Regression Discontinuity Designs Under Interference

Elena Dal Torrione[1] Tiziano Arduini[2] Laura Forastiere[1]

[1]Yale University

[2]Tor Vergata University of Rome

EEA 25

Aug 27th 2025

TOR VERGATA
UNIVERSITÀ DEGLI STUDI DI ROMA

Regression Discontinuity Design (RDD)

Treatment D_i is assigned to units based on whether a **score variable** X_i exceeds a known **cutoff** c.

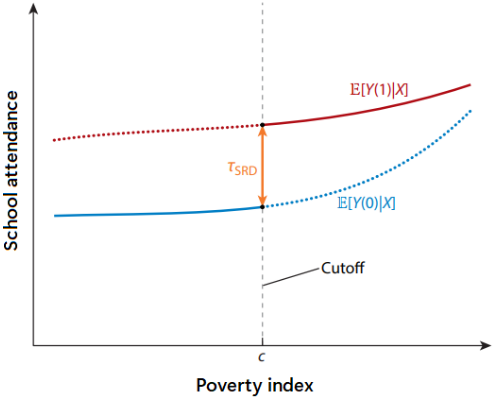


Figure: Source: Cattaneo and Titiunik (2022).

Interference

A unit's outcome is influenced by the treatment of other units.

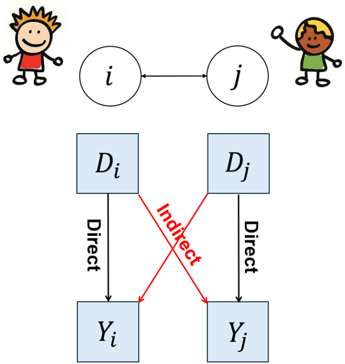


Figure: Direct and indirect effects.

Interference in RDDs

- Spillover effects in RDDs:
  - Evidence in cash transfer programs from a two-stage partial population design (Angelucci and De Giorgi, 2009; Lalive and Cattaneo, 2009).
  - Evaluation of vaccination programs (Basta and Halloran, 2019).

- Few recent works on the topic:
  - Aronow, Basta, and Halloran (2017)
  - Auerbach, Cai, and Rafi (2024)
  - Borusyak and Kolerman-Shemer (2024)

Our contribution: formal framework for network interference in the continuity-based approach to RDDs.

Setting

Let $\boldsymbol{G} = (\mathcal{N}, \mathcal{E})$ be a network consisting of a set of n nodes $\mathcal{N}$ and edges $\mathcal{E}$.

- Interference set: $\mathcal{S}_i \subseteq (\mathcal{N} \setminus \{i\})$ (e.g., friends, village members).
- Treatment vector: $\boldsymbol{D} = (D_i, \boldsymbol{D}_{\mathcal{S}_i}, \boldsymbol{D}_{\mathcal{N}\setminus\{i,\mathcal{S}_i\}})$ with values $\boldsymbol{d} = (d_i, \boldsymbol{d}_{\mathcal{S}_i}, \boldsymbol{d}_{\mathcal{N}\setminus\{i,\mathcal{S}_i\}})$.
- Score vector: $\boldsymbol{X} = (X_i, \boldsymbol{X}_{\mathcal{S}_i}, \boldsymbol{X}_{\mathcal{N}\setminus\{i,\mathcal{S}_i\}})$ with values $\boldsymbol{x} = (x_i, \boldsymbol{x}_{\mathcal{S}_i}, \boldsymbol{x}_{\mathcal{N}\setminus\{i,\mathcal{S}_i\}})$.
- Exposure mapping: $g : \{0,1\}^{|\mathcal{S}_i|} \to \mathcal{G}_i$.
- Neighborhood treatment: $G_i = g(\boldsymbol{D}_{\mathcal{S}_i})$ (e.g., sum of treated friends).
- Observed outcomes: $Y_i = \sum_{\boldsymbol{d}} \mathbb{1}(\boldsymbol{D} = \boldsymbol{d}) \cdot Y_i(\boldsymbol{d})$ (consistency).

Setting

Let $\boldsymbol{G} = (\mathcal{N}, \mathcal{E})$ be a network consisting of a set of n nodes $\mathcal{N}$ and edges $\mathcal{E}$.

- Interference set: $\mathcal{S}_i \subseteq (\mathcal{N} \setminus \{i\})$ (e.g., friends, village members).
- Treatment vector: $\boldsymbol{D} = (D_i, \boldsymbol{D}_{\mathcal{S}_i}, \boldsymbol{D}_{\mathcal{N}\setminus\{i,\mathcal{S}_i\}})$ with values $\boldsymbol{d} = (d_i, \boldsymbol{d}_{\mathcal{S}_i}, \boldsymbol{d}_{\mathcal{N}\setminus\{i,\mathcal{S}_i\}})$.
- Score vector: $\boldsymbol{X} = (X_i, \boldsymbol{X}_{\mathcal{S}_i}, \boldsymbol{X}_{\mathcal{N}\setminus\{i,\mathcal{S}_i\}})$ with values $\boldsymbol{x} = (x_i, \boldsymbol{x}_{\mathcal{S}_i}, \boldsymbol{x}_{\mathcal{N}\setminus\{i,\mathcal{S}_i\}})$.
- Exposure mapping: $g : \{0,1\}^{|\mathcal{S}_i|} \to \mathcal{G}_i$.
- Neighborhood treatment: $G_i = g(\boldsymbol{D}_{\mathcal{S}_i})$ (e.g., sum of treated friends).
- Observed outcomes: $Y_i = \sum_{\boldsymbol{d}} \mathbb{1}(\boldsymbol{D} = \boldsymbol{d}) \cdot Y_i(\boldsymbol{d})$ (consistency).

Assumption: Potential outcomes depend on the effective treatment (D_i, G_i):

$$Y_i(D_i = d, G_i = g) = Y_i(d, g) \quad \forall i$$

Average causal effect: $\mathbb{E}[Y_i(d,g) - Y_i(d',g')]$.

Effective treatment assignment

- D_i is given by the **individual treatment rule**: $D_i = \mathbb{1}(X_i \geq c) = t(X_i)$.

Effective treatment assignment

- D_i is given by the **individual treatment rule**: $D_i = \mathbb{1}(X_i \geq c) = t(X_i)$.
- G_i is a deterministic function of $\boldsymbol{X}_{\mathcal{S}_i}$ through $\boldsymbol{D}_{\mathcal{S}_i}$:

$$G_i = g(\boldsymbol{D}_{\mathcal{S}_i}) = g(\{t(X_j)\}_{j \in \mathcal{S}_i}) = e(\boldsymbol{X}_{\mathcal{S}_i})$$

hence, $G_i = e(\boldsymbol{X}_{\mathcal{S}_i})$ is the **neighborhood treatment rule**.

Effective treatment assignment

- D_i is given by the **individual treatment rule**: $D_i = \mathbb{1}(X_i \geq c) = t(X_i)$.
- G_i is a deterministic function of $\boldsymbol{X}_{\mathcal{S}_i}$ through $\boldsymbol{D}_{\mathcal{S}_i}$:

$$G_i = g(\boldsymbol{D}_{\mathcal{S}_i}) = g(\{t(X_j)\}_{j \in \mathcal{S}_i}) = e(\boldsymbol{X}_{\mathcal{S}_i})$$

  hence, $G_i = e(\boldsymbol{X}_{\mathcal{S}_i})$ is the **neighborhood treatment rule**.

- Then the **effective treatment assignment rule** is

$$(D_i, G_i) = \big(t(X_i), e(\boldsymbol{X}_{\mathcal{S}_i})\big) = \boldsymbol{F}(X_i, \boldsymbol{X}_{\mathcal{S}_i})$$

Effective treatment assignment

- D_i is given by the **individual treatment rule**: $D_i = \mathbb{1}(X_i \geq c) = t(X_i)$.
- G_i is a deterministic function of $\boldsymbol{X}_{\mathcal{S}_i}$ through $\boldsymbol{D}_{\mathcal{S}_i}$:

$$G_i = g(\boldsymbol{D}_{\mathcal{S}_i}) = g(\{t(X_j)\}_{j \in \mathcal{S}_i}) = e(\boldsymbol{X}_{\mathcal{S}_i})$$

  hence, $G_i = e(\boldsymbol{X}_{\mathcal{S}_i})$ is the **neighborhood treatment rule**.

- Then the **effective treatment assignment rule** is

$$(D_i, G_i) = \big(t(X_i), e(\boldsymbol{X}_{\mathcal{S}_i})\big) = \boldsymbol{F}(X_i, \boldsymbol{X}_{\mathcal{S}_i})$$

$\Longrightarrow$ Interference induces a **multiscore RDD** (Reardon and Robinson, 2012; Wong, Steiner, and Cook, 2013).

Excursus: typical multiscore RDDs

The treatment is a deterministic function of two or more *individual* covariates, usually $D_i = \mathbb{1}(X_{i1} \geq c_1, X_{i2} \geq c_2)$.

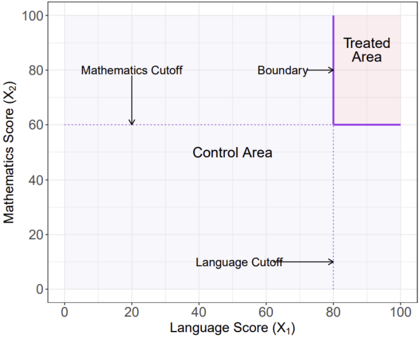


Figure: Multiscore RDD with two scores. Source: (Cattaneo, Idrobo, and Titiunik, 2024a)

Properties of the RDD under interference

Because (D_i, G_i) depends deterministically on $(X_i, \boldsymbol{X}_{\mathcal{S}_i})$:

Properties of the RDD under interference

Because (D_i, G_i) depends deterministically on $(X_i, \boldsymbol{X}_{\mathcal{S}_i})$:

- **Lack of overlap**:

$$\mathbb{P}(\, D_i = d, G_i = g \mid (X_i, \boldsymbol{X}_{\mathcal{S}_i}) = (x_i, \boldsymbol{x}_{\mathcal{S}_i})\,) = \begin{cases} 1 & (x_i, \boldsymbol{x}_{\mathcal{S}_i}) \in \mathcal{X}_{i\mathcal{S}_i}(d, g) \\ 0 & (x_i, \boldsymbol{x}_{\mathcal{S}_i}) \notin \mathcal{X}_{i\mathcal{S}_i}(d, g) \end{cases}$$

Properties of the RDD under interference

Because (D_i, G_i) depends deterministically on $(X_i, \boldsymbol{X}_{\mathcal{S}_i})$:

- **Lack of overlap**:

$$\mathbb{P}(\, D_i = d, G_i = g \,|\, (X_i, \boldsymbol{X}_{\mathcal{S}_i}) = (x_i, \boldsymbol{x}_{\mathcal{S}_i}) \,) = \begin{cases} 1 & (x_i, \boldsymbol{x}_{\mathcal{S}_i}) \in \mathcal{X}_{i\mathcal{S}_i}(d, g) \\ 0 & (x_i, \boldsymbol{x}_{\mathcal{S}_i}) \notin \mathcal{X}_{i\mathcal{S}_i}(d, g) \end{cases}$$

  We never observe units under different effective treatment statutes having the same values of $(X_i, \boldsymbol{X}_{\mathcal{S}_i})$.

Properties of the RDD under interference

Because (D_i, G_i) depends deterministically on $(X_i, \boldsymbol{X}_{\mathcal{S}_i})$:

- **Lack of overlap**:

$$\mathbb{P}(\, D_i = d, G_i = g \,|\, (X_i, \boldsymbol{X}_{\mathcal{S}_i}) = (x_i, \boldsymbol{x}_{\mathcal{S}_i})\,) = \begin{cases} 1 & (x_i, \boldsymbol{x}_{\mathcal{S}_i}) \in \mathcal{X}_{i\mathcal{S}_i}(d,g) \\ 0 & (x_i, \boldsymbol{x}_{\mathcal{S}_i}) \notin \mathcal{X}_{i\mathcal{S}_i}(d,g) \end{cases}$$

  We never observe units under different effective treatment statutes having the same values of $(X_i, \boldsymbol{X}_{\mathcal{S}_i})$.

- **Unconfoundedness**: $Y_i(d,g) \perp\!\!\!\perp D_i,\ G_i \,|\, X_i,\ \boldsymbol{X}_{\mathcal{S}_i}$.

Properties of the RDD under interference

Because (D_i, G_i) depends deterministically on $(X_i, \boldsymbol{X}_{\mathcal{S}_i})$:

- **Lack of overlap**:

$$\mathbb{P}(\, D_i = d, G_i = g \,|\, (X_i, \boldsymbol{X}_{\mathcal{S}_i}) = (x_i, \boldsymbol{x}_{\mathcal{S}_i})\,) = \begin{cases} 1 & (x_i, \boldsymbol{x}_{\mathcal{S}_i}) \in \mathcal{X}_{i\mathcal{S}_i}(d, g) \\ 0 & (x_i, \boldsymbol{x}_{\mathcal{S}_i}) \notin \mathcal{X}_{i\mathcal{S}_i}(d, g) \end{cases}$$

  We never observe units under different effective treatment statutes having the same values of $(X_i, \boldsymbol{X}_{\mathcal{S}_i})$.

- **Unconfoundedness**: $Y_i(d, g) \perp\!\!\!\perp D_i,\ G_i \,|\, X_i,\ \boldsymbol{X}_{\mathcal{S}_i}$. Conditional on $(X_i, \boldsymbol{X}_{\mathcal{S}_i})$ no variation is left in the effective treatment assignment (D_i, G_i).

Properties of the RDD under interference

Because (D_i, G_i) depends deterministically on $(X_i, \boldsymbol{X}_{\mathcal{S}_i})$:

- **Lack of overlap**:

$$\mathbb{P}(\, D_i = d, G_i = g \,|\, (X_i, \boldsymbol{X}_{\mathcal{S}_i}) = (x_i, \boldsymbol{x}_{\mathcal{S}_i})\,) = \begin{cases} 1 & (x_i, \boldsymbol{x}_{\mathcal{S}_i}) \in \mathcal{X}_{i\mathcal{S}_i}(d, g) \\ 0 & (x_i, \boldsymbol{x}_{\mathcal{S}_i}) \notin \mathcal{X}_{i\mathcal{S}_i}(d, g) \end{cases}$$

  We never observe units under different effective treatment statutes having the same values of $(X_i, \boldsymbol{X}_{\mathcal{S}_i})$.

- **Unconfoundedness**: $Y_i(d, g) \perp\!\!\!\perp D_i,\ G_i \,|\, X_i,\ \boldsymbol{X}_{\mathcal{S}_i}$. Conditional on $(X_i, \boldsymbol{X}_{\mathcal{S}_i})$ no variation is left in the effective treatment assignment (D_i, G_i).

$\Longrightarrow$ we can identify causal effects at the effective treatment boundaries!

Causal estimands

- **Boundary point average causal effects**:

$$\tau_{d,g|d',g'}(\bar{x}_i, \bar{\boldsymbol{x}}_{\mathcal{S}_i}) = \mathbb{E}[\, Y_i(d,g) - Y_i(d',g') \,|\, (X_i, \boldsymbol{X}_{\mathcal{S}_i}) = (\bar{x}_i, \bar{\boldsymbol{x}}_{\mathcal{S}_i})\,]$$

for $(\bar{x}_i, \bar{\boldsymbol{x}}_{\mathcal{S}_i}) \in \bar{\mathcal{X}}_{i\mathcal{S}_i}(d, g \,|\, d', g')$.

:

.

Causal estimands

- **Boundary point average causal effects**:

$$\tau_{d,g|d',g'}(\bar{x}_i, \bar{\boldsymbol{x}}_{\mathcal{S}_i}) = \mathbb{E}[\, Y_i(d,g) - Y_i(d',g') \,|\, (X_i, \boldsymbol{X}_{\mathcal{S}_i}) = (\bar{x}_i, \bar{\boldsymbol{x}}_{\mathcal{S}_i}) \,]$$

for $(\bar{x}_i, \bar{\boldsymbol{x}}_{\mathcal{S}_i}) \in \bar{\mathcal{X}}_{i\mathcal{S}_i}(d, g \,|\, d', g')$.

- **Boundary average causal effects**:

$$\begin{aligned} &\tau_{d,g|d',g'}\left(\bar{\mathcal{X}}_{i\mathcal{S}_i}(d, g \,|\, d', g')\right) \\ &\qquad = \mathbb{E}[\, Y_i(d,g) - Y_i(d',g') \,|\, (X_i, \boldsymbol{X}_{\mathcal{S}_i}) \in \bar{\mathcal{X}}_{i\mathcal{S}_i}(d, g \,|\, d', g') \,] \end{aligned}$$

.

Causal estimands

- **Boundary point average causal effects**:

$$\tau_{d,g|d',g'}(\bar{x}_i, \bar{\boldsymbol{x}}_{\mathcal{S}_i}) = \mathbb{E}[\, Y_i(d,g) - Y_i(d',g') \,|\, (X_i, \boldsymbol{X}_{\mathcal{S}_i}) = (\bar{x}_i, \bar{\boldsymbol{x}}_{\mathcal{S}_i})\,]$$

for $(\bar{x}_i, \bar{\boldsymbol{x}}_{\mathcal{S}_i}) \in \bar{\mathcal{X}}_{i\mathcal{S}_i}(d,g \,|\, d',g')$.

- **Boundary average causal effects**:

$$\tau_{d,g|d',g'}\left(\bar{\mathcal{X}}_{i\mathcal{S}_i}(d,g \,|\, d',g')\right)$$
$$= \mathbb{E}[\, Y_i(d,g) - Y_i(d',g') \,|\, (X_i, \boldsymbol{X}_{\mathcal{S}_i}) \in \bar{\mathcal{X}}_{i\mathcal{S}_i}(d,g \,|\, d',g')\,]$$
$$= \frac{\int_{\bar{\mathcal{X}}_{is_i}(d,g \,|\, d',g')} \tau_{d,g|d',g'}(x_i, \boldsymbol{x}_{\mathcal{S}_i}) f(x_i, \boldsymbol{x}_{\mathcal{S}_i}) d(x_i, \boldsymbol{x}_{\mathcal{S}_i})}{\int_{\bar{\mathcal{X}}_{is_i}(d,g \,|\, d',g')} f(x_i, \boldsymbol{x}_{\mathcal{S}_i}) d(x_i, \boldsymbol{x}_{\mathcal{S}_i})}$$

where $f(x_i, \boldsymbol{x}_{\mathcal{S}_i})$ is the joint density of $(x_i, \boldsymbol{x}_{\mathcal{S}_i})$.

Example

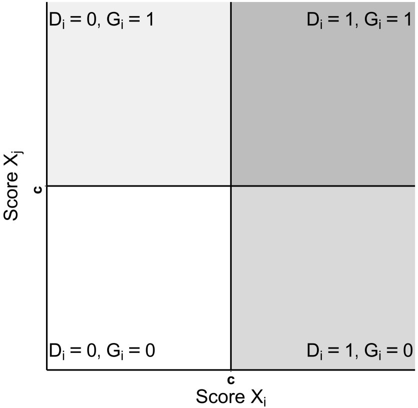


Figure: Multiscore space with one interfering unit.

$\tau_{01|00}(\bar{x}_i, \bar{x}_j)$, $\tau_{01|00}(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))$

Example

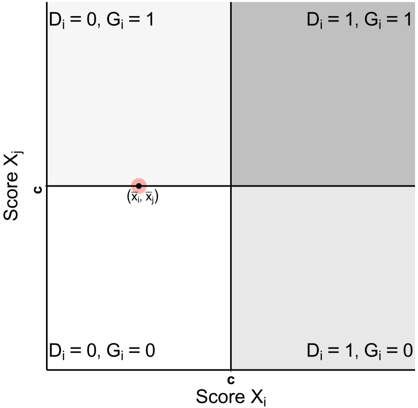


Figure: Multiscore space with one interfering unit.

$\tau_{01|00}(\bar{x}_i, \bar{x}_j),\ \tau_{01|00}(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))$

Example

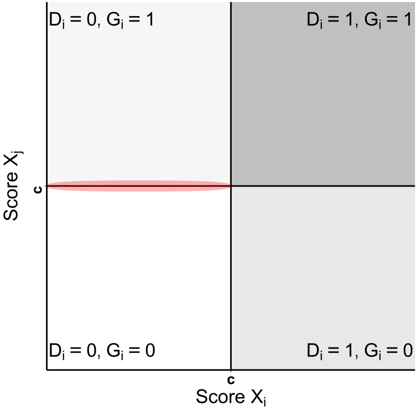


Figure: Multiscore space with one interfering unit.

$\tau_{01|00}(\bar{x}_i, \bar{x}_j),\ \tau_{01|00}(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))$

Key assumptions

Let $\mathcal{B}_\epsilon(\bar{\mathcal{X}}_{i\mathcal{S}_i}(d, g \mid d', g'))$ be the set of points within distance ϵ from $\bar{\mathcal{X}}_{i\mathcal{S}_i}(d, g \mid d', g')$.

Assumption 1 (Identification)

Assume that $(X_i, \boldsymbol{X}_{\mathcal{S}_i})$ has bounded support. For all $i \in \mathcal{N}$, for all d, $d' \in \{0, 1\}$ and for all g, $g' \in \mathcal{G}_i$:

1) *Score density positivity: $f(x_i, \boldsymbol{x}_{\mathcal{S}_i}) > 0$ for all $(x_i, \boldsymbol{x}_{\mathcal{S}_i}) \in \mathcal{B}_\epsilon(\bar{\mathcal{X}}_{i\mathcal{S}_i}(d, g \mid d', g'))$.*
2) *Outcome continuity: $\mathbb{E}[\,Y_i(d, g) \mid (X_i, \boldsymbol{X}_{\mathcal{S}_i}) = (x_i, \boldsymbol{x}_{\mathcal{S}_i})]$ is continuous at all $(x_i, \boldsymbol{x}_{\mathcal{S}_i}) \in \mathcal{X}_{i\mathcal{S}_i}$.*
3) *Score density continuity: $f(x_i, \boldsymbol{x}_{\mathcal{S}_i})$ is continuous at all $(x_i, \boldsymbol{x}_{\mathcal{S}_i}) \in \mathcal{X}_{i\mathcal{S}_i}$.*
4) *Boundedness: $\left|\mathbb{E}[\,Y_i(d, g) \mid (X_i, \boldsymbol{X}_{\mathcal{S}_i}) = (x_i, \boldsymbol{x}_{\mathcal{S}_i})]\right|$ and $f(x_i, \boldsymbol{x}_{\mathcal{S}_i})$ are bounded.*

Identification of boundary effects

Define the following:

$$\mathcal{B}_{\epsilon}^{d,g}(\bar{\mathcal{X}}_{i\mathcal{S}_i}(d,g \mid d',g')) = \mathcal{B}_{\epsilon}(\bar{\mathcal{X}}_{i\mathcal{S}_i}(d,g \mid d',g')) \cap \mathcal{X}_{i\mathcal{S}_i}(d,g)$$
$$\mathcal{B}_{\epsilon}^{d',g'}(\bar{\mathcal{X}}_{i\mathcal{S}_i}(d,g \mid d',g')) = \mathcal{B}_{\epsilon}(\bar{\mathcal{X}}_{i\mathcal{S}_i}(d,g \mid d',g')) \cap \mathcal{X}_{i\mathcal{S}_i}(d',g').$$

Identification of boundary effects

Define the following:

$$\mathcal{B}_{\epsilon}^{d,g}(\bar{\mathcal{X}}_{i\mathcal{S}_i}(d,g \mid d',g')) = \mathcal{B}_{\epsilon}(\bar{\mathcal{X}}_{i\mathcal{S}_i}(d,g \mid d',g')) \cap \mathcal{X}_{i\mathcal{S}_i}(d,g)$$

$$\mathcal{B}_{\epsilon}^{d',g'}(\bar{\mathcal{X}}_{i\mathcal{S}_i}(d,g \mid d',g')) = \mathcal{B}_{\epsilon}(\bar{\mathcal{X}}_{i\mathcal{S}_i}(d,g \mid d',g')) \cap \mathcal{X}_{i\mathcal{S}_i}(d',g').$$

Theorem 1 (Identification)

If Assumption 1 holds then

$$\begin{aligned}
&\tau_{d,g|d',g'}(\bar{\mathcal{X}}_{i\mathcal{S}_i}(d,g \mid d',g')) \\
&\quad = \lim_{\epsilon \to 0} \mathbb{E}[\, Y_i \mid (X_i, \boldsymbol{X}_{\mathcal{S}_i}) \in \mathcal{B}_{\epsilon}^{d,g}(\bar{\mathcal{X}}_{i\mathcal{S}_i}(d,g \mid d',g'))\,] \\
&\quad - \lim_{\epsilon \to 0} \mathbb{E}[\, Y_i \mid (X_i, \boldsymbol{X}_{\mathcal{S}_i}) \in \mathcal{B}_{\epsilon}^{d',g'}(\bar{\mathcal{X}}_{i\mathcal{S}_i}(d,g \mid d',g'))\,]
\end{aligned}$$

Identification under continuity

Suppose Assumption 1 holds in the previous example with $\mathcal{S}_i = \{j\}$.

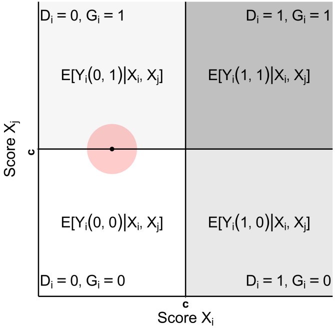


Figure: RDD with one interfering unit.

- $\mathcal{B}_\epsilon(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))$ is a neighborhood of the boundary $\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00)$.
- $\mathcal{B}^{01}_\epsilon(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))$ is its intersection with $\mathcal{X}_{i\mathcal{S}_i}(0,1)$.
- $\mathcal{B}^{00}_\epsilon(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))$ is its intersection with $\mathcal{X}_{i\mathcal{S}_i}(0,0)$.

Identification under continuity

Suppose Assumption 1 holds in the previous example with $\mathcal{S}_i = \{j\}$.

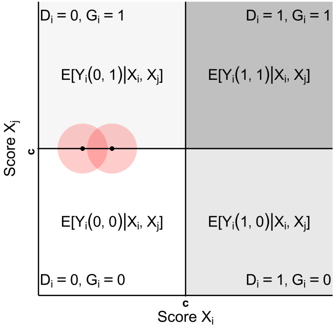


Figure: RDD with one interfering unit.

- $\mathcal{B}_\epsilon(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))$ is a neighborhood of the boundary $\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00)$.
- $\mathcal{B}^{01}_\epsilon(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))$ is its intersection with $\mathcal{X}_{i\mathcal{S}_i}(0,1)$.
- $\mathcal{B}^{00}_\epsilon(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))$ is its intersection with $\mathcal{X}_{i\mathcal{S}_i}(0,0)$.

Identification under continuity

Suppose Assumption 1 holds in the previous example with $\mathcal{S}_i = \{j\}$.

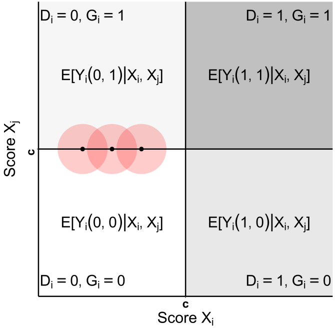


Figure: RDD with one interfering unit.

- $\mathcal{B}_\epsilon(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))$ is a neighborhood of the boundary $\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00)$.
- $\mathcal{B}^{01}_\epsilon(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))$ is its intersection with $\mathcal{X}_{i\mathcal{S}_i}(0,1)$.
- $\mathcal{B}^{00}_\epsilon(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))$ is its intersection with $\mathcal{X}_{i\mathcal{S}_i}(0,0)$.

$$\implies \tau_{01|00}(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00)) = \lim_{\epsilon\to 0} \mathbb{E}[\, Y_i \mid (X_i, X_j) \in \mathcal{B}^{10}_\epsilon(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))\,] - \lim_{\epsilon\to 0} \mathbb{E}[\, Y_i \mid (X_i, X_j) \in \mathcal{B}^{00}_\epsilon(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))\,]$$

Identification under continuity

Suppose Assumption 1 holds in the previous example with $\mathcal{S}_i = \{j\}$.

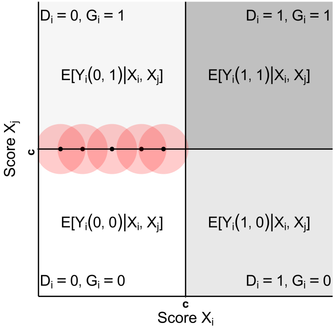


Figure: RDD with one interfering unit.

- $\mathcal{B}_\epsilon(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))$ is a neighborhood of the boundary $\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00)$.
- $\mathcal{B}^{01}_\epsilon(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))$ is its intersection with $\mathcal{X}_{i\mathcal{S}_i}(0,1)$.
- $\mathcal{B}^{00}_\epsilon(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))$ is its intersection with $\mathcal{X}_{i\mathcal{S}_i}(0,0)$.

$$\implies \tau_{01|00}(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00)) = \lim_{\epsilon\to 0} \mathbb{E}[\, Y_i \mid (X_i, X_j) \in \mathcal{B}^{10}_\epsilon(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))\,] - \lim_{\epsilon\to 0} \mathbb{E}[\, Y_i \mid (X_i, X_j) \in \mathcal{B}^{00}_\epsilon(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))\,]$$

Identification under continuity

Suppose Assumption 1 holds in the previous example with $\mathcal{S}_i = \{j\}$.

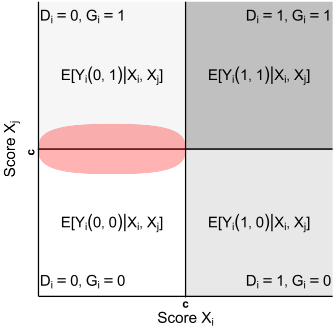


Figure: RDD with one interfering unit.

- $\mathcal{B}_\epsilon(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))$ is a neighborhood of the boundary $\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00)$.
- $\mathcal{B}^{01}_\epsilon(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))$ is its intersection with $\mathcal{X}_{i\mathcal{S}_i}(0,1)$.
- $\mathcal{B}^{00}_\epsilon(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))$ is its intersection with $\mathcal{X}_{i\mathcal{S}_i}(0,0)$.

$$\implies \tau_{01|00}(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00)) = \lim_{\epsilon\to 0} \mathbb{E}[\, Y_i \mid (X_i, X_j) \in \mathcal{B}^{10}_\epsilon(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))\,] - \lim_{\epsilon\to 0} \mathbb{E}[\, Y_i \mid (X_i, X_j) \in \mathcal{B}^{00}_\epsilon(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))\,]$$

Identification under continuity

Suppose Assumption 1 holds in the previous example with $\mathcal{S}_i = \{j\}$.

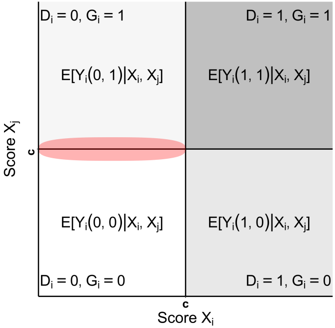


Figure: RDD with one interfering unit.

- $\mathcal{B}_\epsilon(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))$ is a neighborhood of the boundary $\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00)$.
- $\mathcal{B}^{01}_\epsilon(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))$ is its intersection with $\mathcal{X}_{i\mathcal{S}_i}(0,1)$.
- $\mathcal{B}^{00}_\epsilon(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))$ is its intersection with $\mathcal{X}_{i\mathcal{S}_i}(0,0)$.

$$\implies \tau_{01|00}(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00)) = \lim_{\epsilon\to 0} \mathbb{E}[\,Y_i \,|\, (X_i, X_j) \in \mathcal{B}^{10}_\epsilon(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))\,] - \lim_{\epsilon\to 0} \mathbb{E}[\,Y_i \,|\, (X_i, X_j) \in \mathcal{B}^{00}_\epsilon(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))\,]$$

Identification under continuity

Suppose Assumption 1 holds in the previous example with $\mathcal{S}_i = \{j\}$.

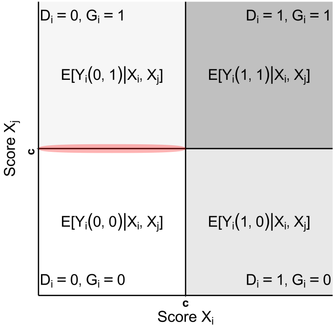


Figure: RDD with one interfering unit.

- $\mathcal{B}_\epsilon(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))$ is a neighborhood of the boundary $\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00)$.
- $\mathcal{B}_\epsilon^{01}(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))$ is its intersection with $\mathcal{X}_{i\mathcal{S}_i}(0,1)$.
- $\mathcal{B}_\epsilon^{00}(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))$ is its intersection with $\mathcal{X}_{i\mathcal{S}_i}(0,0)$.

$$\implies \tau_{01|00}(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00)) = \lim_{\epsilon \to 0} \mathbb{E}[\, Y_i \,|\, (X_i, X_j) \in \mathcal{B}_\epsilon^{10}(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))\,] - \lim_{\epsilon \to 0} \mathbb{E}[\, Y_i \,|\, (X_i, X_j) \in \mathcal{B}_\epsilon^{00}(\bar{\mathcal{X}}_{i\mathcal{S}_i}(01|00))\,]$$

RDD local linear regression

Idea: approximate the outcome regression functions on each side of the cutoff by a local linear fit using kernel-weigthed observations within a bandwidth h (see, e.g. Calonico, Cattaneo, and Titiunik, 2014).

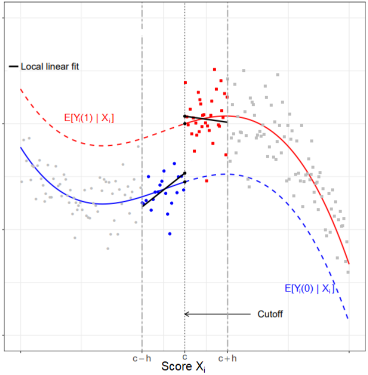
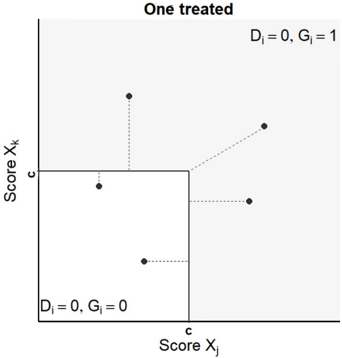


Figure: Source: Cattaneo, Idrobo and Titiunik (2020).

Distance-based local linear estimation

Use the **minimum Euclidean distance** from $\bar{\mathcal{X}}_{iS_i}(d, g \mid d', g')$ as the one-dimensional score.



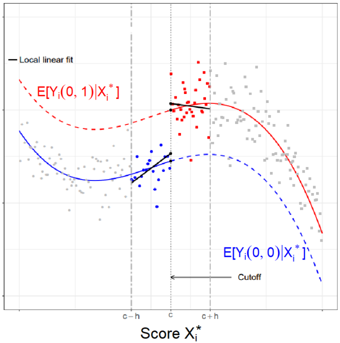


Figure: Distance-based local linear regression.

Dependency neighborhoods

The dependency neighborhood $\boldsymbol{N}_i$ for observation i is a set of indices such that observation i is independent of observation j, for any $j \notin \boldsymbol{N}_i$.

Dependency neighborhoods

The dependency neighborhood $\boldsymbol{N}_i$ for observation i is a set of indices such that observation i is independent of observation j, for any $j \notin \boldsymbol{N}_i$.

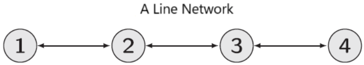


Figure: Source: Leung (2020).

Example: $\boldsymbol{N}_i$ contains all units that are at most two links apart from i.

Dependency neighborhoods

The dependency neighborhood $\boldsymbol{N}_i$ for observation i is a set of indices such that observation i is independent of observation j, for any $j \notin \boldsymbol{N}_i$.

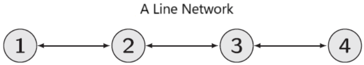


Figure: Source: Leung (2020).

Example: $\boldsymbol{N}_i$ contains all units that are at most two links apart from i.

Dependency neighborhoods can be are represented in a dependency graph $\boldsymbol{W}$ with entry $w_{ij} = 1$ if $j \in \boldsymbol{N}_i$ and $w_{ij} = 0$ otherwise.

Data dependence

- The variables $(\{Y_i(d,g)\}_{i,d,g}, \boldsymbol{X}, \boldsymbol{D},)$ are jointly distributed according to $P(\boldsymbol{G})$, a data-generating distribution depending on $\boldsymbol{G}$.
- We observe the complete network $\boldsymbol{G}$. Each node associated with variables $O_i = (Y_i, X_i, \boldsymbol{X}_{\mathcal{S}_i}, D_i, G_i)$.
- O_i has dependency neighborhoods $\boldsymbol{N}_i \subseteq \mathcal{N}$, $i = 1, \ldots, n$, where $\boldsymbol{N}_i$ is a function of $\boldsymbol{G}$. Let $\boldsymbol{W}$ be the associated dependency graph.

Data dependence

- The variables $(\{Y_i(d,g)\}_{i,d,g}, \boldsymbol{X}, \boldsymbol{D},)$ are jointly distributed according to $P(\boldsymbol{G})$, a data-generating distribution depending on $\boldsymbol{G}$.
- We observe the complete network $\boldsymbol{G}$. Each node associated with variables $O_i = (Y_i, X_i, \boldsymbol{X}_{\mathcal{S}_i}, D_i, G_i)$.
- O_i has dependency neighborhoods $\boldsymbol{N}_i \subseteq \mathcal{N}$, $i = 1, \ldots, n$, where $\boldsymbol{N}_i$ is a function of $\boldsymbol{G}$. Let $\boldsymbol{W}$ be the associated dependency graph.

Assumption 2 (Local Dependence (Leung, 2020))

Assume a sequence of networks $\boldsymbol{G}_n$ and dependency graphs $\boldsymbol{W}_n$. Then $n^{-1}\sum_i |\boldsymbol{N}_{i,n}|^3$ and $n^{-1}\sum_i \sum_{j\neq i}(\boldsymbol{W}_n^3)_{ij}$ are bounded.

Asymptotic normality

Let $\bar{s}$ be the number of components of the score vector $(x_i, \boldsymbol{x}_{\mathcal{S}_i})$ lying at the cutoff in $\bar{\mathcal{X}}_{i\mathcal{S}_i}(d, g \mid d', g')$

Theorem 2 (Asymptotic normality)

Under Assumptions 1- 2 and other regularity assumptions. Let $\bar{\mathcal{X}}_{i\mathcal{S}_i}(d, g \mid d', g')$ *be of dimension* $|\mathcal{S}_i| + 1 - \bar{s}_i$*, with* $\bar{s}_i = \bar{s}$ *for all* i*. If* $nh_n^{\bar{s}} \to \infty$*,* $h_n \to 0$ *and* $h_n = O(n^{1/(\bar{s}+4)})$ *then*

$$\sqrt{nh_n^{\bar{s}}}\Big([\hat{\tau}_{d,g|d',g'}(h_n) - \tau_{d,g|d',g'}(\bar{\mathcal{X}}_{i\mathcal{S}_i}(d, g \mid d', g'))] - h_n^2 B_{d,g|d',g'}\Big) \xrightarrow{d} \mathcal{N}\Big(0, V_{d,g|d',g'}\Big)$$

with $B_{d,g|d',g'} = \frac{\mu^{(2)}_{d,g}(0) - \mu^{(2)}_{d',g'}(0)}{2!}(\Gamma^{-1}\theta)_1$ *and* $V_{d,g|d',g'} = \frac{\bar{\omega}_{d,g}(0) + \bar{\omega}_{d',g'}(0)}{\bar{\tilde{f}}(0)^2}(\Gamma^{-1}\Psi\Gamma^{-1})_{11}$ *and* Γ*,* θ *and* Ψ *depend on the kernel.*

Asymptotic normality

Let $\bar{s}$ be the number of components of the score vector $(x_i, \boldsymbol{x}_{\mathcal{S}_i})$ lying at the cutoff in $\bar{\mathcal{X}}_{i\mathcal{S}_i}(d, g \mid d', g')$

Theorem 2 (Asymptotic normality)

Under Assumptions 1- 2 and other regularity assumptions. Let $\bar{\mathcal{X}}_{i\mathcal{S}_i}(d, g \mid d', g')$ be of dimension $|\mathcal{S}_i| + 1 - \bar{s}_i$, with $\bar{s}_i = \bar{s}$ for all i. If $nh_n^{\bar{s}} \to \infty$, $h_n \to 0$ and $h_n = O(n^{1/(\bar{s}+4)})$ then

$$\sqrt{nh_n^{\bar{s}}}\Big([\hat{\tau}_{d,g|d',g'}(h_n) - \tau_{d,g|d',g'}(\bar{\mathcal{X}}_{i\mathcal{S}_i}(d, g \mid d', g'))] - h_n^2 B_{d,g|d',g'}\Big) \xrightarrow{d} \mathcal{N}\Big(0, V_{d,g|d',g'}\Big)$$

with $B_{d,g|d',g'} = \frac{\mu^{(2)}_{d,g}(0) - \mu^{(2)}_{d',g'}(0)}{2!}(\Gamma^{-1}\theta)_1$ *and* $V_{d,g|d',g'} = \frac{\bar{\omega}_{d,g}(0) + \bar{\omega}_{d',g'}(0)}{\bar{\bar{f}}(0)^2}(\Gamma^{-1}\Psi\Gamma^{-1})_{11}$ *and* Γ, θ *and* Ψ *depend on the kernel.*

Data dependence vanishes asymptotically (see e.g., Masry and Fan, 1997).

Progresa/Oportunidades data

The Mexican PROGRESA/Oportunidaded program paid a subsidy to households in exchange for regular school attendance.

- Units of analysis: children from 320 treated villages attending grade 3-6 of primary school in 1997-1998.
- Score variable: household poverty index.
- Treatment variable: subsidy receipt.
- Outcome variable: school enrolment in 1998-1999.
- Interference sets: children of the same gender in the same village and grade.
- Exposure mapping: $G_i = \frac{\sum_{j \in \mathcal{S}_i} D_j}{|\mathcal{S}_i|}$, proportion of treated neighbors.

Results

	Estimate	S.E.	C.I.	p-value	$N_{d,g}$	$N_{d',g'}$
$\hat{\tau}_{10\|00}(h_n)$	0.243	0.073	(0.100, 0.387)	0.001	190	123
$\hat{\tau}_{11\|01}(h_n)$	0.024	0.043	(-0.061, 0.109)	0.578	216	325
$\hat{\tau}_{01\|00}(h_n)$	0.081	0.103	(-0.121, 0.283)	0.433	160	143
$\hat{\tau}_{10\|01}(h_n)$	-0.046	0.043	(-0.130, 0.038)	0.286	161	363
$\hat{\tau}_{1\|0}(h_n)$	0.074	0.030	(0.015, 0.133)	0.016	730	987
$\hat{\tau}_{indirect}(h_n)$	0.024	0.011	(0.002 , 0.045)	0.030	448	584

Table: Estimates for PROGRESA/Oportunidades data with MSE-optimal bandwidths. S.E. = estimated standard errors, C.I. = Normal approximation confidence intervals. p-value = Normal approximation two-sided p-value. $N_{d,g}$, $N_{d',g'}$ = Effective sample sizes. Number of observations = 3,111.

Boundaries

Simulations

Conclusions

- RDDs under interference are characterized by multivariate scores and boundaries. Causal effects are identified at the boundaries under generalized continuity assumptions.
- Under our assumptions, the typical RDD identification strategy identifies an overall direct effect at the cutoff, which can then be estimated using the standard RDD estimator.
- Boundary causal effects can be estimated with a distance-based local linear estimator, using the proposed variance estimator to account for network dependence.

Appendix

Characterization of effective treatment boundaries

Let $\mathcal{D}_{i\mathcal{S}_i}(d,g) = \{(d_i, \boldsymbol{d}_{\mathcal{S}_i}) \in \mathcal{D}_{i\mathcal{S}_i} \; s.t. \, (d_i, g(\boldsymbol{d}_{\mathcal{S}_i})) = (d,g)\}$, with generic element $(d_i^{d,g}, \boldsymbol{d}_{\mathcal{S}_i}^{d,g})$.

Theorem 3

The effective treatment boundary is given by the image of the following correspondence: $h : \mathcal{D}_{i\mathcal{S}_i}(d,g) \times \mathcal{D}_{i\mathcal{S}_i}(d',g') \twoheadrightarrow \mathcal{X}_{i\mathcal{S}_i}$ *such that*

$$h(d_i^{d,g}, \boldsymbol{d}_{\mathcal{S}_i}^{d',g'}, d_i^{d',g'}, \boldsymbol{d}_{\mathcal{S}_i}^{d',g'}) = \begin{cases} x_z \geq c & d_z^{d,g} = d_z^{d',g'} = 1 \\ x_z \leq c & d_z^{d,g} = d_z^{d',g'} = 0 \\ x_z = c & d_z^{d,g} \neq d_z^{d',g'} \end{cases}$$

for $z \in (\{i\} \cup \mathcal{S}_i)$, *i.e.,* $\bar{\mathcal{X}}_{i\mathcal{S}_i}(d,g \,|\, d',g') = \bigcup h(d_i^{d,g}, \boldsymbol{d}_{\mathcal{S}_i}^{d',g'}, d_i^{d',g'}, \boldsymbol{d}_{\mathcal{S}_i}^{d',g'})$

Note that Theorem 3 also tells us that $\bar{\mathcal{X}}_{i\mathcal{S}_i}(d,g \,|\, d',g')$ is always non empty.

Example

$\bar{\mathcal{X}}_{i\mathcal{S}_i}(d, g \mid d', g')$ can be identified by applying h to each pair of elements $(d_i^{d,g}, \boldsymbol{d}_{\mathcal{S}_i}^{d,g})$ and $(d_i^{d',g'}, \boldsymbol{d}_{\mathcal{S}_i}^{d',g'})$ and taking the union of the resulting sets.

- Suppose $\mathcal{S}_i = 2$ for each i, and we want to estimate the boundary effect of, say, having all neighbors treated, $G_i = 1$, and having no neighbors treated, $G_i = 0$, for someone ineligible $(D_i = 0)$.
  - $\mathcal{D}_{i\mathcal{S}_i}(0, 1) = \{(0, 1, 1)\}$
  - $\mathcal{D}_{i\mathcal{S}_i}(0, 0) = \{(0, 0, 0)\}$

- Then $h\big((0, 1, 1), (0, 0, 0)\big) = \{(x_i, \boldsymbol{x}_{\mathcal{S}_i}) \in \mathcal{X}_{i\mathcal{S}_i} : x_i \leq c \text{ and } x_j = c \text{ for } j \in \mathcal{S}_i\}$.

Go back

Distance-based estimator

Let $\boldsymbol{Y} = [Y_1, ..., Y_n]'$ and $\tilde{X}_n = [\tilde{X}_1, ..., \tilde{X}_n]^T$, and let $\tilde{X}(h_n) = [\boldsymbol{r}(\tilde{X}_1/h_n), ..., \boldsymbol{r}(\tilde{X}_n/h_n)]$ with $r(u) = [1, u]$. Define:

- $\boldsymbol{w}_{d,g}(h_n)$ with element $w_{i,d,g}(h_n) = \mathbb{1}(T_i = 1, \tilde{X}_i \geq 0) \cdot K_{h_n}(\tilde{X}_i)$, and $\boldsymbol{w}_{d',g'}(h_n)$ with elements $w_{i,d',g'}(h_n) = \mathbb{1}(T_i = 1, \tilde{X}_i < 0) \cdot K_{h_n}(\tilde{X}_i)$
- $W_{d,g}(h_n) = \text{diag}\big(\boldsymbol{w}_{d,g}(h_n)\big)$, and $W_{d',g'}(h_n) = \text{diag}\big(\boldsymbol{w}_{d',g'}(h_n)\big)$
- $\Gamma_{d,g}(h_n) = \tilde{X}(h_n)^T W_{d,g}(h_n)\tilde{X}(h_n)/n$, and $\Gamma_{d',g'}(h_n) = \tilde{X}(h_n)^T W_{d',g'}(h_n)\tilde{X}(h_n)/n$

Letting $H(h_n) = \text{diag}(1, h_n^{-1})$, the solution to the previous minimization problem is

$$[\hat{\mu}_{d,g}(h_n), \hat{\mu}^{(1)}_{d,g}(h_n)]^T = H(h_n)\Gamma^{-1}_{d,g}(h_n)\tilde{X}(h_n)^T W_{d,g}(h_n)\boldsymbol{Y}/n$$
$$[\hat{\mu}_{d',g'}(h_n), \hat{\mu}^{(1)}_{d,g}(h_n)]^T = H(h_n)\Gamma^{-1}_{d',g'}(h_n)\tilde{X}(h_n)^T W_{d',g'}(h_n)\boldsymbol{Y}/n$$

The estimator for the boundary causal effect is

$$\hat{\tau}_{d,g|d',g'}(h_n) = \hat{\mu}_{d,g}(h_n) - \hat{\mu}_{d',g'}(h_n)$$

Go back

Description of Treatment Boundaries

	Boundary	All observations		Treatment		Control	
		N	Range	N	Range	N	Range
$\tau_{10\|00}$	$X_i = 0,\ X_j < 0\ \forall j \in \mathcal{S}_i$	597	[-527.75, 365.50]	233	[0.00, 365.50]	364	[-527.75, -0.50]
$\tau_{11\|01}$	$X_i = 0,\ X_j \geq 0\ \forall j \in \mathcal{S}_i$	1303	[-542.50, 379.50]	957	[0.50, 379.50]	346	[-542.50, -0.50]
$\tau_{01\|00}$	$X_i < 0,\ X_j = 0\ \forall j \in \mathcal{S}_i$	710	[-700.72, 478.44]	346	[0.00, 478.44]	364	[-700.72, -0.50]
$\tau_{11\|10}$	$X_i \geq 0,\ X_j = 0\ \forall j \in \mathcal{S}_i$	1190	[-542.50, 575.94]	957	[0.50, 575.94]	233	[-542.50, -0.50]

Table: Boundaries and sample sizes of each effect for PROGRESA/Oportunidades data. The range refers to the minimum distance score.

Go back

Simulation Setup

- Z clusters of size 3 where $\{1z, 2z, 3z\}$ is the set of units in cluster z.
- Interference set $\mathcal{S}_{iz} = \{jz, kz\}$ and exposure mapping $G_{iz} = \mathbb{1}\{\sum_{j \in \mathcal{S}_{iz}} D_j > 0\}$.
- $Y_{iz} = m(X_{iz}, \boldsymbol{X}_{\mathcal{S}_{iz}}) + \epsilon_{iz}$, where

$$m(x_{iz}, \boldsymbol{x}_{\mathcal{S}_{iz}}) = \begin{cases} 7.2 + 3.2x_{iz} + 7.2x_{iz}^2 + 1.2(x_{jz} + x_{kz}) + 0.2x_{jz}x_{kz}, & \text{if } D_{iz} = 1, G_{iz} = 0 \\ 6.7 + 2.9x_{iz} + 3.2x_{iz}^2 + 2.3(x_{jz} + x_{kz}) + 0.2x_{jz}x_{kz}, & \text{if } D_{iz} = 1, G_{iz} = 1 \\ 0.9 + 1.7x_{iz} - 1.5x_{iz}^2 + 1.2(x_{jz} + x_{kz}) + 0.2x_{jz}x_{kz}, & \text{if } D_{iz} = 0, G_{iz} = 0 \\ 3.5 + 1.2x_{iz} - 3.1x_{iz}^2 + 2.3(x_{jz} + x_{kz}) + 0.2x_{jz}x_{kz}, & \text{if } D_{iz} = 0, G_{iz} = 1 \end{cases}$$

- $[X_{1z}, X_{2z}, X_{3z}]^T \sim TN(\mu, \Sigma, -5, 5)$ with $\mu = (0, 0, 0)$ and $\Sigma = \begin{bmatrix} 1 & 0.5 & 0.8 \\ 0.5 & 1 & 0.8 \\ 0.8 & 0.8 & 1 \end{bmatrix}$
- $\epsilon_{iz} = \frac{1}{2}(\tilde{\eta}_{jz} + \tilde{\eta}_{zk}) + \eta_{iz}$ where $j \in \mathcal{S}_{iz}$ $\tilde{\eta}_{jz} = \eta_{jz}[2 \cdot \mathbb{1}(D_{iz} + D_{jz} = 2) + 4 \cdot \mathbb{1}(D_{iz} + D_{jz} = 0)) - 2 \cdot \mathbb{1}(D_{iz} + D_{jz} = 1)]$ and η are i.i.d.n.

Go back

Simulation Results

		Bias	S.D.	S.E.	S.E.$_{i.i.d.}$	C.R.	C.R.$_{i.i.d.}$	$\bar{N}_{d,g}$	$\bar{N}_{d',g'}$
	$\hat{\tau}_{10\|00}(h_n)$	0.122	3.022	2.367	2.319	0.884	0.879	26.085	42.674
	$\hat{\tau}_{11\|01}(h_n)$	-0.076	1.368	1.262	1.210	0.946	0.934	106.200	78.335
N = 750	$\hat{\tau}_{01\|00}(h_n)$	-0.069	2.975	2.462	1.902	0.902	0.819	78.180	99.396
	$\hat{\tau}_{11\|10}(h_n)$	-0.132	2.892	2.235	2.207	0.883	0.883	46.547	26.877
	$\hat{\tau}_{10\|00}(h_n)$	-0.140	1.922	1.662	1.609	0.911	0.905	55.658	95.344
	$\hat{\tau}_{11\|01}(h_n)$	-0.136	0.978	0.898	0.852	0.928	0.915	218.810	160.431
N = 1500	$\hat{\tau}_{01\|00}(h_n)$	-0.010	2.022	1.742	1.330	0.916	0.822	161.721	208.540
	$\hat{\tau}_{11\|10}(h_n)$	0.043	1.862	1.608	1.584	0.918	0.916	99.019	56.490
	$\hat{\tau}_{10\|00}(h_n)$	-0.117	1.308	1.153	1.107	0.924	0.912	116.704	203.776
	$\hat{\tau}_{11\|01}(h_n)$	-0.129	0.699	0.639	0.606	0.917	0.904	433.762	319.482
N = 3000	$\hat{\tau}_{01\|00}(h_n)$	0.009	1.348	1.218	0.926	0.931	0.833	332.142	427.524
	$\hat{\tau}_{11\|10}(h_n)$	0.054	1.346	1.139	1.120	0.914	0.902	208.597	116.592
	$\hat{\tau}_{10\|00}(h_n)$	-0.144	0.893	0.813	0.776	0.930	0.914	239.157	423.078
	$\hat{\tau}_{11\|01}(h_n)$	-0.119	0.502	0.461	0.437	0.922	0.901	837.709	623.791
N = 6000	$\hat{\tau}_{01\|00}(h_n)$	0.014	0.932	0.862	0.653	0.943	0.846	669.710	867.682
	$\hat{\tau}_{11\|10}(h_n)$	0.003	0.927	0.811	0.797	0.924	0.917	420.011	234.515

Table: Simulation results. One thousand replications for sample size $N = \{750, 1500, 3000, 6000\}$ and number of clusters $N_k = \{250, 300, 1000, 2000\}$.

Go back