

# 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:

- “Estimation and Inference in Boundary Discontinuity Designs”, arXiv:2211.10805.
- “Causal Treatment Effect Aggregation in Boundary Discontinuity Designs”, arXiv:..
- “rd2d: Causal Inference in Boundary Discontinuity Designs”, arXiv:2505.07989.

# Outline

## 1. Introduction

## 2. Theoretical Results

### 2.1. Distance-Based Methods

### 2.2. Location-Based Methods

## 3. Empirical Application

## 4. Aggregation Along the Boundary

## 5. Conclusion

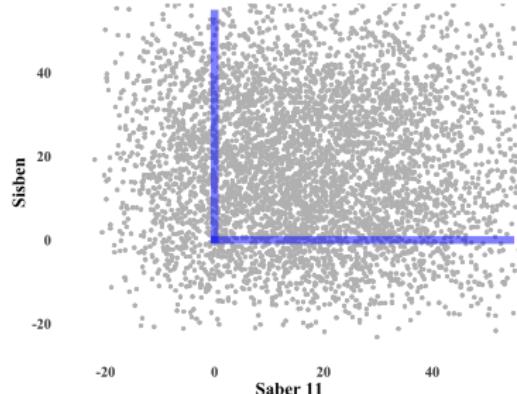
## Introduction

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

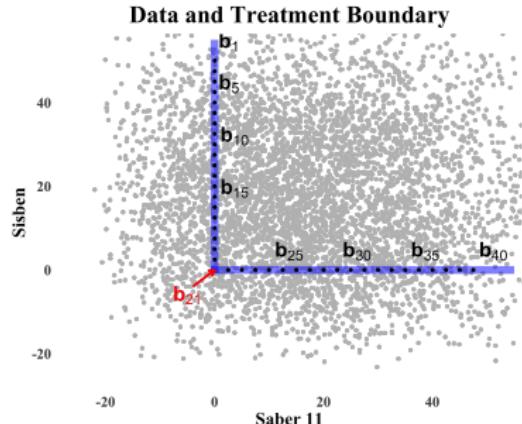
- ▶ Multi-dimensional Regression Discontinuity (RD) designs.
  - ▶ Multi-score RD designs / Geographic RD designs.
- ▶ Two modern approaches for analysis in practice:
  - ▶ Local regression based on univariate distance to boundary.
  - ▶ Local regression based on bivariate location relative to boundary.
- ▶ Aggregation along the boundary: pooled local regression analysis.
- ▶ 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/>

## Data and Treatment Boundary



- ▶ Ser Pilo Paga (SPP) Colombian policy program; students  $i = 1, 2, \dots, n$ .
- ▶  $\mathbf{X}_i = (\text{SABER11}_i, \text{SISBEN}_i)^\top$ ;  $\text{SABER11}_i$  = exam score and  $\text{SISBEN}_i$  = wealth index.
- ▶  $\mathcal{B} = \{\text{SABER11} \geq 0 \text{ and } \text{SISBEN} = 0\} \cup \{\text{SABER11} = 0 \text{ and } \text{SISBEN} \geq 0\}$ .
- ▶  $(Y_i(0), Y_i(1), \mathbf{X}_i)$ ,  $i = 1, 2, \dots, n$ , 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.



- ▶ Causal treatment effect along the assignment boundary:

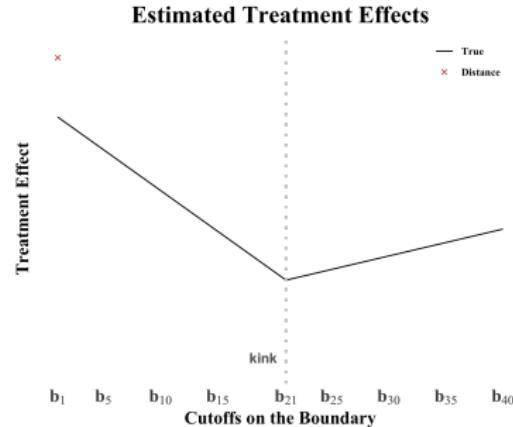
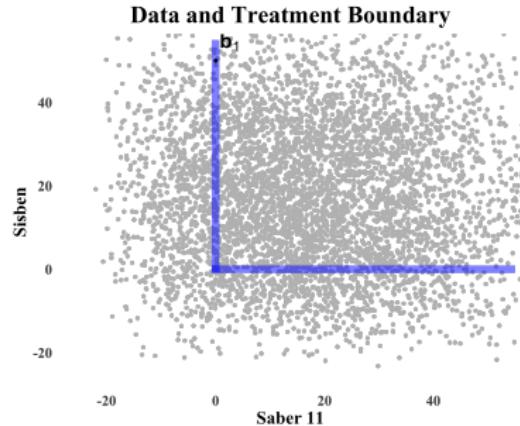
$$\tau(\mathbf{x}) = \mathbb{E}[Y_i(1) - Y_i(0) | \mathbf{X}_i = \mathbf{x}], \quad \mathbf{x} \in \mathcal{B}.$$

- ▶ Estimation and Inference Approaches:

- ▶ Local regression based on univariate distance to boundary:

$$D_i(\mathbf{x}) = d(\mathbf{X}_i, \mathbf{x})(\mathbf{1}(\mathbf{X}_i \in \mathcal{A}_1) - \mathbf{1}(\mathbf{X}_i \in \mathcal{A}_0)), \quad \mathbf{x} \in \mathcal{B}.$$

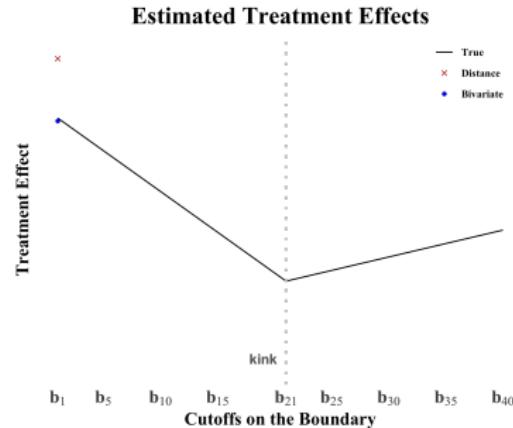
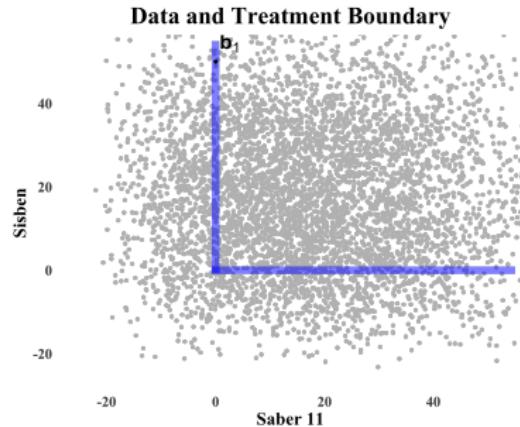
- ▶ Local regression based on bivariate location relative to boundary.



- Distance-based Estimator:  $\hat{\gamma}_{\text{dis}}(\mathbf{b}_1) = \mathbf{e}_1^\top \hat{\gamma}_1(\mathbf{b}_1) - \mathbf{e}_1^\top \hat{\gamma}_0(\mathbf{b}_1)$ , where

$$\hat{\gamma}_t(\mathbf{x}) = \arg \min_{\gamma} \sum_{i=1}^n (Y_i - \mathbf{r}_p(D_i(\mathbf{x}))^\top \gamma)^2 k\left(\frac{D_i(\mathbf{x})}{h}\right) \mathbf{1}(D_i(\mathbf{x}) \in \mathcal{I}_t).$$

- $\mathbf{r}_p(u) = (1, u, u^2, \dots, u^p)^\top$ .
- $k(\cdot)$  univariate kernel, and  $h$  bandwidth.
- $\mathcal{I}_0 = (-\infty, 0)$  and  $\mathcal{I}_1 = [0, \infty)$ .
- $D_i(\mathbf{x}) = d(\mathbf{X}_i, \mathbf{x})(\mathbf{1}(\mathbf{X}_i \in \mathcal{A}_1) - \mathbf{1}(\mathbf{X}_i \in \mathcal{A}_0))$ .
- $\mathbf{x} \in \mathcal{B}$  and  $t \in \{0, 1\}$ .

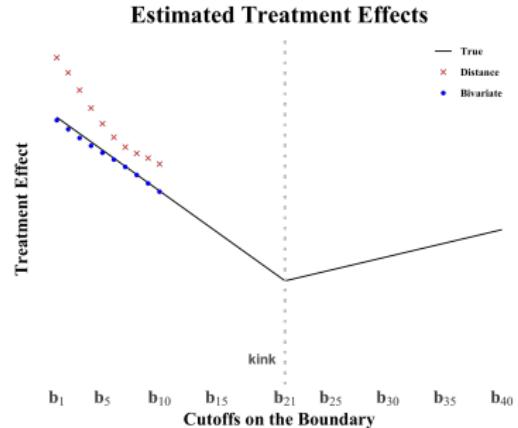
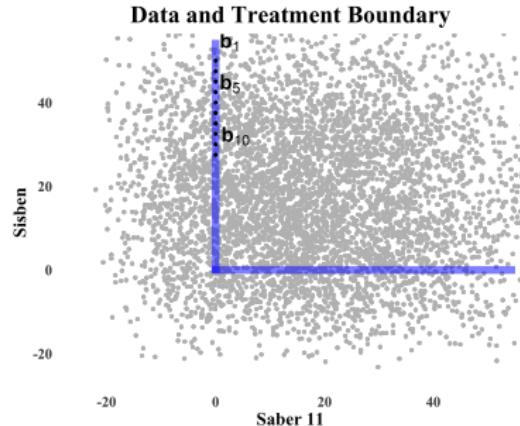


- ▶ Location-based Estimator:  $\hat{\tau}(b_1) = \mathbf{e}_1^\top \hat{\beta}_1(b_1) - \mathbf{e}_1^\top \hat{\beta}_0(b_1)$  for  $\mathbf{x} \in \mathcal{B}$ ,

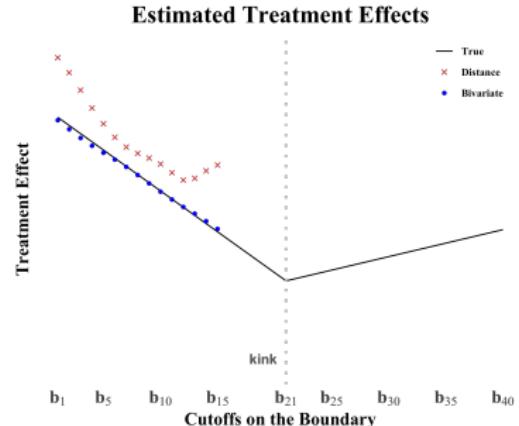
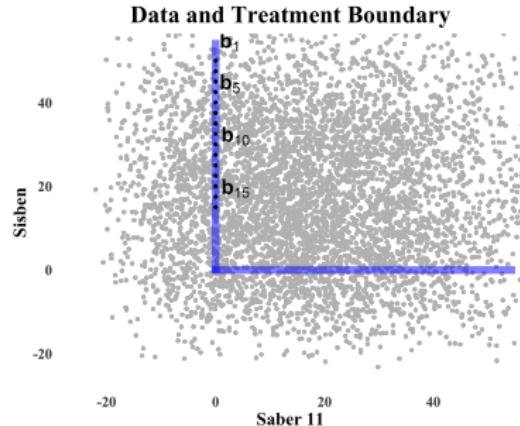
$$\hat{\beta}_t(\mathbf{x}) = \arg \min_{\boldsymbol{\beta}} \sum_{i=1}^n (Y_i - \mathbf{R}_p(\mathbf{X}_i - \mathbf{x})^\top \boldsymbol{\beta})^2 K\left(\frac{\mathbf{X}_i - \mathbf{x}}{h}\right) \mathbf{1}(\mathbf{X}_i \in \mathcal{A}_t).$$

$$\mathbf{R}_p(\mathbf{u}) = (1, u_1, u_2, u_1^2, u_2^2, u_1 u_2, \dots, u_1^p, u_2^p)^\top.$$

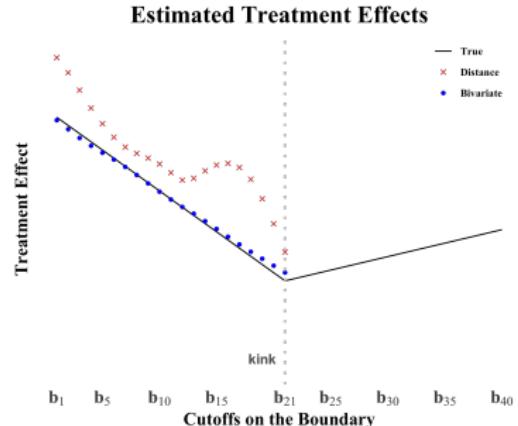
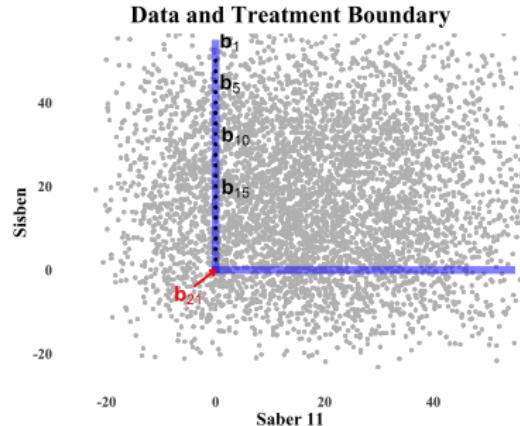
- ▶  $K(\cdot)$  bivariate kernel, and  $h$  bandwidth.
- ▶  $\mathcal{A}_0$  = control region and  $\mathcal{A}_1$  = treatment region.
- ▶  $\mathbf{X}_i$  bivariate score.
- ▶  $\mathbf{x} \in \mathcal{B}$  and  $t \in \{0, 1\}$ .



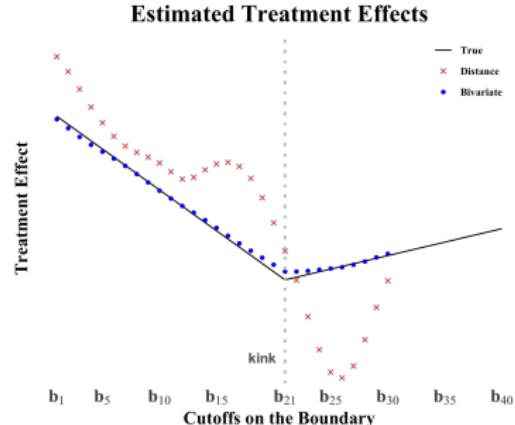
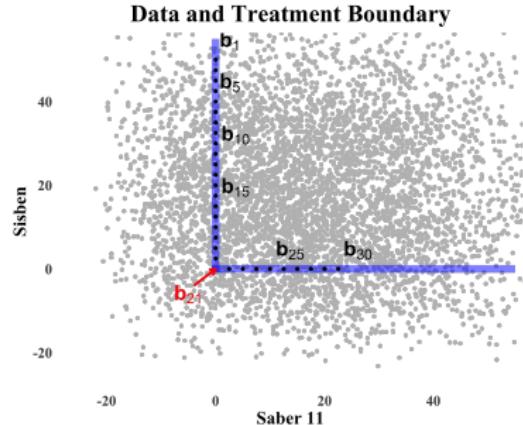
- ▶ Distance-based Estimator:  $\hat{\tau}_{\text{dis}}(\mathbf{x})$ .
- ▶ Location-based Estimator:  $\hat{\tau}(\mathbf{x})$ .
- ▶ Evaluation points along  $\mathcal{B}$ :  $\mathbf{x} \in \{\mathbf{b}_1, \dots, \mathbf{b}_{10}\}$ .



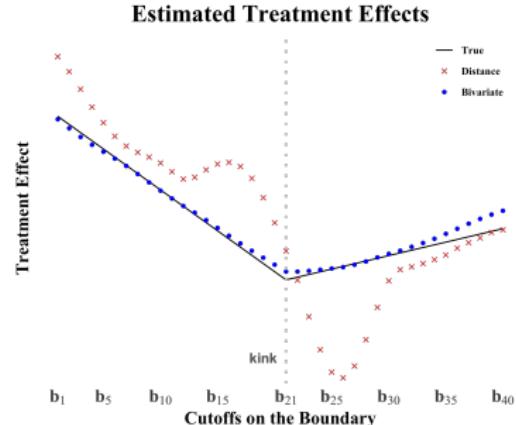
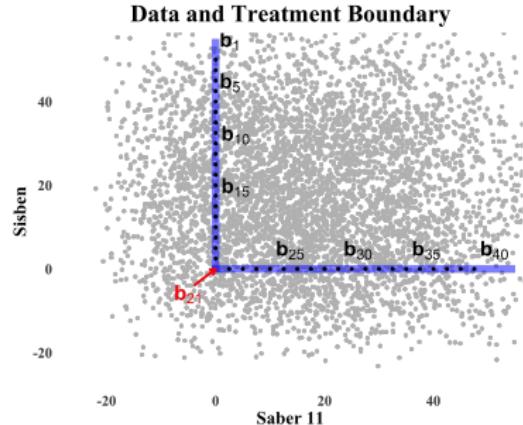
- ▶ Distance-based Estimator:  $\hat{\tau}_{\text{dis}}(\mathbf{x})$ .
- ▶ Location-based Estimator:  $\hat{\tau}(\mathbf{x})$ .
- ▶ Evaluation points along  $\mathcal{B}$ :  $\mathbf{x} \in \{\mathbf{b}_1, \dots, \mathbf{b}_{15}\}$ .



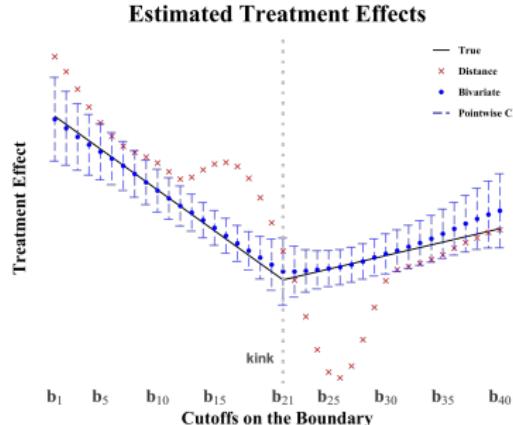
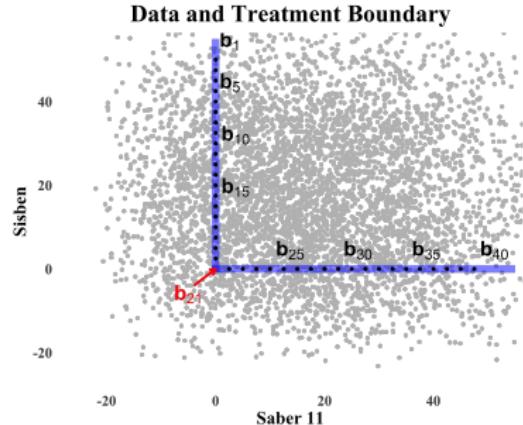
- ▶ Distance-based Estimator:  $\hat{\tau}_{\text{dis}}(\mathbf{x})$ .
- ▶ Location-based Estimator:  $\hat{\tau}(\mathbf{x})$ .
- ▶ Evaluation points along  $\mathcal{B}$ :  $\mathbf{x} \in \{\mathbf{b}_1, \dots, \mathbf{b}_{21}\}$ .



- ▶ Distance-based Estimator:  $\hat{\tau}_{\text{dis}}(\mathbf{x})$ .
- ▶ Location-based Estimator:  $\hat{\tau}(\mathbf{x})$ .
- ▶ Evaluation points along  $\mathcal{B}$ :  $\mathbf{x} \in \{\mathbf{b}_1, \dots, \mathbf{b}_{30}\}$ .



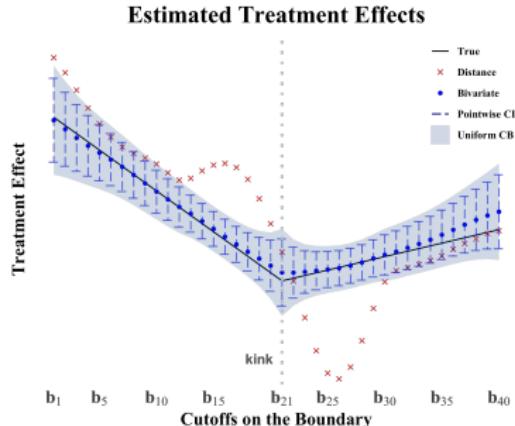
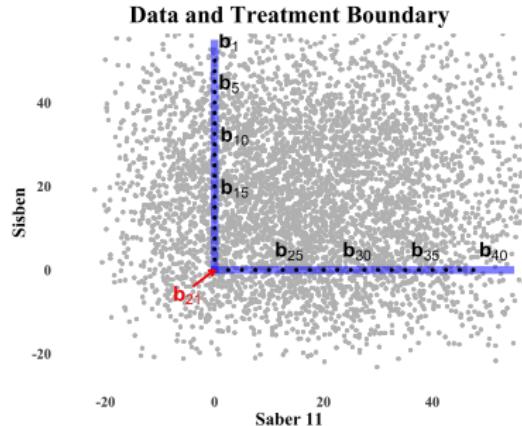
- ▶ Distance-based Estimator:  $\hat{\tau}_{\text{dis}}(\mathbf{x})$ .
- ▶ Location-based Estimator:  $\hat{\tau}(\mathbf{x})$ .
- ▶ Evaluation points along  $\mathcal{B}$ :  $\mathbf{x} \in \{\mathbf{b}_1, \dots, \mathbf{b}_{40}\}$ .



- ▶ 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  $\mathbf{x} \in \{b_1, \dots, b_{40}\}$ ,

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

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



- ▶ 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  $\mathbf{x} \in \mathcal{B}$ ,

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

- ▶  $q_\alpha = \inf\{c > 0 : \mathbb{P}[\sup_{\mathbf{x} \in \mathcal{B}} |\hat{Z}_n(\mathbf{x})| \geq c | \text{data}] \leq \alpha\}$ .
- ▶  $(\hat{Z}_n : \mathbf{x} \in \mathcal{B})$  is a Gaussian process conditional on data, with  $\mathbb{E}[\hat{Z}_n(\mathbf{x}_1) | \text{data}] = 0$  and an estimated covariance function  $\mathbb{E}[\hat{Z}_n(\mathbf{x}_1)\hat{Z}_n(\mathbf{x}_2) | \text{data}]$  for all  $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{B}$ .

# Outline

1. Introduction

2. Theoretical Results

    2.1. Distance-Based Methods

    2.2. Location-Based Methods

3. Empirical Application

4. Aggregation Along the Boundary

5. Conclusion

## Contributions

- ▶ Analysis based on univariate distance to boundary:  $\hat{\tau}_{\text{dis}}(\mathbf{x})$ .
  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.
- ▶ Analysis based on bivariate location relative to boundary:  $\hat{\tau}(\mathbf{x})$ .
  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.
- ▶ New strong approximation result for empirical processes.

# Outline

1. Introduction

2. Theoretical Results

    2.1. Distance-Based Methods

    2.2. Location-Based Methods

3. Empirical Application

4. Aggregation Along the Boundary

5. 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.**  $\hat{\tau}_{\text{dis}}(\mathbf{x}) = \mathbf{e}_1^\top \hat{\gamma}_1(\mathbf{x}) - \mathbf{e}_1^\top \hat{\gamma}_0(\mathbf{x})$  for  $\mathbf{x} \in \mathcal{B}$ , where

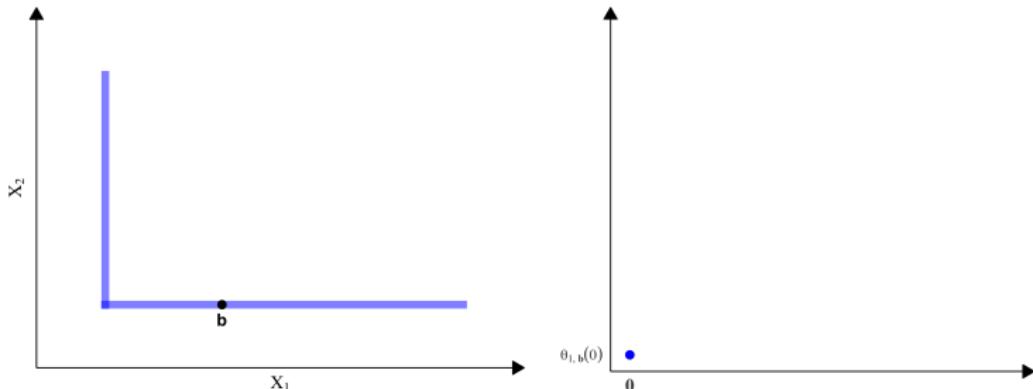
$$\hat{\gamma}_t(\mathbf{x}) = \arg \min_{\boldsymbol{\gamma}} \sum_{i=1}^n (Y_i - \mathbf{r}_p(D_i(\mathbf{x}))^\top \boldsymbol{\gamma})^2 k\left(\frac{D_i(\mathbf{x})}{h}\right) \mathbf{1}(D_i(\mathbf{x}) \in \mathcal{J}_t).$$

- ▶ **Assumption.** 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  $k(u) = \mathbf{1}(u \in [-1, 1])$ .
  - ▶  $\liminf_{h \downarrow 0} \inf_{\mathbf{x} \in \mathcal{B}} \int_{\mathcal{A}_t} k_h(d(\mathbf{u}, \mathbf{x})) d\mathbf{u} \gtrsim 1$ .
- ▶ **Identification.** For all  $\mathbf{x} \in \mathcal{B}$ ,

$$\tau(\mathbf{x}) = \lim_{r \downarrow 0} \theta_{1,\mathbf{x}}(r) - \lim_{r \uparrow 0} \theta_{0,\mathbf{x}}(r)$$

with

$$\theta_{t,\mathbf{x}}(r) = \mathbb{E}[Y_i | D_i(\mathbf{x}) = r, D_i(\mathbf{x}) \in \mathcal{J}_t].$$

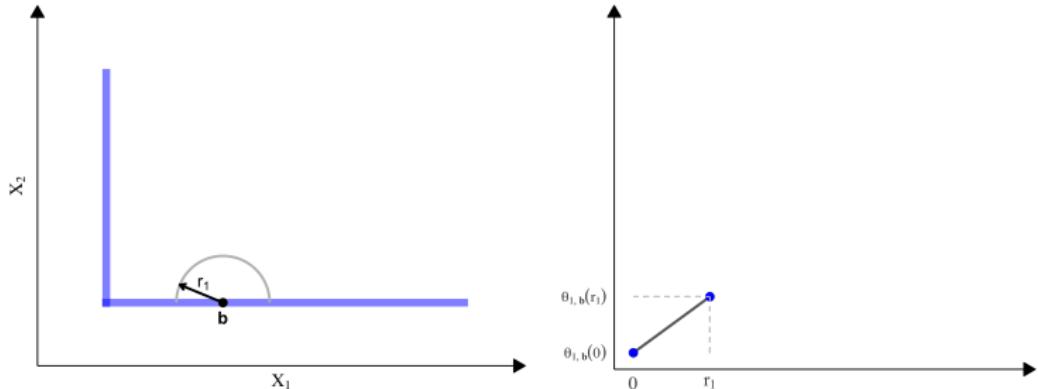


- ▶ **Treatment Group.** Bivariate vs. univariate (distance-induced) expectations:

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

where

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



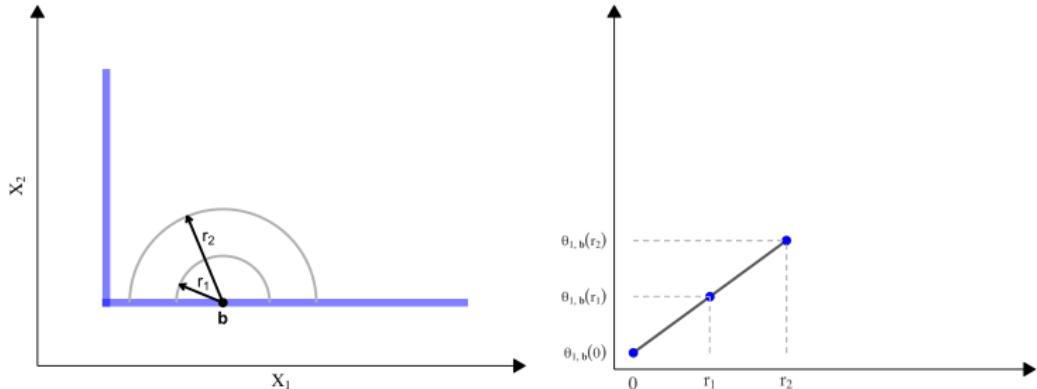
- **Distance range:**  $[0, r_1]$ . If  $\mathbf{x} \mapsto \mathbb{E}[Y(1)|\mathbf{X}_i = \mathbf{b}, \mathbf{X}_i \in \mathcal{A}_1]$  is **smooth**, then

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

is also **smooth**.

- Thus, distance-based local polynomial estimator misspecification bias is

$$\mathbb{E}[\mathbf{e}_1^\top \hat{\gamma}_1(\mathbf{x})|\mathbf{D}(\mathbf{b})] - \theta_{t,\mathbf{b}}(r) \lesssim_{\mathbb{P}} h^{p+1}.$$



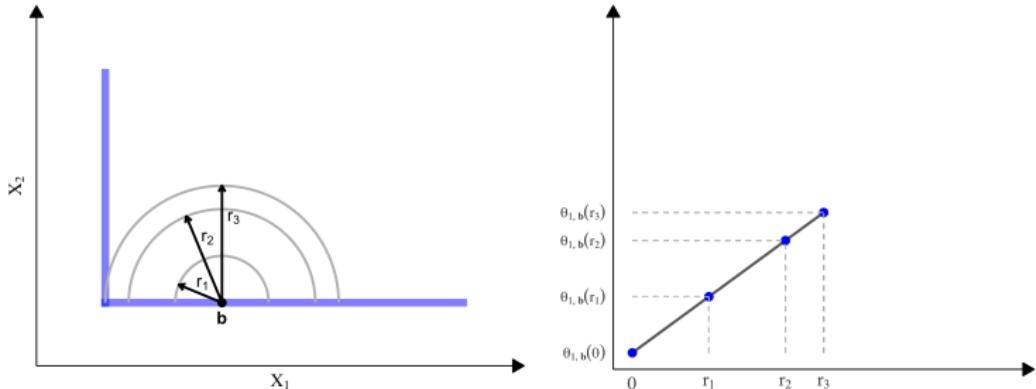
- **Distance range:**  $[0, r_2]$ . If  $\mathbf{x} \mapsto \mathbb{E}[Y(1)|\mathbf{X}_i = \mathbf{b}, \mathbf{X}_i \in \mathcal{A}_1]$  is **smooth**, then

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

is also **smooth**.

- Thus, distance-based local polynomial estimator misspecification bias is

$$\mathbb{E}[\mathbf{e}_1^\top \hat{\gamma}_1(\mathbf{x})|\mathbf{D}(\mathbf{b})] - \theta_{t,\mathbf{b}}(r) \lesssim_{\mathbb{P}} h^{p+1}.$$

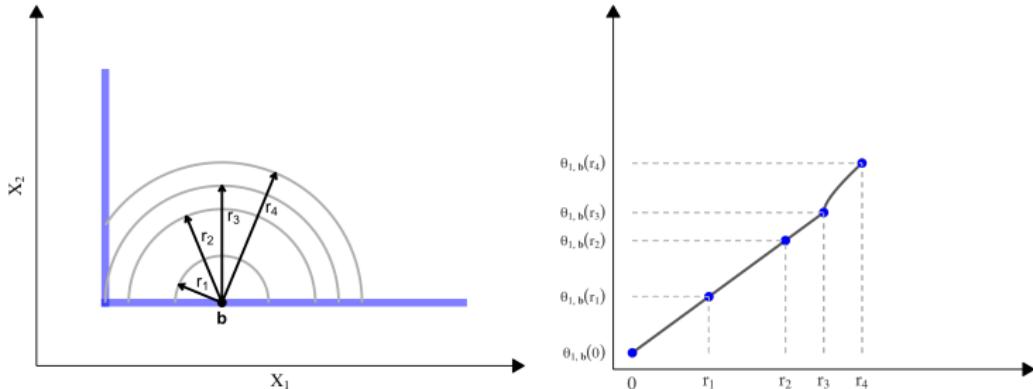


- **Distance range:**  $[0, r_3]$ . If  $\mathbf{x} \mapsto \mathbb{E}[Y(1)|\mathbf{X}_i = \mathbf{b}, \mathbf{X}_i \in \mathcal{A}_1]$  is **smooth**, then

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

is also **smooth**.

- **Smoothness.**  $r \mapsto \theta_{t,\mathbf{b}}(r)$  is **locally to zero**  $(p+1)$ th smooth.

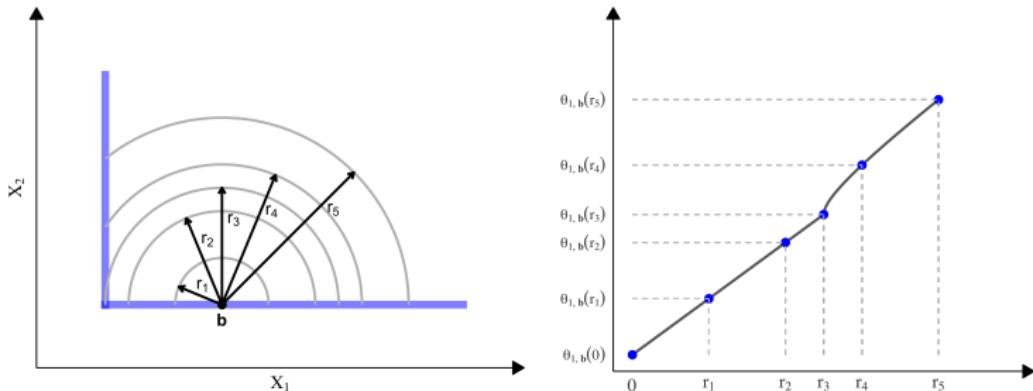


- **Distance range:**  $[0, r_4]$ . If  $\mathbf{x} \mapsto \mathbb{E}[Y(1)|\mathbf{X}_i = \mathbf{b}, \mathbf{X}_i \in \mathcal{A}_1]$  is **smooth**, then

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

is not **smooth**.

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



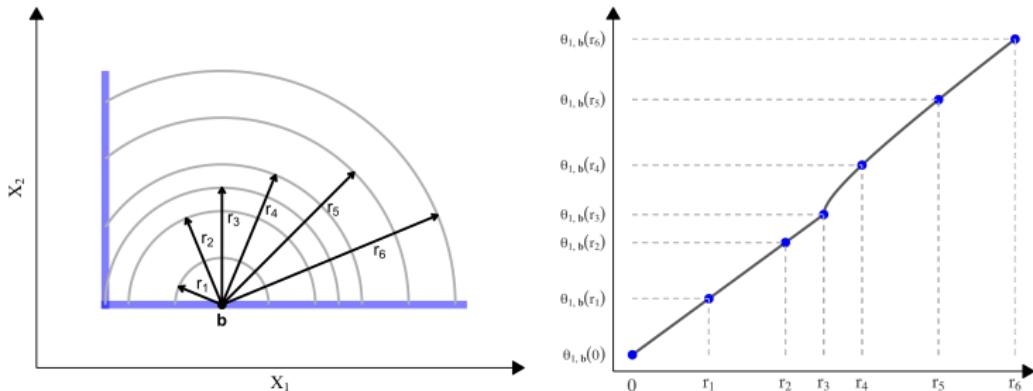
- **Distance range:**  $[0, r_5]$ . Distance-based local polynomial estimator misspecification bias is

$$\mathbb{E}[\mathbf{e}_1^\top \hat{\gamma}_1(\mathbf{x}) | \mathbf{D}(\mathbf{b})] - \theta_{t,\mathbf{b}}(r) \lesssim_{\mathbb{P}} h$$

regardless  $p$  used, that is, not of order  $h^{p+1}$  as expected given underlying 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!
- The closer to a kink point on  $\mathcal{B}$ , the smaller the bandwidth  $h$  must be.

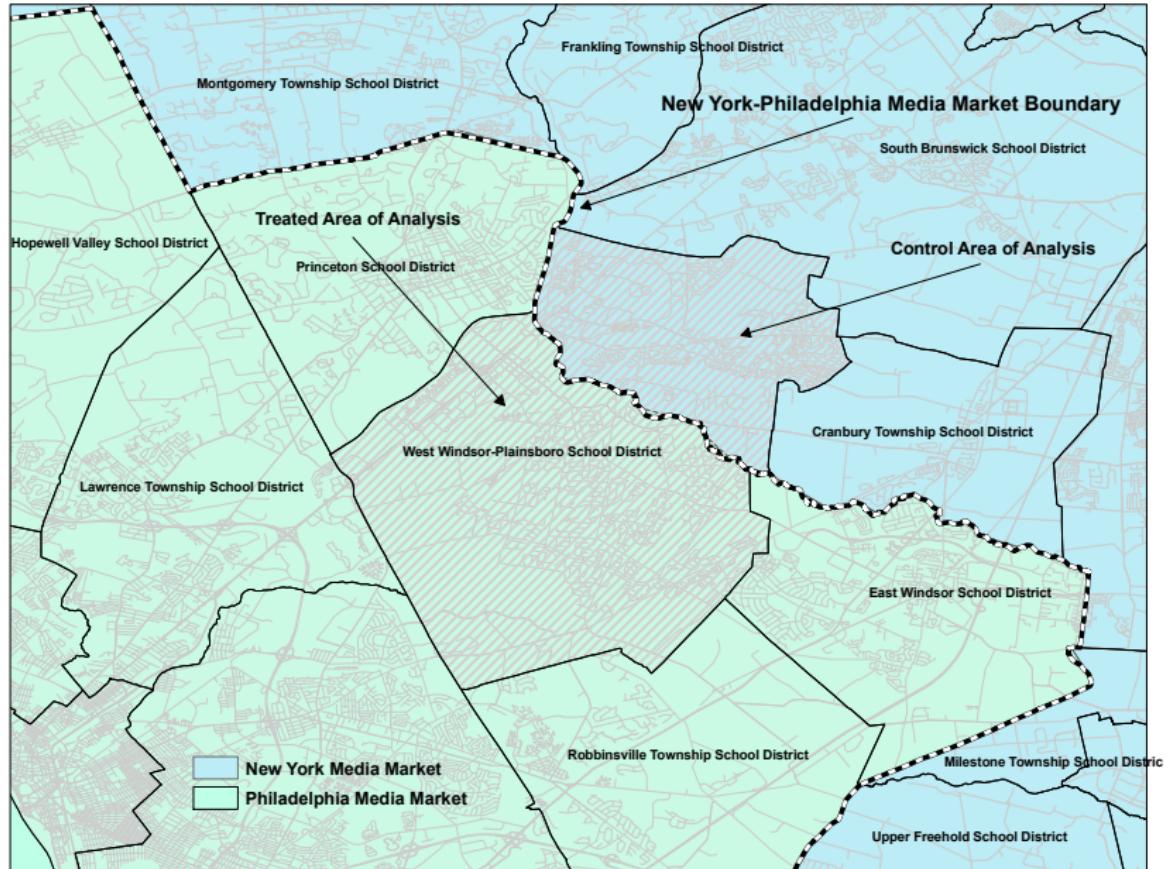


► **Uniform Analysis.** Under minimal regularity conditions, and for any  $p \geq 1$ ,

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

where  $\mathfrak{B}_n(\mathbf{x})$  denotes the bias of the distance-based estimator.

- 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  $(\mathbf{W}_n = (Y_i, \mathbf{X}_i^\top)^\top : 1 \leq i \leq n)$ .

## Other Results for Distance-Based Methods

- ▶ **Regularity Condition.**  $\sup_{\mathbf{x} \in \mathcal{X}} \mathbb{E}[|Y_i(t)|^{2+v} | \mathbf{X}_i = \mathbf{x}] < \infty$  for some  $v \geq 2$ .
- ▶ **Convergence Rates.** Under minimal regularity conditions,

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

and

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

- ▶ **Pointwise Inference.** Ignoring the potential bias problem when  $\mathcal{B}$  is non-smooth, paper establishes distribution theory with valid standard errors for each  $\mathbf{x} \in \mathcal{B}$ . This result is fairly standard, up to handling  $\mathcal{B}$ .
- ▶ **Uniform Inference.** Ignoring the potential bias problem when  $\mathcal{B}$  is non-smooth, paper establishes feasible uniform distribution theory via simulation. This result requires new technical tools, and requires careful handling of  $\mathcal{B}$ . More details later.
- ▶ **Practice.** Valid and invalid practices based on standard univariate RD designs methods.

# Outline

1. Introduction

2. Theoretical Results

    2.1. Distance-Based Methods

    2.2. Location-Based Methods

3. Empirical Application

4. Aggregation Along the Boundary

5. 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.**  $\hat{\tau}(\mathbf{b}_1) = \mathbf{e}_1^\top \hat{\beta}_1(\mathbf{b}_1) - \mathbf{e}_1^\top \hat{\beta}_0(\mathbf{b}_1)$  for  $\mathbf{x} \in \mathcal{B}$ ,

$$\hat{\beta}_t(\mathbf{x}) = \arg \min_{\boldsymbol{\beta}} \sum_{i=1}^n (Y_i - \mathbf{R}_p(\mathbf{X}_i - \mathbf{x})^\top \boldsymbol{\beta})^2 K\left(\frac{\mathbf{X}_i - \mathbf{x}}{h}\right) \mathbf{1}(\mathbf{X}_i \in \mathcal{A}_t).$$

► **Assumption.** Let  $t \in \{0, 1\}$ .

- $K : \mathbb{R}^2 \rightarrow [0, \infty)$  compact supported & Lipschitz continuous, or  $K(\mathbf{u}) = \mathbf{1}(\mathbf{u} \in [-1, 1]^2)$ .
- $\liminf_{h \downarrow 0} \inf_{\mathbf{x} \in \mathcal{B}} \int_{\mathcal{A}_t} K_h(\mathbf{u} - \mathbf{x}) d\mathbf{u} \gtrsim 1$ .

► **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.

## Point Estimation Results for Location-Based Methods

- ▶ **Regularity Condition.**  $\sup_{\mathbf{x} \in \mathcal{X}} \mathbb{E}[|Y_i(t)|^{2+v} | \mathbf{X}_i = \mathbf{x}] < \infty$  for some  $v \geq 2$ .
- ▶ **Convergence Rates.** Under minimal regularity conditions,

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

and

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

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

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

and

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

- ▶ Standard bandwidth selection methods developed in the paper.

## Inference Results for Location-Based Methods

- ▶ **Feasible t-test.** Using standard least squares algebra,  $\widehat{T}(\mathbf{x}) = \frac{\widehat{\tau}(\mathbf{x}) - \tau(\mathbf{x})}{\sqrt{\widehat{\Omega}_{\mathbf{x}, \mathbf{x}}}}$ .
- ▶ **Uncertainty Quantification.** Confidence intervals and confidence bands,
$$\widehat{I}(\mathbf{x}; \alpha) = \left[ \widehat{\tau}(\mathbf{x}) - \varphi_\alpha \sqrt{\widehat{\Omega}_{\mathbf{x}}} , \widehat{\tau}(\mathbf{x}) + \varphi_\alpha \sqrt{\widehat{\Omega}_{\mathbf{x}}} \right], \quad \mathbf{x} \in \mathcal{B},$$
- ▶ **Pointwise Inference.** By standard CLT result, for each  $\mathbf{x} \in \mathcal{B}$ , set  $\varphi_\alpha = \Phi^{-1}(1 - \alpha/2)$ .
- ▶ **Uniform Inference.** Note that
$$\mathbb{P}[\tau(\mathbf{x}) \in \widehat{I}(\mathbf{x}; \alpha), \text{ for all } \mathbf{x} \in \mathcal{B}] = \mathbb{P}\left[\sup_{\mathbf{x} \in \mathcal{B}} |\widehat{T}(\mathbf{x})| \leq \varphi_\alpha\right].$$
  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.
  2. Deduce the distribution of  $\sup_{\mathbf{x} \in \mathcal{B}} |\widehat{T}(\mathbf{x})|$ .
  3. Using simulations, set  $\varphi_\alpha = \inf\{c > 0 : \mathbb{P}[\sup_{\mathbf{x} \in \mathcal{B}} |\widehat{Z}_n(\mathbf{x})| \geq c | \text{data}] \leq \alpha\}$ .
- ▶ **Implementation and Bias.** (I)MSE-optimal bandwidth selection for point estimation, robust bias correction for inference.

# Outline

1. Introduction

2. Theoretical Results

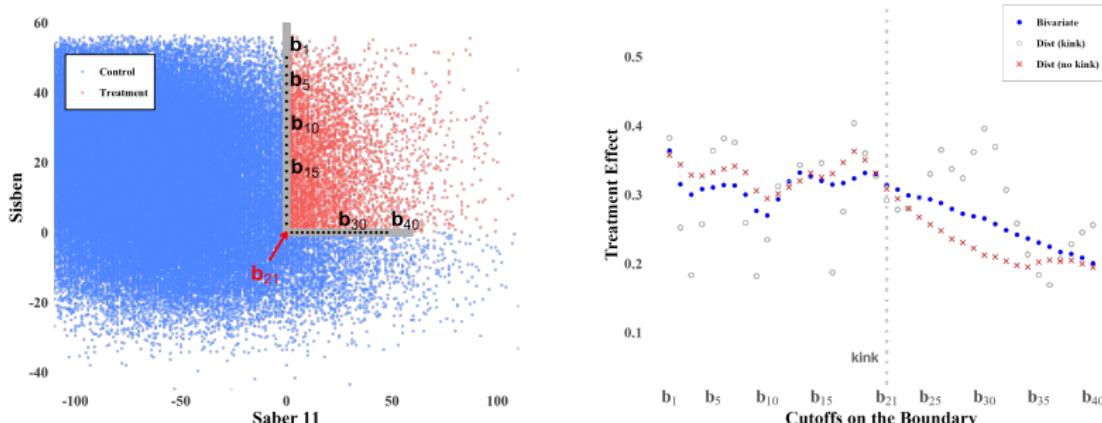
    2.1. Distance-Based Methods

    2.2. Location-Based Methods

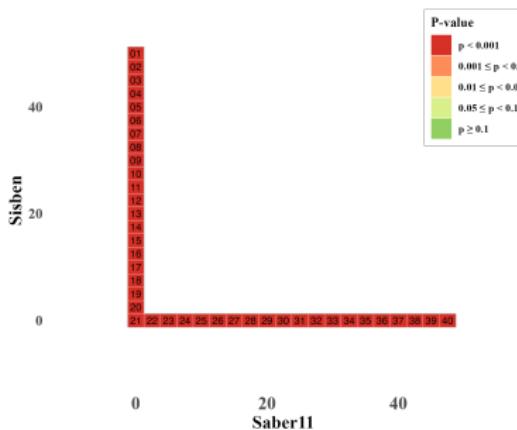
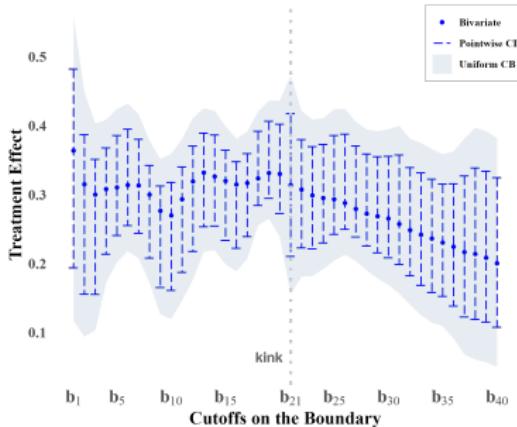
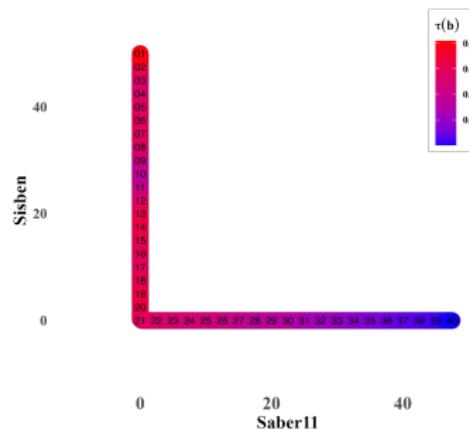
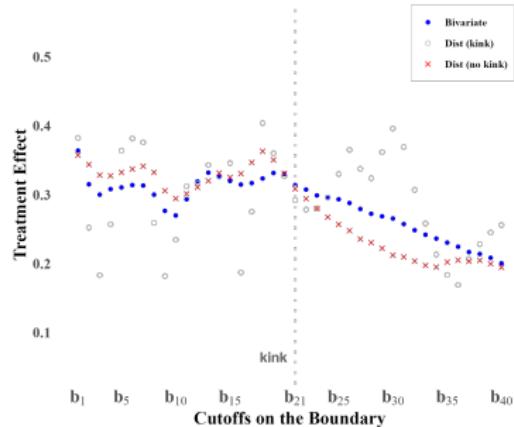
3. Empirical Application

4. Aggregation Along the Boundary

5. Conclusion



- ▶ Ser Pilo Paga (SPP) Colombian policy program; students  $i = 1, 2, \dots, n$ .
- ▶  $\mathbf{X}_i = (\text{SABER11}_i, \text{SISBEN}_i)^\top$ ;  $\text{SABER11}_i$  = exam score and  $\text{SABER11}_i$  = 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. Theoretical Results

    2.1. Distance-Based Methods

    2.2. Location-Based Methods

3. Empirical Application

4. Aggregation Along the Boundary

5. Conclusion

## Inference Results for Location-Based Methods

- The *aggregated average treatment effect* (AATE) along the boundary  $\mathcal{B}$  is

$$\tau_{\text{AATE}, \mathcal{B}} = \frac{\int_{\mathcal{B}} \tau(\mathbf{b}) w(\mathbf{b}) d\mathbf{b}}{\int_{\mathcal{B}} w(\mathbf{b}) d\mathbf{b}}$$

and a generic plug-in “estimator” thereof is

$$\hat{\tau}_{\text{AATE}, \mathbf{b}} = \frac{\sum_{j=1}^J \hat{\tau}(\mathbf{b}_j) w(\mathbf{b}_j)}{\sum_{j=1}^J w(\mathbf{b}_j)},$$

for choice of boundary cutoff points  $\mathbf{b} = (\mathbf{b}_1, \dots, \mathbf{b}_J)^\top$  over  $\mathcal{B}$ .

- **Distribution theory.** Under regularity conditions,

$$\frac{\hat{\tau}_{\text{AATE}, \mathbf{b}} - \tau_{\text{AATE}, \mathcal{B}}}{\sqrt{\mathbf{w}^\top \widehat{\mathbf{V}}_{\mathbf{b}} \mathbf{w} / (nh^2)}} \rightsquigarrow \mathcal{N}(0, 1)$$

where  $\mathbf{w} = (w(\mathbf{b}_1), \dots, w(\mathbf{b}_J))^\top$ .

- **Comments.**

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

# Outline

1. Introduction

2. Theoretical Results

    2.1. Distance-Based Methods

    2.2. Location-Based Methods

3. Empirical Application

4. Aggregation Along the Boundary

5. Conclusion

## Conclusion

- ▶ Multi-dimensional RD designs are widely used across disciplines.
- ▶ Methodological and formal results lagging behind its popularity in practice.
- ▶ We offer a through treatment of Boundary Discontinuity Designs.
  - ▶ Distance-based methods may exhibit large bias when  $\mathcal{B}$  is non-smooth.
  - ▶ Location-based methods do not suffer of this drawback.
  - ▶ We develop pointwise and uniform estimation and inference methods.
- ▶ rd2d package for R.

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