# RD designs with multiple cutoffs or scores

## Northwestern Causal Inference Workshop

Gonzalo Vazquez-Bare

Department of Economics, UC Santa Barbara

August 1, 2024

# Overview

- Multiple cutoffs:
  - Cutoffs change across regions, time periods, etc
  - All units receive the same treatment when they exceed their cutoff
- Cumulative cutoffs:
  - Treatment is multivalued
  - Different dosage of treatment depending on value of $X_i$
  - E.g. $D_i = \mathbb{1}(X_i \leq c_1) + 2\mathbb{1}(c_1 < X_i \leq c_2)$
- Multiple scores:
  - Treatment assigned based on multiple running variables
  - E.g. scholarship if both math and language scores above a cutoff

# RD with multiple cutoffs: some examples

Table: Progresa (Mexico)

| Region | Cutoff |
|---|---|
| 27 | 691.0 |
| 6 | 751.0 |
| 5 | 751.5 |
| 4 | 753.0 |
| 3 | 759.4 |
| 28 | 853.3 |

Table: P-900 (Chile)

| Region | Cutoff |
|---|---|
| 7 | 42.4 |
| 6,8 | 43.4 |
| 13 | 46.4 |
| 9 | 47.4 |
| 2,5,10 | 49.4 |
| 1,3,4 | 51.4 |

# RD with multiple cutoffs

- Common empirical approach: normalizing-and-pooling
  - $C_i \in \mathcal{C}$ (random) cutoff faced by unit $i$
  - Discrete cutoffs: $\mathcal{C} = \{c_0, c_1, ..., c_J\}$
  - Re-centered running variable: $\tilde{X}_i = X_i - C_i$
  - Pooled estimand:

$$\tau^p = \lim_{x \downarrow 0} \mathbb{E}[Y_i | \tilde{X}_i = x] - \lim_{x \uparrow 0} \mathbb{E}[Y_i | \tilde{X}_i = x]$$

- What parameter is this approach identifying?

# Identification under the pooling approach

$$\tau^p = \lim_{x \downarrow 0} \mathbb{E}[Y_i | \tilde{X}_i = x] - \lim_{x \uparrow 0} \mathbb{E}[Y_i | \tilde{X}_i = x]$$

## Identification under pooling (CKTV, 2016)

If the CEFs and $f_{X|C}(x|c)$ are continuous at the cutoffs,

$$\tau^p = \sum_{c \in \mathcal{C}} \mathbb{E}[Y_i(1) - Y_i(0) | X_i = c, C_i = c] \omega(c)$$

where

$$\omega(c) = \frac{f_{X|C}(c|c) \mathbb{P}[C_i = c]}{\sum_{c \in \mathcal{C}} f_{X|C}(c|c) \mathbb{P}[C_i = c]}$$

# Empirical example: Progresa

Table: Pooled and separate RD estimates

| | Cutoff | Estimate | s.e. | Obs. | Weights |
|---|---|---|---|---|---|
| Pooled | 0 | 24.66 | 7.66 | 734 | - |
| Region 3 | 759.4 | 31.64 | 12.91 | 119 | .16 |
| Region 4 | 753.0 | 15.12 | 17.05 | 270 | .21 |
| Region 5 | 751.5 | 24.42 | 9.54 | 474 | .52 |
| Region 6 | 751.0 | 27.98 | 24.65 | 63 | .11 |

# Exploiting multiple cutoffs

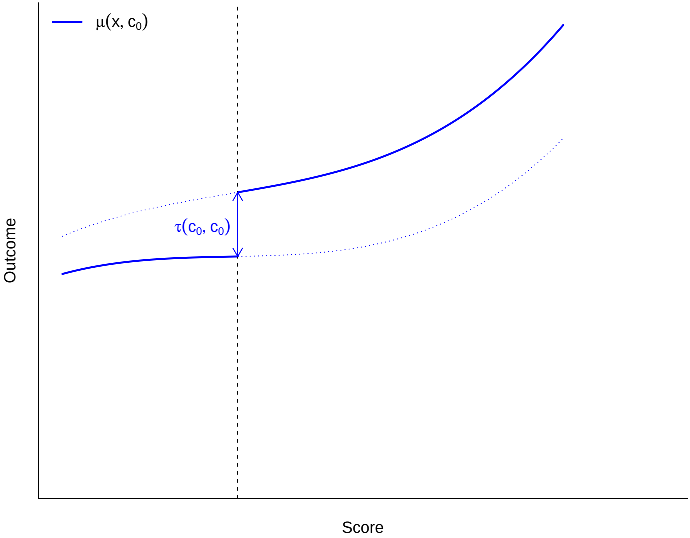
- Two drawbacks of the pooling approach:
  - policy relevance: combines TEs *for different populations*
  - discards variation that can identify parameters of interest
- What are the parameters of interest in this context?
- Potential CEFs:

$$\mu_d(x, c) = \mathbb{E}[Y_i(d) | X_i = x, C_i = c], \qquad d \in \{0, 1\}$$

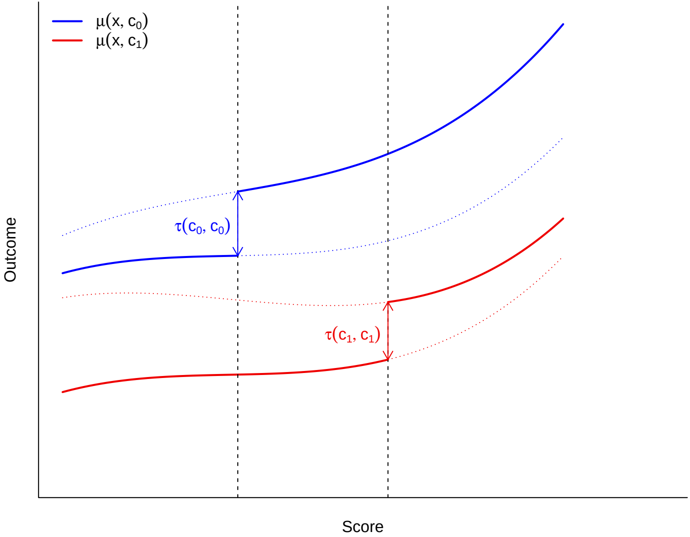
- (Conditional) ATE:

$$\tau(x, c) = \mathbb{E}[Y_i(1) - Y_i(0) | X_i = x, C_i = c] = \mu_1(x, c) - \mu_0(x, c)$$

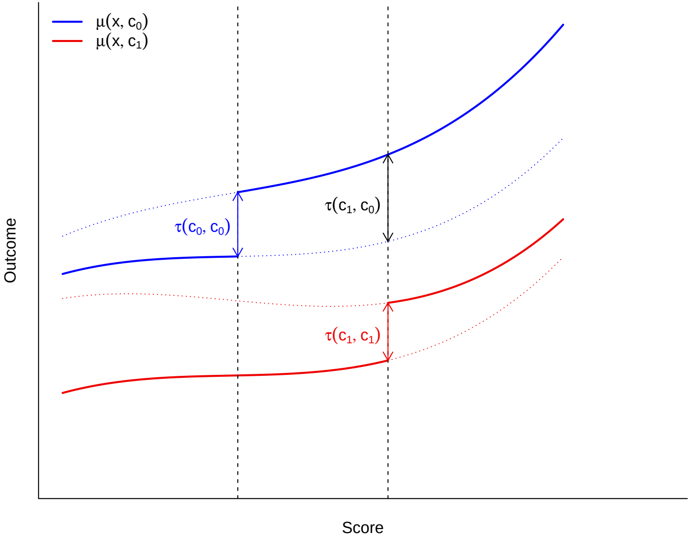
# Multiple cutoffs: parameters of interest



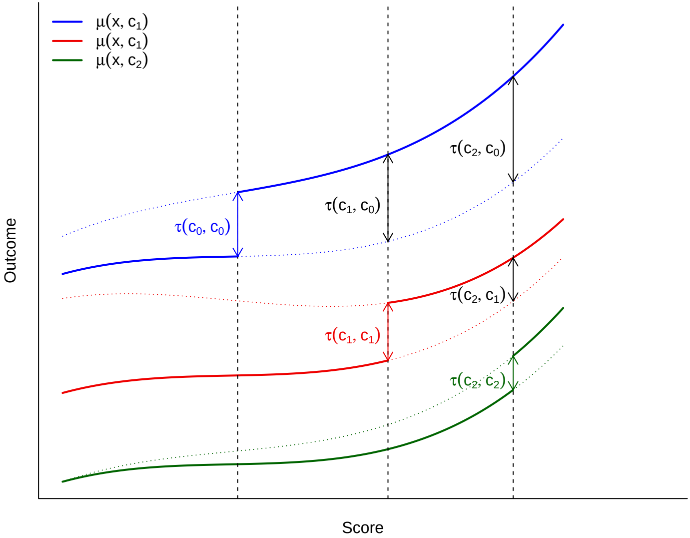
# Multiple cutoffs: parameters of interest



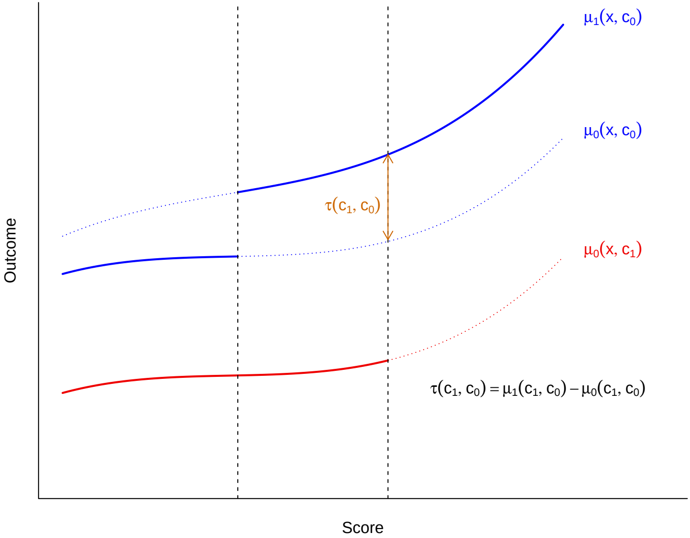
# Multiple cutoffs: parameters of interest



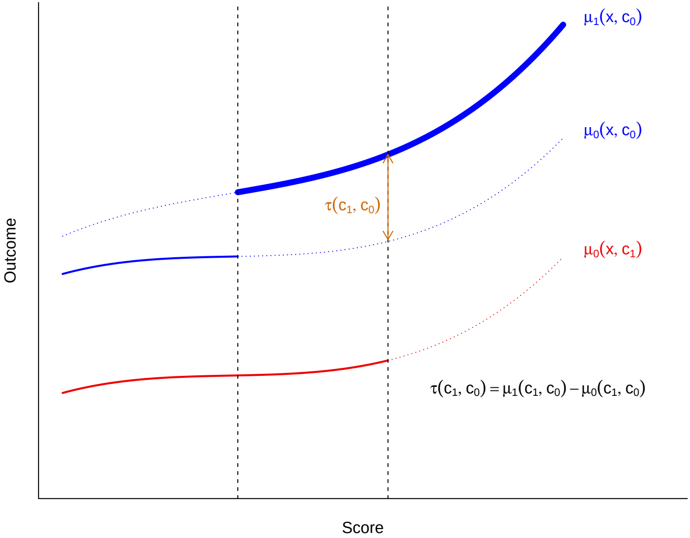
# Multiple cutoffs: parameters of interest



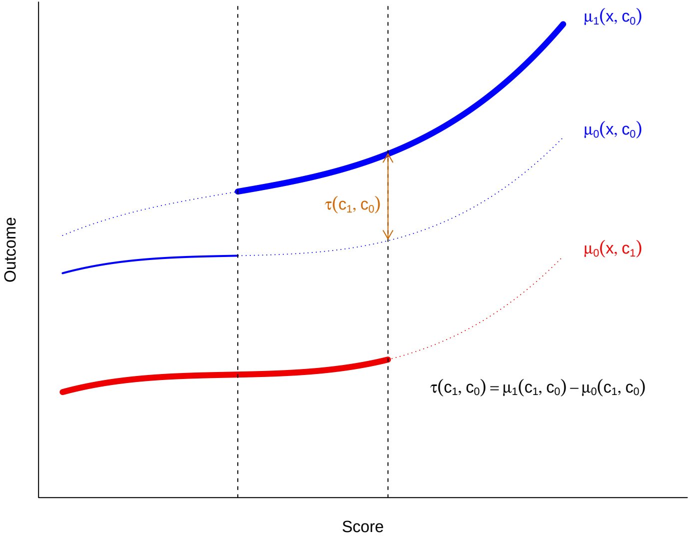
# Intuition: two-cutoff case



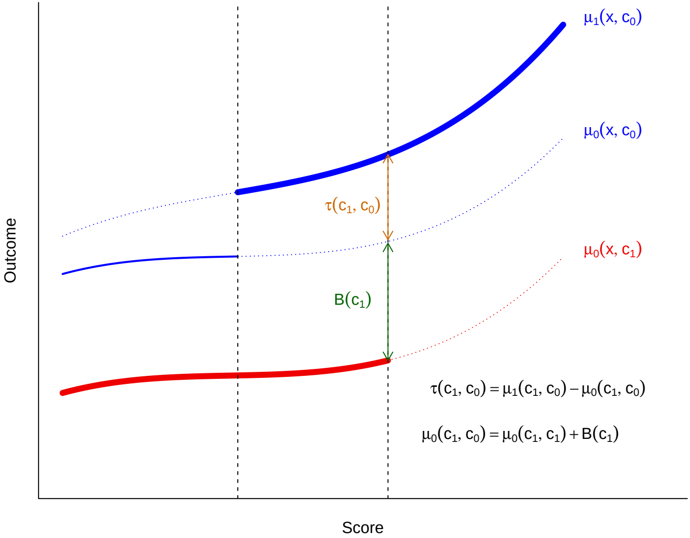
# Intuition: two-cutoff case



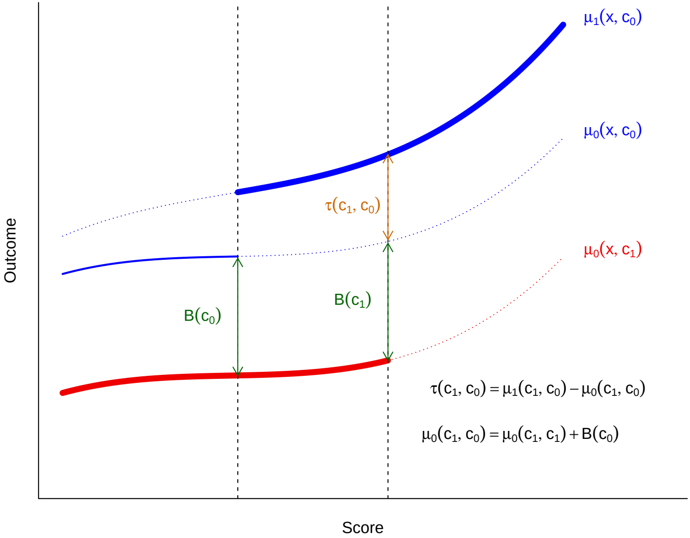
# Intuition: two-cutoff case



# Intuition: two-cutoff case



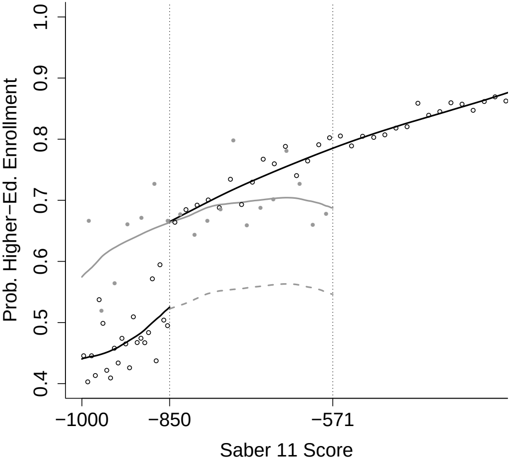
# Intuition: two-cutoff case



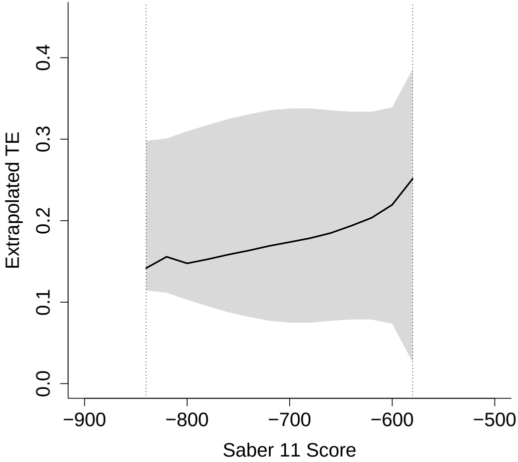
# Empirical illustration: the ACCES program in Colombia

- Cattaneo, Keele, Titiunik and Vazquez-Bare (JASA, forthcoming)
  - Reanalysis of Melguizo, Sanchez and Velasco (2016)
  - Subsidized loan program for post-secondary education
- Eligibility discontinuity: *SABER 11* exam
- Eligibility cutoff varies by year and department (129 cutoffs)
- We select two cutoffs with largest sample sizes
- Nonparametrically estimate effects between the two cutoffs

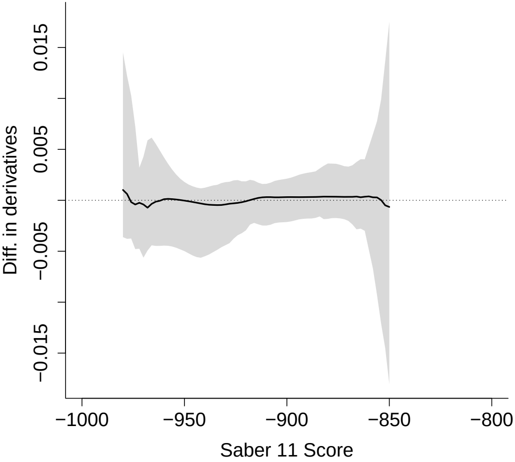
# Effect of ACCES: enrollment in higher ed



# Effect of ACCES: enrollment in higher ed



# Assessing the parallel trends assumption



# RD with cumulative cutoffs: example

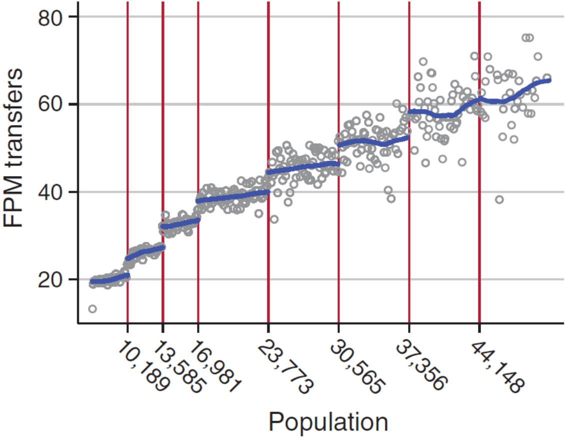
Brollo et al (2013, AER): The Political Resource Curse

- Effect of federal transfers on corruption in Brazil
- Federal transfers jump at population thresholds

Table: Fundo de Participaçao dos Municipios (Brazil)

| Population | Transfer |
|---|---|
| Below 10,189 | 0.6 |
| 10,189-13,584 | 0.8 |
| 13,585-16,980 | 1 |
| 16,981,23772 | 1.2 |
| ⋮ | ⋮ |

# RD with cumulative cutoffs: example



# RD with cumulative cutoffs: parameters

- Multivalued treatment $D_i \in \{d_1, d_2, \ldots, d_J\}$
- Effect of switching from one dosage to the next one:

$$\tau_j = \mathbb{E}[Y_i(d_j) - Y_i(d_{j-1})|X_i = c_j]$$

- Under continuity assumptions,

$$\tau_j = \lim_{x \downarrow c_j} \mathbb{E}[Y_i|X_i = x] - \lim_{x \uparrow c_j} \mathbb{E}[Y_i|X_i = x]$$

# RD with cumulative cutoffs: estimation and inference

- Can use robust bias-corrected techniques cutoff by cutoff
- Unit $i$ is “control” for some units, “treated” for others
- Bandwidth selection:
  - Ensure bandwidths are non-overlapping or
  - Joint estimation accounting for overlap

# RD with multiple scores

- Treatment assigned based on several running variables
- Bivariate score: $\mathbf{X}_i = (X_{1i}, X_{2i})$
- Suppose treatment is assigned if both scores exceed a cutoff:

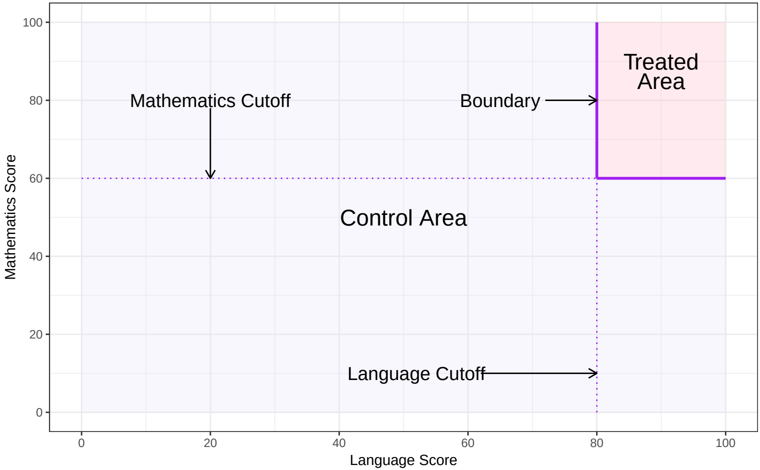
$$D_i = \mathbb{1}(X_{1i} \geq b_1)\mathbb{1}(X_{2i} \geq b_2)$$

- Multidimensional RD parameter:

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

  - ATE at each point in the boundary set $\mathcal{B}$

# RD with multiple scores: example



# RD with multiple scores: identification

- Under continuity assumptions,

$$\tau(\mathbf{b}) = \lim_{\substack{d(\mathbf{x},\mathbf{b})\to 0 \\ \mathbf{x}\in\mathcal{B}_t}} \mathbb{E}[Y_i|\mathbf{X}_i = \mathbf{x}] - \lim_{\substack{d(\mathbf{x},\mathbf{b})\to 0 \\ \mathbf{x}\in\mathcal{B}_c}} \mathbb{E}[Y_i|\mathbf{X}_i = \mathbf{x}]$$

  - $\mathcal{B}_t$ = treated region
  - $\mathcal{B}_c$ = control region

- Need to define a notion of distance $d(\mathbf{x}, \mathbf{b})$

# RD with multiple scores: estimation

- Estimating a whole curve of TE may not be feasible
- Alternative approach: pooling
  - Define distance measure $d(\cdot, \cdot)$
  - Normalize running variable as distance to closest boundary point
  - Run RD on (unidimensional) normalized running variable $\tilde{X}_i$

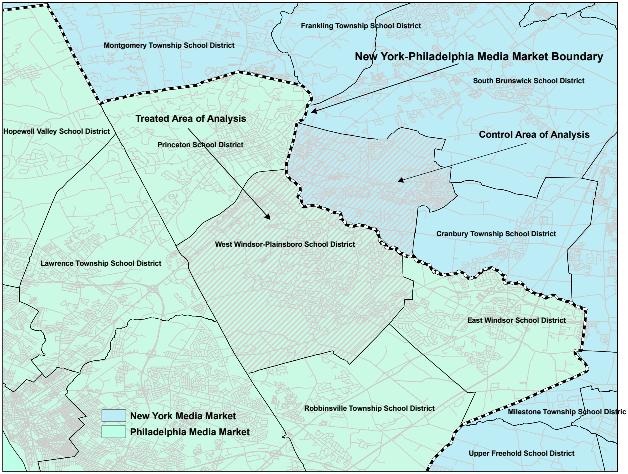
# RD with multiple scores: estimation

- Intermediate approach:
  - Define a set of boundary points of interest
  - E.g. $\{(80,60),(100,60),(80,80)\}$
  - Define a distance measure $d(\cdot,\cdot)$
  - Estimate a pooled RD at each of the specified points
- Helps explore heterogeneity in TEs across boundary points

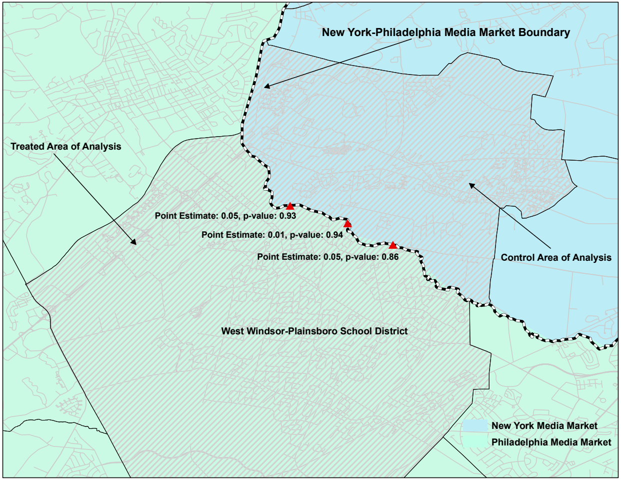
# RD with multiple scores: geographic RD

- Common case of multi-score RD:
    - E.g. treatment and control areas are adjacent counties, municipalities
- Set $\mathcal{B}$ is a geographic boundary separating treated and control areas
- Score $\mathbf{X}_i =$ geographical coordinates (latitude and longitude)

# Geographic RD



# Geographic RD



# Geographic RD: some concerns

- Compound treatments
  - Many policies changing simultaneously at geographic boundary
- Running variable manipulation
  - Units may be able to move from treated to control area or vice versa
- Spillovers / interference
  - Treatment can indirectly affect units in untreated areas
- See Keele and Titiunik (2015, PA)