation: Distance-Based vs. Location-Based



- ▶ Estimators: $\hat{\tau}_{\text{dis}}(\mathbf{x})$ and $\hat{\tau}(\mathbf{x})$, uniformly in $\mathbf{x} \in \mathcal{B}$.
- ▶ Uncertainty Quantification: Confidence Bands. Uniformly in $\mathbf{x} \in \mathcal{B}$,

$$\hat{\mathbb{I}}(\mathbf{x}; \alpha) = \left[\hat{\tau}(\mathbf{x}) - q_\alpha \sqrt{\hat{\Omega}_{\mathbf{x}}}, \hat{\tau}(\mathbf{x}) + q_\alpha \sqrt{\hat{\Omega}_{\mathbf{x}}} \right].$$

- ▶ $q_\alpha = \inf\{c > 0 : \mathbb{P}[\sup_{\mathbf{x} \in \mathcal{B}} |\hat{Z}_n(\mathbf{x})| \geq c | \text{data}] \leq \alpha\}$.
- ▶ $(\hat{Z}_n : \mathbf{x} \in \mathcal{B})$ is a Gaussian process conditional on data.

Summary of Results

- ▶ Pooled Boundary Average Treatment Effects: $\hat{\tau}$.
 1. Necessary and sufficient conditions for identification.
 2. Sufficient conditions for local polynomial debiasing.
 3. Estimation and inference.
 4. Similarities and differences with standard univariate RD designs.
- ▶ Univariate distance-based methods: $\hat{\tau}_{\text{dis}}(\mathbf{x})$.
 1. Sufficient conditions for identification.
 2. “Large” misspecification bias when \mathcal{B} is non-smooth (e.g., near a kink).
 3. “Small” misspecification bias when \mathcal{B} is smooth.
 4. Pointwise and uniform convergence rates and distribution theory.
 5. Discuss connects and differences with standard univariate RD designs.
- ▶ Bivariate location-based methods: $\hat{\tau}(\mathbf{x})$.
 1. Identification, estimation, and inference (pointwise and uniform over \mathcal{B}) are standard.
 2. Additional (mild) regularity on \mathcal{B} is needed.
 3. New methods for analysis of Boundary Discontinuity Designs.
- ▶ Tools: convergence over manifolds, minimax estimation, and strong approximation.

Outline

1. Introduction
2. Boundary Average Treatment Effects
3. Distance-Based Methods
4. Location-Based Methods
5. Empirical Application
6. Conclusion

BATE: Technical Setup

- **Goal:** Properly define existence of and converge to

$$\tau_w = \frac{\int_{\mathcal{B}} \tau(\mathbf{b}) w(\mathbf{b}) d\mathbf{b}}{\int_{\mathcal{B}} w(\mathbf{b}) d\mathbf{b}}, \quad \tau(\mathbf{b}) = \mathbb{E}[Y_i(1) - Y_i(0) | \mathbf{X}_i = \mathbf{b}].$$

- **Challenge:** Assignment boundary \mathcal{B} can be non-smooth.

- Recall: $D_i = (2 \cdot T_i - 1)d(\mathbf{X}_i, \mathcal{B})$ distance to boundary \mathcal{B} ; $d(\mathbf{x}, \mathcal{B}) = \inf_{\mathbf{b} \in \mathcal{B}} d(\mathbf{x}, \mathbf{b})$.

- **Approach:** Geometric measure theory.

- \mathcal{B} is a 1-dimensional rectifiable curve.
- $\int_{\mathcal{B}} m(\mathbf{b}) d\mathbf{b} = \int_{\mathcal{B}} m d\mathfrak{H}^1$, with \mathfrak{H}^1 the 1-dimensional Hausdorff measure.

- **Question:** When does the following limit of integrals is well-defined?

$$\lim_{\epsilon \downarrow 0} \int_{\mathcal{T}(\epsilon)} \frac{1}{\epsilon} g\left(\frac{d(\mathbf{x}, \mathcal{B})}{\epsilon}\right) m(\mathbf{x}) d\mathbf{x},$$

where

$$\mathcal{T}(\epsilon) = \{\mathbf{x} \in \mathcal{X} : d(\mathbf{x}, \mathcal{B}) \leq \epsilon\}, \quad \epsilon \geq 0.$$

BATE: Workhorse Technical Tool

$$\mathcal{T}(\epsilon) = \{\mathbf{x} \in \mathcal{X} : d(\mathbf{x}, \mathcal{B}) \leq \epsilon\}, \quad d(\mathbf{x}, \mathcal{B}) = \inf_{\mathbf{b} \in \mathcal{B}} d(\mathbf{x}, \mathbf{b}).$$

► Technical Lemma.

- $d : \mathbb{R}^2 \mapsto [0, \infty)$ satisfies $\|\mathbf{x}_1 - \mathbf{x}_2\| \lesssim d(\mathbf{x}_1, \mathbf{x}_2) \lesssim \|\mathbf{x}_1 - \mathbf{x}_2\|$ for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$.
- $J_2 d(\cdot, \mathcal{B}) \neq 0$ \mathfrak{m} -almost everywhere on \mathcal{X} , and $\int_{\mathcal{X}} |\frac{m(\mathbf{x})}{J_2 d(\mathbf{x}, \mathcal{B})}| d\mathbf{x} < \infty$.
- There exists a constant $c_{\mathcal{B}}$ such that $\lim_{\epsilon \downarrow 0} M(\epsilon) = c_{\mathcal{B}} \cdot M(0)$ is finite, where

$$M(r) = \int_{d(\mathbf{x}, \mathcal{B})=r} \frac{m(\mathbf{x})}{J_2 d(\mathbf{x}, \mathcal{B})} d\mathfrak{H}^1(\mathbf{x}), \quad r \geq 0.$$

Then,

$$\lim_{\epsilon \downarrow 0} \int_{\mathcal{T}(\epsilon)} \frac{1}{\epsilon} g\left(\frac{d(\mathbf{x}, \mathcal{B})}{\epsilon}\right) m(\mathbf{x}) d\mathbf{x} = c_{\mathcal{B}} \cdot \int_0^1 g(s) ds \cdot \int_{\mathcal{B}} \frac{m(\mathbf{x})}{J_2 d(\mathbf{x}, \mathcal{B})} d\mathfrak{H}^1(\mathbf{x}).$$

BATE: Workhorse Technical Tool – Proof

- ▶ The function $\mathbf{x} \mapsto d(\mathbf{x}, \mathcal{B})$ is C_u -Lipschitz.
- ▶ The level sets $\mathcal{L}(\epsilon) = \{\mathbf{x} \in \mathcal{X} : d(\mathbf{x}, \mathcal{B}) = \epsilon\}$, $\epsilon \geq 0$, are 1-dimensional rectifiable sets.
- ▶ The function $\mathbf{x} \mapsto \epsilon^{-1} g\left(\frac{d(\mathbf{x}, \mathcal{B})}{\epsilon}\right) m(\mathbf{x}) \text{J}_2 d(\mathbf{x}, \mathcal{B})^{-1}$ is m -summable over \mathcal{X} .
- ▶ Using the **Coarea formula**,

$$\begin{aligned}
 \lim_{\epsilon \downarrow 0} \int_{\mathcal{T}(\epsilon)} \frac{1}{\epsilon} g\left(\frac{d(\mathbf{x}, \mathcal{B})}{\epsilon}\right) m(\mathbf{x}) d\mathbf{x} &= \lim_{\epsilon \downarrow 0} \int_{\mathcal{T}(\epsilon)} \frac{1}{\epsilon} g\left(\frac{d(\mathbf{x}, \mathcal{B})}{\epsilon}\right) m(\mathbf{x}) \frac{1}{\text{J}_2 d(\mathbf{x}, \mathcal{B})} \text{J}_2 d(\mathbf{x}, \mathcal{B}) d\mathbf{x} \\
 &= \lim_{\epsilon \downarrow 0} \int_0^\epsilon \int_{d(\mathbf{x}, \mathcal{B})=r} \frac{1}{\epsilon} g\left(\frac{r}{\epsilon}\right) \frac{m(\mathbf{x})}{\text{J}_2 d(\mathbf{x}, \mathcal{B})} d\mathfrak{H}^1(\mathbf{x}) dr \\
 &= \lim_{\epsilon \downarrow 0} \int_0^\epsilon \frac{1}{\epsilon} g\left(\frac{r}{\epsilon}\right) \textcolor{violet}{M}(r) dr \\
 &= \lim_{\epsilon \downarrow 0} \int_0^1 g(u) \textcolor{violet}{M}(\epsilon u) du \\
 &= \mathbf{c}_{\mathcal{B}} \cdot \int_0^1 g(s) ds \cdot \textcolor{violet}{M}(0),
 \end{aligned}$$

BATE: Identification, Estimation, and Inference

- Basic approach + poly-expansions of D_i :

$$(\widehat{\zeta}, \widehat{\tau}, \widehat{\beta}) = \arg \min_{\zeta, \tau, \beta} \sum_{i=1}^n (Y_i - \zeta - T_i \tau - \mathbf{q}_p(D_i, T_i)^\top \beta)^2 \mathbf{1}(|D_i| \leq h).$$

- **Identification.**

- (i) $r \mapsto \mathbb{E}[Y_i | D_i = r, T_i = t] = \mathbb{E}[Y_i(t) | \mathcal{A}(\mathbf{X}_i, \mathcal{B}) = |r|]$ is continuous at 0, for $t = 0, 1$.
- (ii) $J_2 d(\cdot, \mathcal{B}) \neq 0$ \mathfrak{m} -almost everywhere on \mathcal{X} , and $\int_{\mathcal{X}} |\frac{f(\mathbf{x})}{J_2 d(\mathbf{x}, \mathcal{B})}| d\mathbf{x} < \infty$.
- (iii) $\lim_{\epsilon \downarrow 0} F(\epsilon) = F(0)$ is finite, where $F(r) = \int_{d(\mathbf{x}, \mathcal{B})=r} \frac{f(\mathbf{x})}{J_2 d(\mathbf{x}, \mathcal{B})} d\mathfrak{H}^1(\mathbf{x})$, $r \geq 0$.

Then,

$$\widehat{\tau} \xrightarrow{\mathbb{P}} \tau = \frac{\int_{\mathcal{B}} \tau(\mathbf{b}) f(\mathbf{b}) d\mathbf{b}}{\int_{\mathcal{B}} f(\mathbf{b}) d\mathbf{b}}, \quad \tau(\mathbf{b}) = \mathbb{E}[Y_i(1) - Y_i(0) | \mathbf{X}_i = \mathbf{b}]$$

- **Convergence Rate and Feasible CLT.**

$$|\widehat{\tau} - \tau| \lesssim_{\mathbb{P}} \sqrt{\frac{1}{nh}} + |\mathfrak{B}_n|, \quad \frac{\widehat{\tau} - \mathbb{E}[\widehat{\tau} | \mathbf{D}]}{\sqrt{\widehat{\mathbb{V}}[\widehat{\tau} | \mathbf{D}]}} \rightsquigarrow \text{Normal}(0, 1).$$

- **Valid Inference.** Need to understand bias \mathfrak{B}_n .

BATE: Verifying Sufficient Conditions + Extensions

► Identification: Sufficient conditions.

- ▶ $d(\mathbf{x}_1, \mathbf{x}_2) = \|\mathbf{x}_1 - \mathbf{x}_2\|$, and \mathcal{B} is piecewise linear.
- ▶ $\mu_0(\mathbf{x}) = \mathbb{E}[Y(0)|\mathbf{X}_i = \mathbf{x}]$, $\mu_1(\mathbf{x}) = \mathbb{E}[Y(1)|\mathbf{X}_i = \mathbf{x}]$, and $f(\mathbf{x})$ are continuous on \mathcal{X} .

Then, identification conditions (i)–(iii) hold for small enough $\epsilon > 0$.

► Estimation and Inference (Bias): Sufficient conditions.

- ▶ $d(\mathbf{x}_1, \mathbf{x}_2) = \|\mathbf{x}_1 - \mathbf{x}_2\|$, and \mathcal{B} is piecewise linear.
- ▶ μ_0 , μ_1 , and f are s -times continuously differentiable on \mathcal{X} .

Then,

$$r \mapsto \mathbb{E}[Y_i|D_i = r, T_i = t] = \mathbb{E}[Y_i(t)|d(\mathbf{X}_i, \mathcal{B}) = |r|]$$

is s -times continuously differentiable on $[0, \epsilon]$. Therefore,

$$|\mathfrak{B}_n| \lesssim_{\mathbb{P}} h^{\min\{s, p+1\}}$$

- ### ► Extensions.
- Boundary-segment FE, fuzzy (IV) setting, etc.

Outline

1. Introduction
2. Boundary Average Treatment Effects
3. Distance-Based Methods
4. Location-Based Methods
5. Empirical Application
6. Conclusion

Distance-Based Methods: Identification

- ▶ **Parameter:** $\tau(\mathbf{x}) = \mathbb{E}[Y_i(1) - Y_i(0)|\mathbf{X}_i = \mathbf{x}]$ for all $\mathbf{x} \in \mathcal{B}$.

- ▶ **Estimator:**

$$(\hat{\zeta}, \hat{\tau}_{\text{dis}}(\mathbf{b}), \hat{\beta}) = \arg \min_{\zeta, \tau, \beta} \sum_{i=1}^n (Y_i - \zeta - T_i \tau - \mathbf{q}_p(D_i(\mathbf{b}), T_i)^\top \beta)^2 k\left(\frac{D_i(\mathbf{b})}{h}\right),$$

- ▶ $D_i(\mathbf{b}) = (2 \cdot T_i - 1) d(\mathbf{x}, \mathbf{b})$ distance to point $\mathbf{b} \in \mathcal{B}$.
- ▶ **Assumptions:** Let $t \in \{0, 1\}$.

- ▶ $d : \mathbb{R}^2 \mapsto [0, \infty)$ satisfies $\|\mathbf{x}_1 - \mathbf{x}_2\| \lesssim d(\mathbf{x}_1, \mathbf{x}_2) \lesssim \|\mathbf{x}_1 - \mathbf{x}_2\|$ for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$.
- ▶ $k : \mathbb{R} \rightarrow [0, \infty)$ is compact supported and Lipschitz continuous, or $k(u) = \mathbf{1}(|u| \leq 1)$.
- ▶ $\liminf_{h \downarrow 0} \inf_{\mathbf{x} \in \mathcal{B}} \int_{\mathcal{A}_t} k(d(\mathbf{u}, \mathbf{x})/h) d\mathbf{u} \gtrsim 1$.

- ▶ **Identification.** For all $\mathbf{b} \in \mathcal{B}$,

$$\begin{aligned}\tau(\mathbf{b}) &= \mathbb{E}[Y_i(1) - Y_i(0)|\mathbf{X}_i = \mathbf{b}] \\ &= \lim_{r \downarrow 0} \mathbb{E}[Y_i|D_i(\mathbf{b}) = r, T_i = t] - \lim_{r \uparrow 0} \mathbb{E}[Y_i|D_i(\mathbf{b}) = r, T_i = t]\end{aligned}$$

Distance-Based Methods: Estimation and Inference Results

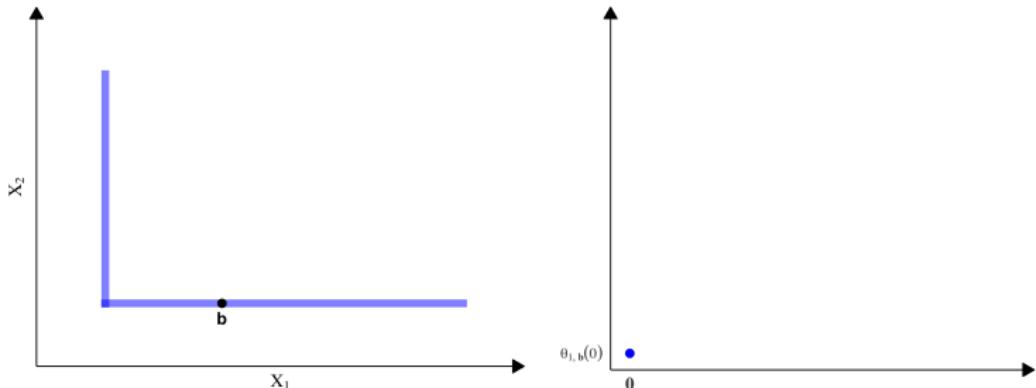
► Convergence Rates.

$$|\widehat{\tau}_{\text{dis}}(\mathbf{x}) - \tau(\mathbf{x})| \lesssim_{\mathbb{P}} \sqrt{\frac{1}{nh^2}} + |\mathfrak{B}_n(\mathbf{x})|, \quad \mathbf{x} \in \mathcal{B}.$$

$$\sup_{\mathbf{x} \in \mathcal{B}} |\widehat{\tau}_{\text{dis}}(\mathbf{x}) - \tau(\mathbf{x})| \lesssim_{\mathbb{P}} \sqrt{\frac{\log n}{nh^2}} + \sup_{\mathbf{x} \in \mathcal{B}} |\mathfrak{B}_n(\mathbf{x})|.$$

- Putting aside the potential bias problem when \mathcal{B} is non-smooth.
 - **Pointwise Inference:** Fairly standard, up to handling \mathcal{B} .
 - **Uniform Inference:** Requires new technical tools, and careful handling of \mathcal{B} .
 - De Giorgi Perimeter: captures notion of “wiggleness” of \mathcal{B} .
- **Practice.** Valid and invalid practices based on standard univariate RD methods.

Distance-Based Methods: Implied Smoothness and Bias



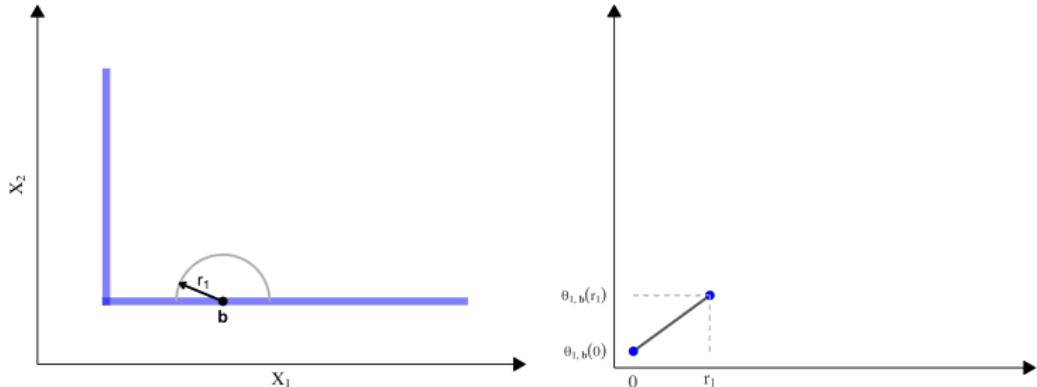
- ▶ **Treatment Group.** Bivariate vs. univariate (distance-induced) expectations:

$$\mathbb{E}[Y(1)|\mathbf{X}_i = \mathbf{b}, \mathbf{X}_i \in \mathcal{A}_1] = \lim_{r \downarrow 0} \theta_{1,\mathbf{b}}(r)$$

where

$$\theta_{t,\mathbf{b}}(r) = \mathbb{E}[Y_i|D_i(\mathbf{b}) = r, D_i(\mathbf{b}) \geq 0] = \mathbb{E}[Y_i(1)|\mathcal{A}(\mathbf{X}_i, \mathbf{x}) = r].$$

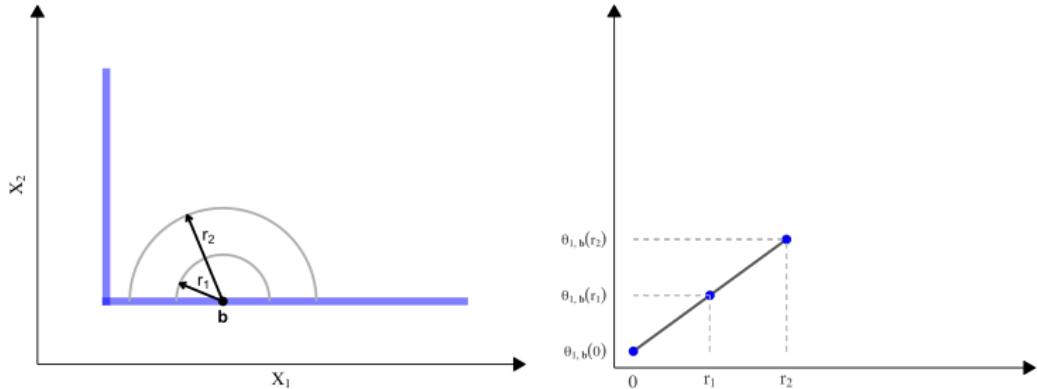
Distance-Based Methods: Implied Smoothness and Bias



- **Distance range:** $[0, r_1]$. If $\mathbf{x} \mapsto \mathbb{E}[Y(1)|\mathbf{X}_i = \mathbf{b}, \mathbf{X}_i \in \mathcal{A}_1]$ is **smooth**, then
$$r \mapsto \theta_{t,\mathbf{b}}(r) = \mathbb{E}[Y_i|D_i(\mathbf{b}) = r, D_i(\mathbf{b}) \geq 0] = \mathbb{E}[Y_i(1)|d(\mathbf{X}_i, \mathbf{x}) = r]$$
is also **smooth**.
- Thus, distance-based local polynomial estimator misspecification bias is

$$|\mathfrak{B}_n(\mathbf{x})| \lesssim_{\mathbb{P}} h^{p+1}.$$

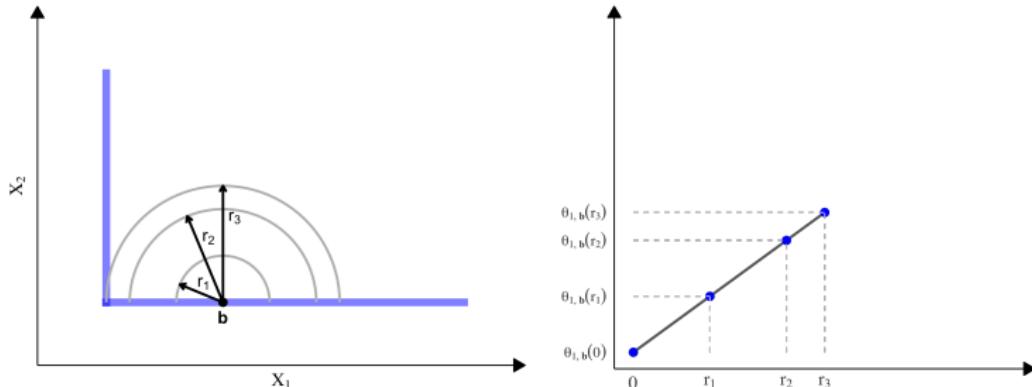
Distance-Based Methods: Implied Smoothness and Bias



- **Distance range:** $[0, r_2]$. If $\mathbf{x} \mapsto \mathbb{E}[Y(1)|\mathbf{X}_i = \mathbf{b}, \mathbf{X}_i \in \mathcal{A}_1]$ is **smooth**, then
$$r \mapsto \theta_{t,\mathbf{b}}(r) = \mathbb{E}[Y_i|D_i(\mathbf{b}) = r, D_i(\mathbf{b}) \geq 0] = \mathbb{E}[Y_i(1)|d(\mathbf{X}_i, \mathbf{x}) = r]$$
is also **smooth**.
- Thus, distance-based local polynomial estimator misspecification bias is

$$|\mathfrak{B}_n(\mathbf{x})| \lesssim_{\mathbb{P}} h^{p+1}.$$

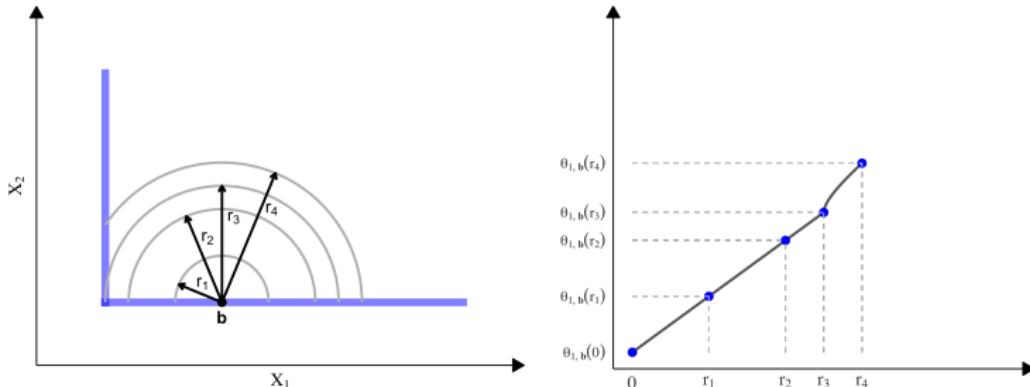
Distance-Based Methods: Implied Smoothness and Bias



- ▶ **Distance range:** $[0, r_3]$. If $\mathbf{x} \mapsto \mathbb{E}[Y(1)|\mathbf{X}_i = \mathbf{b}, \mathbf{X}_i \in \mathcal{A}_1]$ is **smooth**, then
$$r \mapsto \theta_{t,\mathbf{b}}(r) = \mathbb{E}[Y_i|D_i(\mathbf{b}) = r, D_i(\mathbf{b}) \geq 0] = \mathbb{E}[Y_i(1)|d(\mathbf{X}_i, \mathbf{x}) = r]$$
is also **smooth**.
- ▶ Thus, distance-based local polynomial estimator misspecification bias is

$$|\mathfrak{B}_n(\mathbf{x})| \lesssim_{\mathbb{P}} h^{p+1}.$$

Distance-Based Methods: Implied Smoothness and Bias



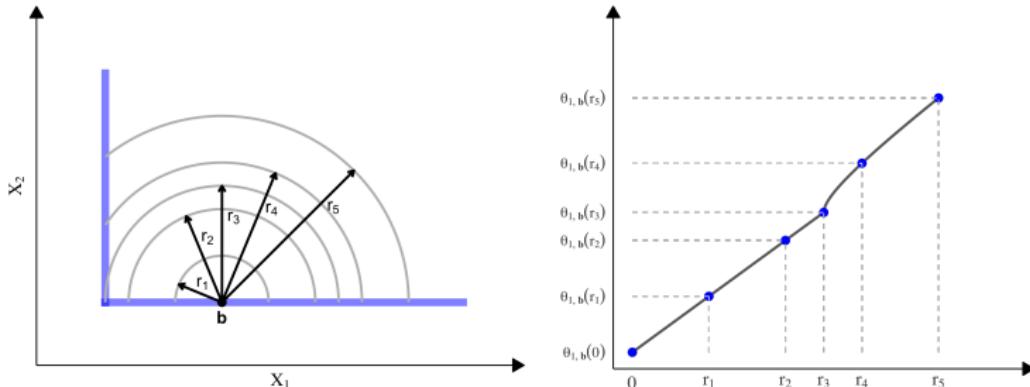
- **Distance range:** $[0, r_4]$. If $\mathbf{x} \mapsto \mathbb{E}[Y(1)|\mathbf{X}_i = \mathbf{b}, \mathbf{X}_i \in \mathcal{A}_1]$ is **smooth**, then

$$r \mapsto \theta_{t,\mathbf{b}}(r) = \mathbb{E}[Y_i|D_i(\mathbf{b}) = r, D_i(\mathbf{b}) \geq 0] = \mathbb{E}[Y_i(1)|d(\mathbf{X}_i, \mathbf{x}) = r]$$

is **non-smooth**.

- **Smoothness.** $r \mapsto \theta_{t,\mathbf{b}}(r)$ is **locally to zero Lipschitz**, regardless underlying smoothness.

Distance-Based Methods: Implied Smoothness and Bias



- **Distance range:** $[0, r_5]$. Distance-based local poly estimator misspecification bias is

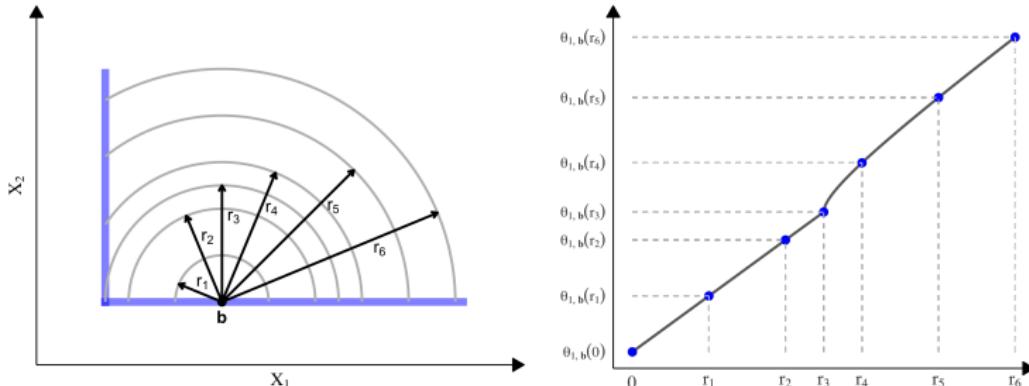
$$|\mathfrak{B}_n(\mathbf{x})| \lesssim_{\mathbb{P}} h$$

regardless p used! Not of order h^{p+1} as expected given standard smoothness.

- **Pointwise Analysis.** Need to choose bandwidth $h \leq r_3 = d(\mathbf{b}, \text{kink})$.

- Bandwidth must vary with $\mathbf{b} \in \mathcal{B}$, depending on “smoothness” of boundary!

Distance-Based Methods: Implied Smoothness and Bias



- **Uniform Analysis.** Under minimal regularity conditions, and for any $p \geq 1$,

$$1 \lesssim \liminf_{n \rightarrow \infty} \sup_{\mathbb{P} \in \mathcal{P}} \sup_{\mathbf{x} \in \mathcal{B}} \frac{\mathfrak{B}_n(\mathbf{x})}{h} \leq \limsup_{n \rightarrow \infty} \sup_{\mathbb{P} \in \mathcal{P}} \sup_{\mathbf{x} \in \mathcal{B}} \frac{\mathfrak{B}_n(\mathbf{x})}{h} \lesssim 1,$$

- Bias cannot be better than order h (Lipschitz continuity) if \mathcal{B} is non-smooth!
- If \mathcal{B} is smooth, then $\sup_{\mathbf{x} \in \mathcal{B}} |\mathfrak{B}_n(\mathbf{x})| \lesssim h^{p+1}$.

Distance-Based Methods: Minimax Result

- ▶ Is the “large” bias with non-smooth \mathcal{B} a general problem? **Yes!**
- ▶ **Impossibility Result.** Under standard regularity conditions:

$$\liminf_{n \rightarrow \infty} n^{1/4} \inf_{T_n \in \mathcal{T}} \sup_{\mathbb{P} \in \mathcal{P}_{NP}} \mathbb{E}_{\mathbb{P}} \left[\sup_{\mathbf{x} \in \mathcal{B}} |T_n(\mathbf{U}_n(\mathbf{x})) - \mu(\mathbf{x})| \right] \gtrsim 1,$$

where

- ▶ \mathcal{T} denotes the class of all distance-based estimators $T_n(\mathbf{U}_n(\mathbf{x}))$ with $\mathbf{U}_n(\mathbf{x}) = [(Y_i, D_i(\mathbf{x}) = \|\mathbf{X}_i - \mathbf{x}\|) : 1 \leq i \leq n]$ for each $\mathbf{x} \in \mathcal{X}$,
- ▶ \mathcal{B} is assumed to be rectifiable, and
- ▶ \mathcal{P}_{NP} includes q -smooth $\mu(\mathbf{x})$ functions.

- ▶ **Stone (1982).** Under the same conditions:

$$\liminf_{n \rightarrow \infty} \left(\frac{n}{\log n} \right)^{\frac{q}{2q+2}} \inf_{S_n \in \mathcal{S}} \sup_{\mathbb{P} \in \mathcal{P}_{NP}} \mathbb{E}_{\mathbb{P}} \left[\sup_{\mathbf{x} \in \mathcal{B}} |S_n(\mathbf{x}; \mathbf{W}_n) - \mu(\mathbf{x})| \right] \gtrsim 1,$$

where

- ▶ \mathcal{S} is the (unrestricted class) of all estimators based on $(\mathbf{W}_n = (Y_i, \mathbf{X}_i^\top)^\top : 1 \leq i \leq n)$.

Outline

1. Introduction
2. Boundary Average Treatment Effects
3. Distance-Based Methods
4. Location-Based Methods
5. Empirical Application
6. Conclusion

Location-Based Methods: Setup

- ▶ **Parameter:** $\tau(\mathbf{x}) = \mathbb{E}[Y_i(1) - Y_i(0)|\mathbf{X}_i = \mathbf{x}]$ for all $\mathbf{x} \in \mathcal{B}$.
- ▶ **Estimator:**

$$(\hat{\zeta}, \hat{\tau}(\mathbf{b}), \hat{\beta}) = \arg \min_{\zeta, \tau, \beta} \sum_{i=1}^n (Y_i - \zeta - T_i \tau - \mathbf{q}_p(\mathbf{X}_i - \mathbf{b}, T_i)^\top \beta)^2 K\left(\frac{\mathbf{X}_i - \mathbf{b}}{h}\right),$$

- ▶ **Assumptions:** Let $t \in \{0, 1\}$.
 - ▶ $K : \mathbb{R}^2 \rightarrow [0, \infty)$ compact supported & Lipschitz continuous, or $K(\mathbf{u}) = \mathbf{1}(\mathbf{u} \in [-1, 1]^2)$.
 - ▶ $\liminf_{h \downarrow 0} \inf_{\mathbf{x} \in \mathcal{B}} \int_{\mathcal{A}_t} K\left(\frac{\mathbf{u} - \mathbf{x}}{h}\right) d\mathbf{u} \gtrsim 1$.
- ▶ **Identification.** For all $\mathbf{b} \in \mathcal{B}$,

$$\tau(\mathbf{b}) = \lim_{\mathbf{x} \rightarrow \mathbf{b}, \mathbf{x} \in \mathcal{A}_1} \mathbb{E}[Y_i | \mathbf{X}_i = \mathbf{x}] - \lim_{\mathbf{x} \rightarrow \mathbf{b}, \mathbf{x} \in \mathcal{A}_0} \mathbb{E}[Y_i | \mathbf{X}_i = \mathbf{x}].$$

- ▶ This is standard from the literature.

Location-Based Methods: Point Estimation

- **Convergence Rates.** Under minimal regularity conditions,

$$|\hat{\tau}(\mathbf{x}) - \tau(\mathbf{x})| \lesssim_{\mathbb{P}} \sqrt{\frac{1}{nh^2}} + h^{p+1}, \quad \mathbf{x} \in \mathcal{B}.$$

$$\sup_{\mathbf{x} \in \mathcal{B}} |\hat{\tau}(\mathbf{x}) - \tau(\mathbf{x})| \lesssim_{\mathbb{P}} \sqrt{\frac{\log n}{nh^2}} + h^{p+1}.$$

- **MSE Expansions.** Under minimal regularity conditions,

$$\mathbb{E}[(\hat{\tau}(\mathbf{x}) - \tau(\mathbf{x}))^2 | \mathbf{X}] = h^{2(p+1)} \mathbf{B}_{\mathbf{x}}^2 + \frac{1}{nh^2} \mathbf{V}_{\mathbf{x}} \quad \mathbf{x} \in \mathcal{B}.$$

$$\int_{\mathcal{B}} \mathbb{E}[(\hat{\tau}(\mathbf{x}) - \tau(\mathbf{x}))^2 | \mathbf{X}] w(\mathbf{x}) d\mathbf{x} = h^{2(p+1)} \int_{\mathcal{B}} \mathbf{B}_{\mathbf{x}}^2 dw(\mathbf{x}) + \frac{1}{nh^2} \int_{\mathcal{B}} \mathbf{V}_{\mathbf{x}} w(\mathbf{x}) d\mathbf{x}.$$

- Standard bandwidth selection methods developed in the paper.

Location-Based Methods: Inference

► Uncertainty Quantification.

$$\widehat{I}(\mathbf{x}; \alpha) = \left[\widehat{\tau}(\mathbf{x}) - \varphi_\alpha \sqrt{\widehat{\Omega}_{\mathbf{x}}}, \widehat{\tau}(\mathbf{x}) + \varphi_\alpha \sqrt{\widehat{\Omega}_{\mathbf{x}}} \right], \quad \mathbf{x} \in \mathcal{B},$$

- **Confidence Interval.** By CLT, for each $\mathbf{x} \in \mathcal{B}$, set $\varphi_\alpha = \Phi^{-1}(1 - \alpha/2)$.
- **Confidence Band.**

$$\mathbb{P}[\tau(\mathbf{x}) \in \widehat{I}(\mathbf{x}; \alpha), \text{ for all } \mathbf{x} \in \mathcal{B}] = \mathbb{P}\left[\sup_{\mathbf{x} \in \mathcal{B}} |\widehat{T}(\mathbf{x})| \leq \varphi_\alpha\right].$$

1. Establish strong approximation for $(\widehat{T}(\mathbf{x}) : \mathbf{x} \in \mathcal{B})$ by $(\widehat{Z}_n : \mathbf{x} \in \mathcal{B})$, a Gaussian process conditional on data. Deduce the distribution of $\sup_{\mathbf{x} \in \mathcal{B}} |\widehat{T}(\mathbf{x})|$.
2. Using simulations, set $\varphi_\alpha = \inf\{c > 0 : \mathbb{P}[\sup_{\mathbf{x} \in \mathcal{B}} |\widehat{Z}_n(\mathbf{x})| \geq c | \text{data}] \leq \alpha\}$.
3. Key regularity condition: De Giorgi Perimeter of \mathcal{B} .

► Implementation and Bias.

- (I)MSE-optimal bandwidth selection for point estimation.
- Robust bias correction for inference.

Boundary Treatment Effects

► Recall: Pooling Approach.

$$\hat{\tau} \rightarrow_{\mathbb{P}} \tau = \frac{\int_{\mathcal{B}} \tau(\mathbf{b}) f(\mathbf{b}) d\mathbf{b}}{\int_{\mathcal{B}} f(\mathbf{b}) d\mathbf{b}}, \quad \tau(\mathbf{b}) = \mathbb{E}[Y_i(1) - Y_i(0) | \mathbf{X}_i = \mathbf{b}].$$

Location-Based Methods.

► Boundary Average Treatment Effects:

$$\hat{\tau}_{\text{loc}, \mathcal{B}} = \frac{\int_{\mathcal{B}} \hat{\tau}(\mathbf{b}) w(\mathbf{b}) d\mathbf{b}}{\int_{\mathcal{B}} w(\mathbf{b}) d\mathbf{b}} \approx \frac{\sum_{j=1}^J \hat{\tau}(\mathbf{b}_j) w(\mathbf{b}_j)}{\sum_{j=1}^J w(\mathbf{b}_j)} \rightarrow_{\mathbb{P}} \tau = \frac{\int_{\mathcal{B}} \tau(\mathbf{b}) w(\mathbf{b}) d\mathbf{b}}{\int_{\mathcal{B}} w(\mathbf{b}) d\mathbf{b}}.$$

- IMSE-optimal bandwidth choice is more natural.
- Choice of $w(\cdot)$ changes causal interpretation.
- Convergence rate may change from $\frac{1}{nh^2}$ to $\frac{1}{nh}$; distribution theory follows.
- Natural connection with pooled OLS analysis (for a specific choice of $w(\cdot)$).

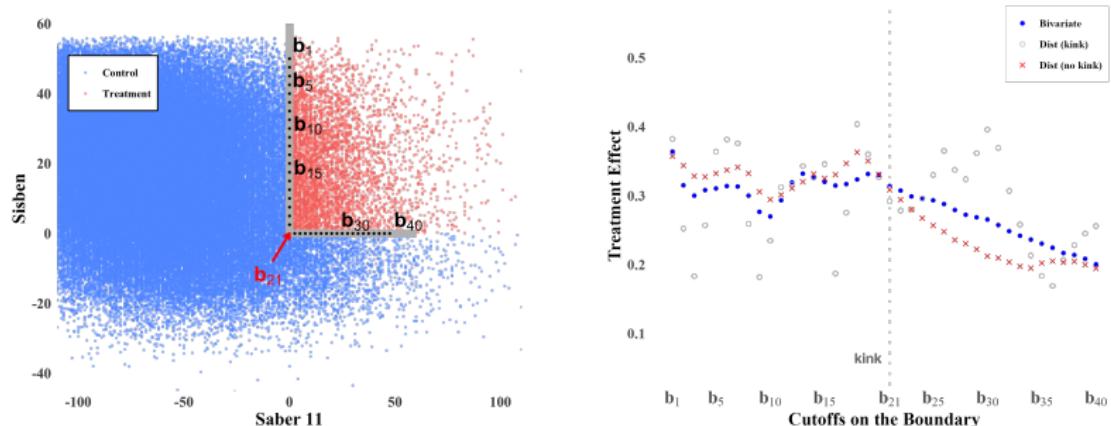
► Boundary Largest Treatment Effect:

$$\hat{\tau}_{\text{loc}, \max} = \sup_{\mathbf{b} \in \mathcal{B}} \hat{\tau}(\mathbf{b}) \approx \max_{j=1, \dots, J} \hat{\tau}(\mathbf{b}_j) \rightarrow_{\mathbb{P}} \tau_{\max} = \sup_{\mathbf{b} \in \mathcal{B}} \tau(\mathbf{b}).$$

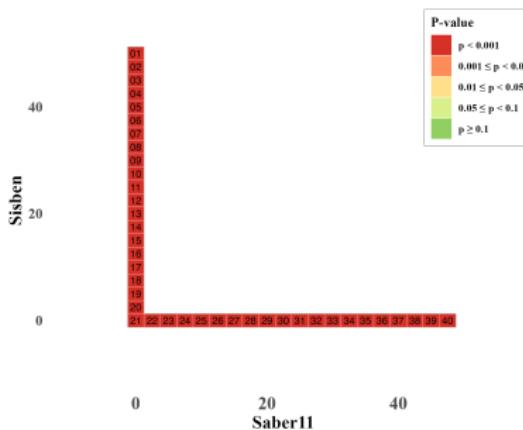
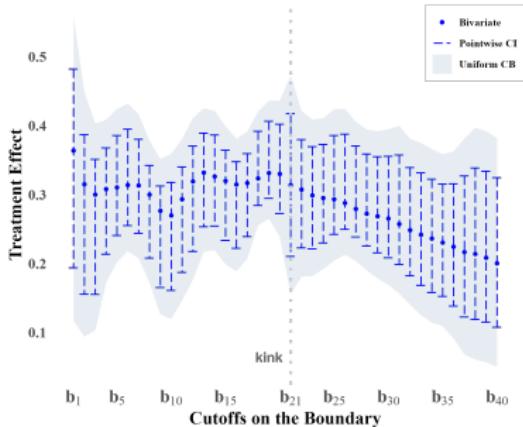
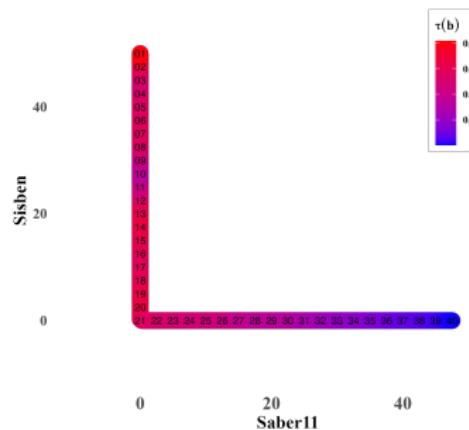
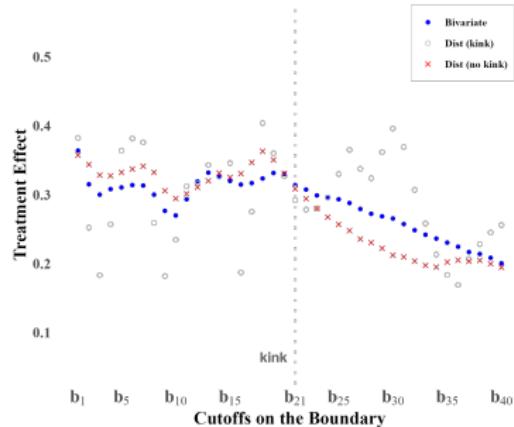
- Inference based on strong approximations and related methods.

Outline

1. Introduction
2. Boundary Average Treatment Effects
3. Distance-Based Methods
4. Location-Based Methods
5. Empirical Application
6. Conclusion



- ▶ Ser Pilo Paga (SPP) Colombian policy program; students $i = 1, 2, \dots, n$.
- ▶ $\mathbf{X}_i = (\text{SABER11}_i, \text{SISBEN}_i)^\top$; SABER11_i = exam score and SABER11_i = wealth index.
- ▶ $\mathcal{B} = \{\text{SABER11} \geq 0 \text{ and } \text{SISBEN} = 0\} \cup \{\text{SABER11} = 0 \text{ and } \text{SISBEN} \geq 0\}$.
- ▶ $(Y_i(0), Y_i(1), \mathbf{X}_i)$, $i = 1, 2, \dots, n$, random sample.
- ▶ $Y_i = \mathbb{1}(\mathbf{X}_i \in \mathcal{A}_0) \cdot Y_i(0) + \mathbb{1}(\mathbf{X}_i \in \mathcal{A}_1) \cdot Y_i(1)$; \mathcal{A}_t group t 's assignment area.



Outline

1. Introduction
2. Boundary Average Treatment Effects
3. Distance-Based Methods
4. Location-Based Methods
5. Empirical Application
6. Conclusion

Conclusion

- ▶ Multi-dimensional RD designs are widely used across disciplines.
- ▶ Methodological and formal results lagging behind its popularity in practice.
- ▶ Through treatment of Boundary Discontinuity Designs:
 - ▶ Boundary average treatment effects using pooled data near \mathcal{B} .
 - ▶ Distance-based methods to each point on \mathcal{B} .
 - ▶ Location-based methods to each point on \mathcal{B} .
 - ▶ Pointwise and uniform estimation and inference methods.
 - ▶ Aggregation of heterogeneous treatment effects along boundary.
- ▶ rd2d package for R.

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