Estimation and Inference in Boundary Discontinuity Designs

Matias D. Cattaneo
Princeton University

Rocio Titiunik
Princeton University

Ruiqi (Rae) Yu
Princeton University

May 2025

Outline

Introduction

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

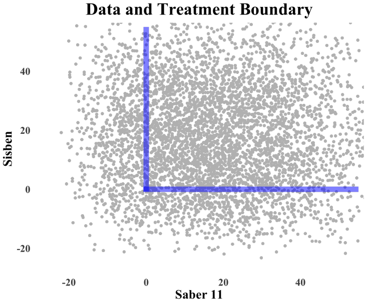
- Multi-dimensional Regression Discontinuity (RD) designs.
  - Multi-score RD designs / Geographic RD designs.

- Two modern approaches for analysis in practice:
  - Local regression based on univariate distance to boundary.
  - Local regression based on bivariate location relative to boundary.

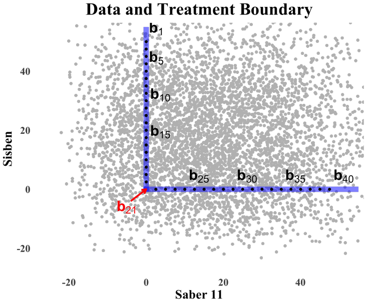
- Aggregation along the boundary: pooled local regression analysis.

- 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/



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



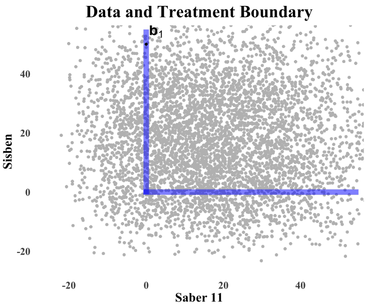
- Causal treatment effect along the assignment boundary:

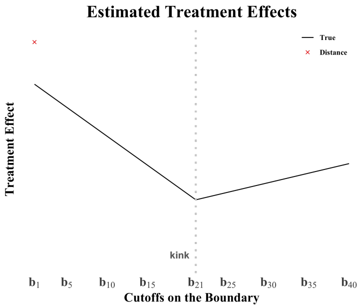
$$\tau(\mathbf{x}) = \mathbb{E}[Y_i(1) - Y_i(0)|\mathbf{X}_i = \mathbf{x}], \qquad \mathbf{x} \in \mathscr{B}.$$

- Estimation and Inference Approaches:
  - Local regression based on <u>univariate</u> distance to boundary:

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

  - Local regression based on <u>bivariate</u> location relative to boundary.

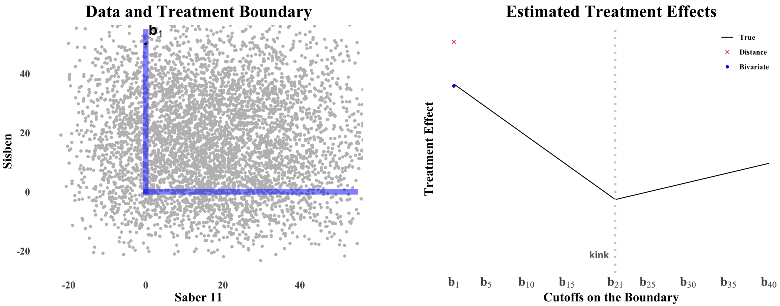




- Distance-based Estimator: $\widehat{\tau}_{\text{dis}}(\mathbf{b}_1) = \mathbf{e}_1^\top \widehat{\boldsymbol{\gamma}}_1(\mathbf{b}_1) - \mathbf{e}_1^\top \widehat{\boldsymbol{\gamma}}_0(\mathbf{b}_1)$, where

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

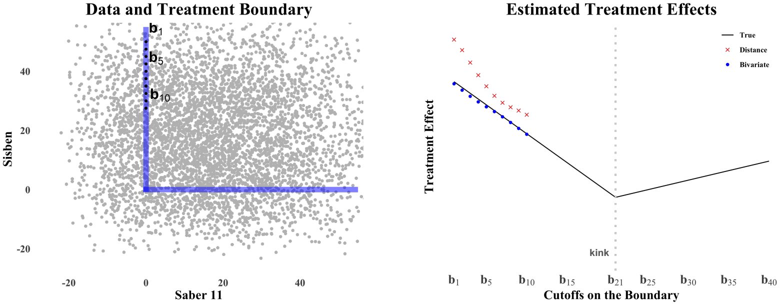
  - $\mathbf{r}_p(u) = (1, u, u^2, \cdots, u^p)^\top$.
  - $k(\cdot)$ univariate kernel, and h bandwidth.
  - $\mathscr{I}_0 = (-\infty, 0)$ and $\mathscr{I}_1 = [0, \infty)$.
  - $D_i(\mathbf{x}) = \mathscr{d}(\mathbf{X}_i, \mathbf{x})(\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_1) - \mathbb{1}(\mathbf{X}_i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



- Location-based Estimator: $\widehat{\tau}(\mathbf{b}_1) = \mathbf{e}_1^\top \widehat{\boldsymbol{\beta}}_1(\mathbf{b}_1) - \mathbf{e}_1^\top \widehat{\boldsymbol{\beta}}_0(\mathbf{b}_1)$ for $\mathbf{x} \in \mathscr{B}$,

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

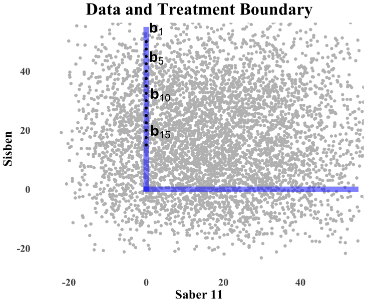
  - $\mathbf{R}_p(\mathbf{u}) = (1, u_1, u_2, u_1^2, u_2^2, u_1 u_2, \cdots, u_1^p, u_2^p)^\top$.
  - $K(\cdot)$ bivariate kernel, and h bandwidth.
  - $\mathscr{A}_0$ = control region and $\mathscr{A}_1$ = treatment region.
  - $\mathbf{X}_i$ bivariate score.
  - $\mathbf{x} \in \mathscr{B}$ and $t \in \{0, 1\}$.

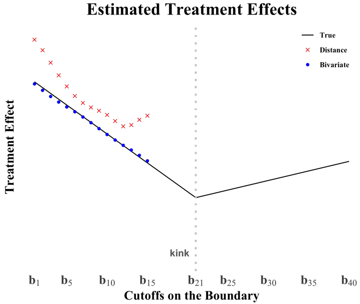


- Distance-based Estimator: $\widehat{\tau}_{\text{dis}}(\mathbf{x})$.

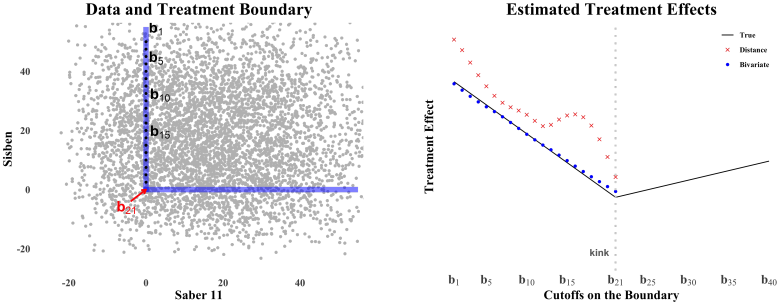
- Location-based Estimator: $\widehat{\tau}(\mathbf{x})$.

- Evaluation points along $\mathscr{B}$: $\mathbf{x} \in \{\mathbf{b}_1, \ldots, \mathbf{b}_{10}\}$.

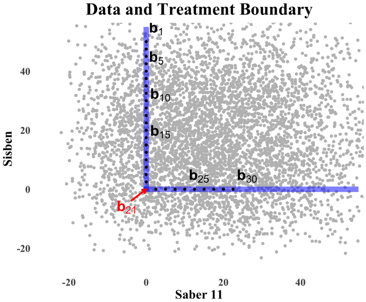


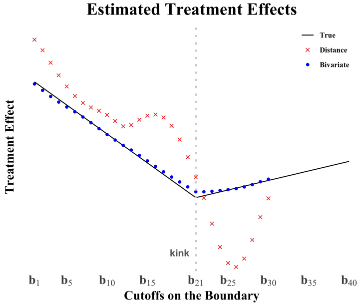


- Distance-based Estimator: $\widehat{\tau}_{\text{dis}}(\mathbf{x})$.
- Location-based Estimator: $\widehat{\tau}(\mathbf{x})$.
- Evaluation points along $\mathscr{B}$: $\mathbf{x} \in \{\mathbf{b}_1, \ldots, \mathbf{b}_{15}\}$.

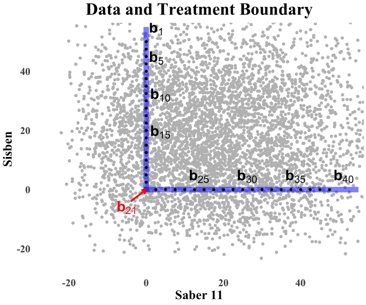


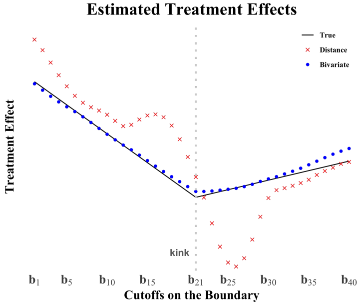
- Distance-based Estimator: $\widehat{\tau}_{\text{dis}}(\mathbf{x})$.
- Location-based Estimator: $\widehat{\tau}(\mathbf{x})$.
- Evaluation points along $\mathscr{B}$: $\mathbf{x} \in \{\mathbf{b}_1, \ldots, \mathbf{b}_{21}\}$.



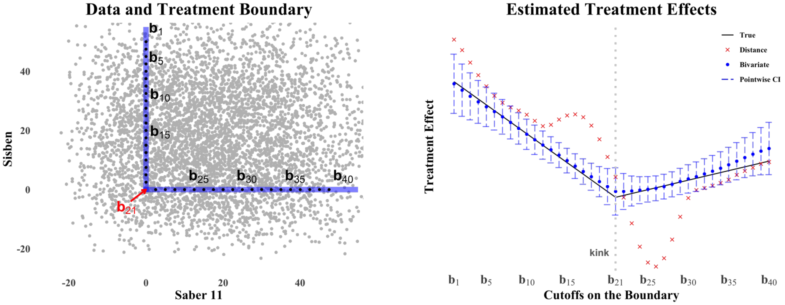


- Distance-based Estimator: $\widehat{\tau}_{\text{dis}}(\mathbf{x})$.
- Location-based Estimator: $\widehat{\tau}(\mathbf{x})$.
- Evaluation points along $\mathscr{B}$: $\mathbf{x} \in \{\mathbf{b}_1, \ldots, \mathbf{b}_{30}\}$.





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

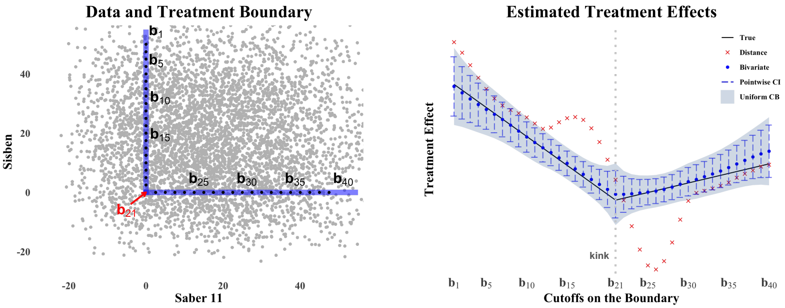


- Estimators: $\widehat{\tau}_{\text{dis}}(\mathbf{x})$ and $\widehat{\tau}(\mathbf{x})$, for each $\mathbf{x} \in \{\mathbf{b}_1, \ldots, \mathbf{b}_{40}\}$.

- Uncertainty Quantification: Confidence Intervals. For each $\mathbf{x} \in \{\mathbf{b}_1, \ldots, \mathbf{b}_{40}\}$,

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

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



- Estimators: $\widehat{\tau}_{\text{dis}}(\mathbf{x})$ and $\widehat{\tau}(\mathbf{x})$, uniformly in $\mathbf{x} \in \mathscr{B}$.

- Uncertainty Quantification: Confidence Bands. Uniformly in $\mathbf{x} \in \mathscr{B}$,

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

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

Outline

Contributions

- Analysis based on <u>univariate</u> distance to boundary: $\widehat{\tau}_{\mathrm{dis}}(\mathbf{x})$.
  1. Sufficient conditions for identification.
  2. "Large" misspecification bias when $\mathscr{B}$ is non-smooth (e.g., near a kink).
  3. "Small" misspecification bias when $\mathscr{B}$ is smooth.
  4. Pointwise and uniform convergence rates and distribution theory.
  5. Discuss connects and differences with standard univariate RD designs.

- Analysis based on <u>bivariate</u> location relative to boundary: $\widehat{\tau}(\mathbf{x})$.
  1. Identification, estimation, and inference (pointwise and uniform over $\mathscr{B}$) are standard.
  2. Additional (mild) regularity on $\mathscr{B}$ is needed.
  3. New methods for analysis of Boundary Discontinuity Designs.

- New strong approximation result for empirical processes.

Outline

Distance-Based Methods: Identification

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

- **Estimator**. $\widehat{\tau}_{\mathrm{dis}}(\mathbf{x}) = \mathbf{e}_1^\top \widehat{\gamma}_1(\mathbf{x}) - \mathbf{e}_1^\top \widehat{\gamma}_0(\mathbf{x})$ for $\mathbf{x} \in \mathscr{B}$, where

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

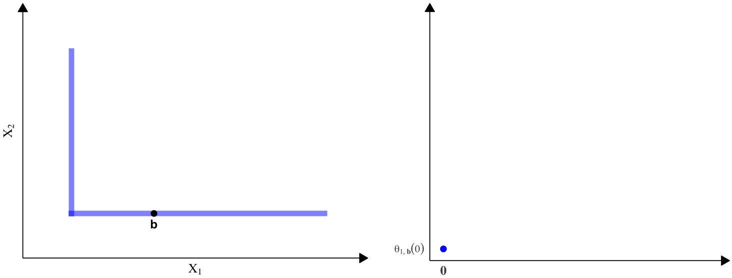
- **Assumption**. Let $t \in \{0, 1\}$.
  - $\mathscr{d} : \mathbb{R}^2 \mapsto [0, \infty)$ satisfies $\|\mathbf{x}_1 - \mathbf{x}_2\| \lesssim \mathscr{d}(\mathbf{x}_1, \mathbf{x}_2) \lesssim \|\mathbf{x}_1 - \mathbf{x}_2\|$ for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathscr{X}$.
  - $k : \mathbb{R} \to [0, \infty)$ is compact supported and Lipschitz continuous, or $k(u) = \mathbb{1}(u \in [-1, 1])$.
  - $\liminf_{h \downarrow 0} \inf_{\mathbf{x} \in \mathscr{B}} \int_{\mathscr{A}_t} k_h(\mathscr{d}(\mathbf{u}, \mathbf{x})) d\mathbf{u} \gtrsim 1$.

- **Identification**. For all $\mathbf{x} \in \mathscr{B}$,

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

with

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

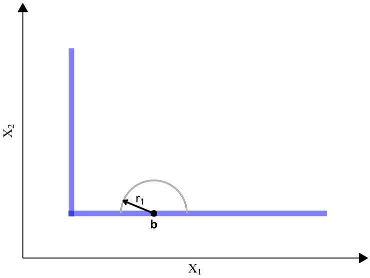


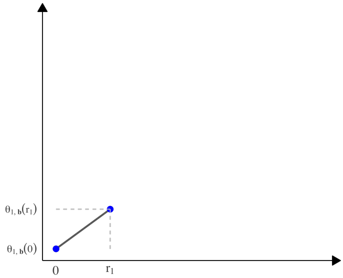
- **Treatment Group**. Bivariate vs. univariate (distance-induced) expectations:

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

where

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





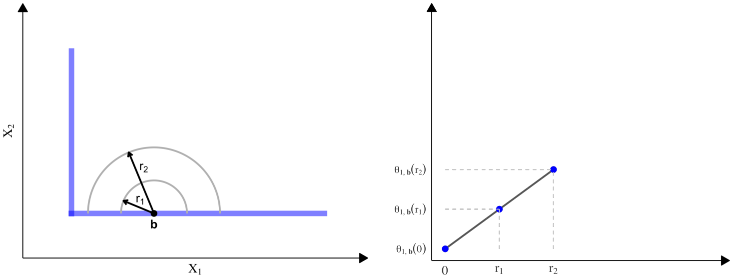
- **Distance range**: $[0, r_1)$. If $\mathbf{x} \mapsto \mathbb{E}[Y(1)|\mathbf{X}_i = \mathbf{b}, \mathbf{X}_i \in \mathscr{A}_1]$ is smooth, then

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

  is also smooth.

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

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



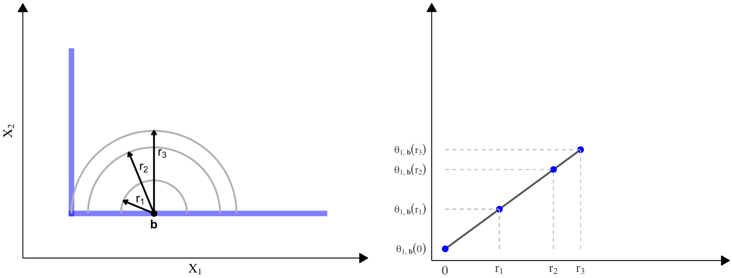
- **Distance range**: $[0, r_2)$. If $\mathbf{x} \mapsto \mathbb{E}[Y(1)|\mathbf{X}_i = \mathbf{b}, \mathbf{X}_i \in \mathscr{A}_1]$ is smooth, then

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

  is also smooth.

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

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

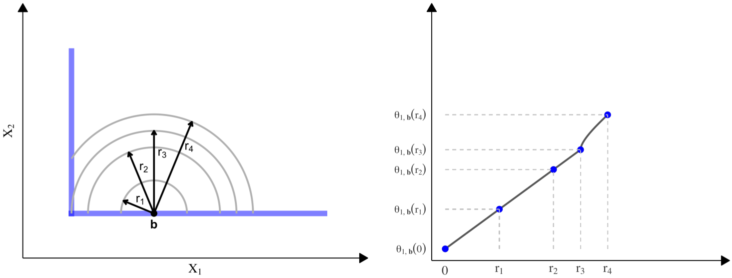


- **Distance range**: $[0, r_3)$. If $\mathbf{x} \mapsto \mathbb{E}[Y(1)|\mathbf{X}_i = \mathbf{b}, \mathbf{X}_i \in \mathscr{A}_1]$ is smooth, then

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

is also smooth.

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

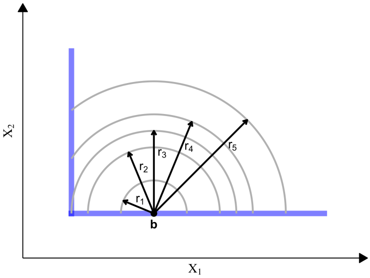


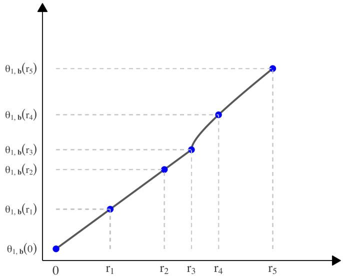
- **Distance range**: $[0, r_4)$. If $\mathbf{x} \mapsto \mathbb{E}[Y(1)|\mathbf{X}_i = \mathbf{b}, \mathbf{X}_i \in \mathscr{A}_1]$ is smooth, then

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

  is not smooth.

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



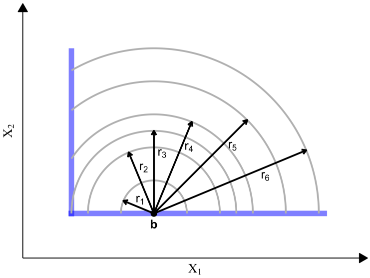


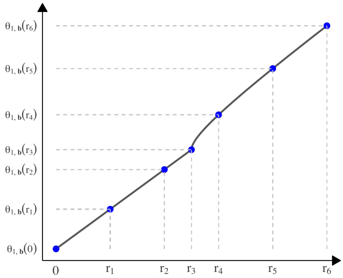
- **Distance range**: $[0, r_5)$. Distance-based local polynomial estimator misspecification bias is

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

regardless p used, that is, not of order h^{p+1} as expected given underlying smoothness.

- **Pointwise Analysis**. Need to choose bandwidth $h \leq r_3 = \mathscr{d}(\mathbf{b}, \text{kink})$.
  - Bandwidth must vary with $\mathbf{b} \in \mathscr{B}$, depending on "smoothness" of boundary!
  - The closer to a kink point on $\mathscr{B}$, the smaller the bandwidth h must be.



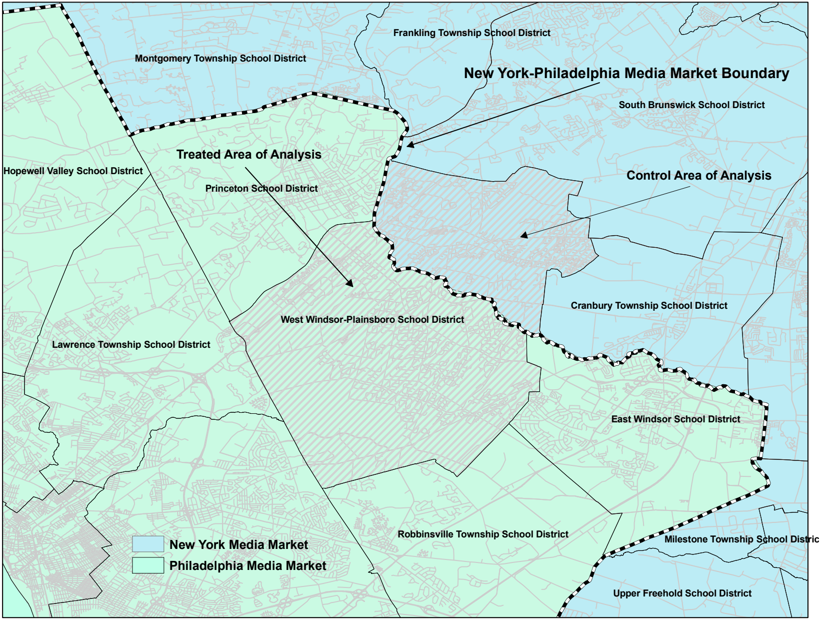


- **Uniform Analysis**. Under minimal regularity conditions, and for any $p \geq 1$,

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

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

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


Frankling Township School District
Montgomery Township School District
New York-Philadelphia Media Market Boundary
South Brunswick School District
Treated Area of Analysis
Hopewell Valley School District
Princeton School District
Control Area of Analysis
Cranbury Township School District
West Windsor-Plainsboro School District
Lawrence Township School District
East Windsor School District
Robbinsville Township School District
Milestone Township School Distric
New York Media Market
Philadelphia Media Market
Upper Freehold School District

Distance-Based Methods: Minimax Result

- Is the "large" bias with non-smooth $\mathscr{B}$ a general problem? Yes!

- **Impossibility Result**. Under standard regularity conditions:

$$\liminf_{n\to\infty} n^{1/4} \inf_{T_n \in \mathscr{T}} \sup_{\mathbb{P}\in\mathscr{P}_{\mathtt{NP}}} \mathbb{E}_{\mathbb{P}}\Big[\sup_{\mathbf{x}\in\mathscr{B}} \big|T_n(\mathbf{U}_n(\mathbf{x})) - \mu(\mathbf{x})\big|\Big] \gtrsim 1,$$

  where

  - $\mathscr{T}$ 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 $\mathbf{x} \in \mathscr{X}$,
  - $\mathscr{B}$ is assumed to be rectifiable, and
  - $\mathscr{P}_{\mathtt{NP}}$ includes q-smooth $\mu(\mathbf{x})$ functions.

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

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

  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**. $\sup_{\mathbf{x}\in\mathscr{X}} \mathbb{E}[|Y_i(t)|^{2+v}|\mathbf{X}_i=\mathbf{x}] < \infty$ for some $v \geq 2$.

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

$$|\widehat{\tau}_{\mathrm{dis}}(\mathbf{x}) - \tau(\mathbf{x})| \lesssim_{\mathbb{P}} \frac{1}{\sqrt{nh^2}} + \frac{1}{n^{\frac{1+v}{2+v}} h^2} + |\mathfrak{B}_n(\mathbf{x})|, \qquad \mathbf{x} \in \mathscr{B},$$

  and

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

- **Pointwise Inference**. Ignoring the potential bias problem when $\mathscr{B}$ is non-smooth, paper establishes distribution theory with valid standard errors for each $\mathbf{x} \in \mathscr{B}$. This result is fairly standard, up to handling $\mathscr{B}$.

- **Uniform Inference**. Ignoring the potential bias problem when $\mathscr{B}$ is non-smooth, paper establishes feasible uniform distribution theory via simulation. This result requires new technical tools, and requires careful handling of $\mathscr{B}$. More details later.

- **Practice**. Valid and invalid practices based on standard univariate RD designs methods.

Outline

Location-Based Methods: Setup

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

- **Estimator**. $\widehat{\tau}(\mathbf{b}_1) = \mathbf{e}_1^\top \widehat{\boldsymbol{\beta}}_1(\mathbf{b}_1) - \mathbf{e}_1^\top \widehat{\boldsymbol{\beta}}_0(\mathbf{b}_1)$ for $\mathbf{x} \in \mathscr{B}$,

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

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

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

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

  This is standard from the literature.

Point Estimation Results for Location-Based Methods

- **Regularity Condition**. $\sup_{\mathbf{x}\in\mathscr{X}} \mathbb{E}[|Y_i(t)|^{2+v}|\mathbf{X}_i = \mathbf{x}] < \infty$ for some $v \geq 2$.

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

$$|\widehat{\tau}(\mathbf{x}) - \tau(\mathbf{x})| \lesssim_{\mathbb{P}} \frac{1}{\sqrt{nh^2}} + \frac{1}{n^{\frac{1+v}{2+v}} h^2} + h^{p+1}, \qquad \mathbf{x} \in \mathscr{B},$$

  and

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

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

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

  and

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

  - Standard bandwidth selection methods developed in the paper.

Inference Results for Location-Based Methods

- **Feasible t-test**. Using standard least squares algebra, $\widehat{\mathsf{T}}(\mathbf{x}) = \frac{\widehat{\tau}(\mathbf{x})-\tau(\mathbf{x})}{\sqrt{\widehat{\Omega}_{\mathbf{x},\mathbf{x}}}}$.

- **Uncertainty Quantification**. Confidence intervals and confidence bands,

$$\widehat{I}(\mathbf{x};\alpha) = \Big[\ \widehat{\tau}(\mathbf{x}) - \mathfrak{q}_\alpha\sqrt{\widehat{\Omega}_{\mathbf{x}}}\ ,\ \widehat{\tau}(\mathbf{x}) + \mathfrak{q}_\alpha\sqrt{\widehat{\Omega}_{\mathbf{x}}}\ \Big], \qquad \mathbf{x}\in\mathscr{B},$$

- **Pointwise Inference**. By standard CLT result, for each $\mathbf{x}\in\mathscr{B}$, set $\mathfrak{q}_\alpha = \Phi^{-1}(1-\alpha/2)$.

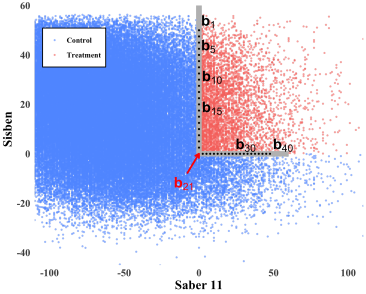
- **Uniform Inference**. Note that

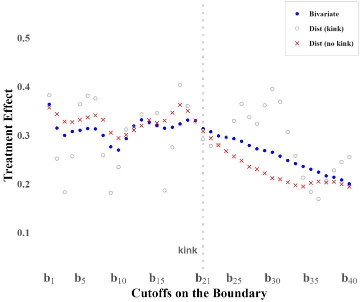
$$\mathbb{P}\Big[\tau(\mathbf{x})\in\widehat{I}(\mathbf{x};\alpha)\ ,\ \text{for all } \mathbf{x}\in\mathscr{B}\Big] = \mathbb{P}\Big[\sup_{\mathbf{x}\in\mathscr{B}}\big|\widehat{\mathsf{T}}(\mathbf{x})\big| \le \mathfrak{q}_\alpha\Big].$$

  1. Establish strong approximation for $(\widehat{\mathsf{T}}(\mathbf{x}) : \mathbf{x}\in\mathscr{B})$ by $(\widehat{Z}_n : \mathbf{x}\in\mathscr{B})$, a Gaussian process conditional on data.
  2. Deduce the distribution of $\sup_{\mathbf{x}\in\mathscr{B}}\big|\widehat{\mathsf{T}}(\mathbf{x})\big|$.
  3. Using simulations, set $\mathfrak{q}_\alpha = \inf\{c>0 : \mathbb{P}[\sup_{\mathbf{x}\in\mathscr{B}}|\widehat{Z}_n(\mathbf{x})| \ge c|\text{data}] \le \alpha\}$.

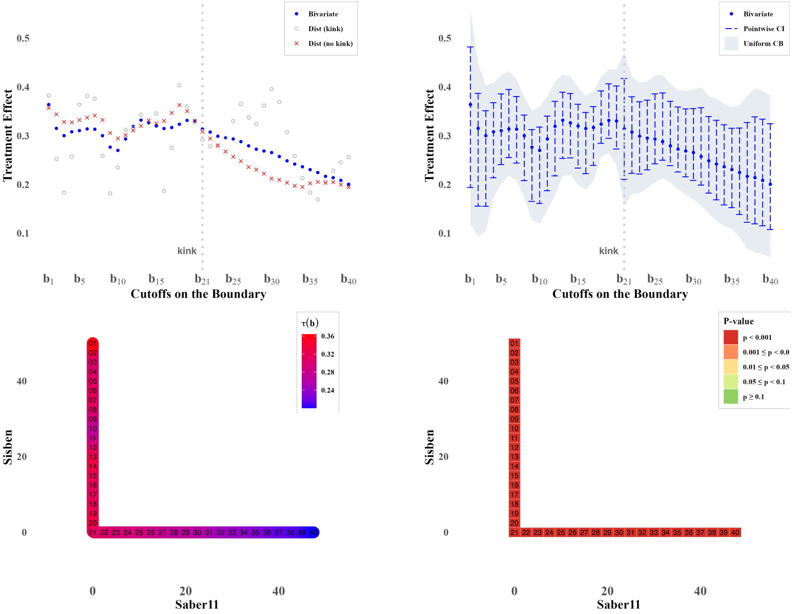
- **Implementation and Bias**. (I)MSE-optimal bandwidth selection for point estimation, robust bias correction for inference.

Outline





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


Bivariate
Dist (kink)
Dist (no kink)
Treatment Effect
kink
Cutoffs on the Boundary
Pointwise CI
Uniform CB
τ(b)
Sisben
Saber11
P-value
p < 0.001
0.01 ≤ p < 0.05
0.05 ≤ p < 0.1
p ≥ 0.1

Outline

Inference Results for Location-Based Methods

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

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

  and a generic plug-in “estimator” thereof is

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

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

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

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

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

- **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

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 $\mathscr{B}$ 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/`