"Isotonic regression discontinuity designs"

Andrii Babii [1] Rohit Kumar [2]

[1]UNC Chapel Hill

[2]Indian Statistical Institute

November 14, 2019

Preview

Motivation

Regression discontinuity designs and shape restrictions.

Preview

Motivation

Regression discontinuity designs and shape restrictions.

Approach

1. New isotonic sharp and fuzzy RDD estimators (iRDD) based on the boundary corrected isotonic regression;
2. Do not estimate tuning parameters.

Preview

Motivation

Regression discontinuity designs and shape restrictions.

Approach

1. New isotonic sharp and fuzzy RDD estimators (iRDD) based on the boundary corrected isotonic regression;
2. Do not estimate tuning parameters.

Results

1. Isotonic regression is inconsistent at the boundary of its support;
2. Non-standard asymptotic approximation for boundary corrected iRDD estimators based on new tightness results;
3. New solution to the bootstrap inconsistency that does not rely on the nonparametric smoothing.

Nonparametric shape restrictions

(Chetverikov, Santos, Shaikh, 2018)

1. Sharper identification: getting point identification and improving partial identification;

Nonparametric shape restrictions

(Chetverikov, Santos, Shaikh, 2018)

1. Sharper identification: getting point identification and improving partial identification;
2. Testable implications based on shape restrictions;

Nonparametric shape restrictions

(Chetverikov, Santos, Shaikh, 2018)

1. Sharper identification: getting point identification and improving partial identification;
2. Testable implications based on shape restrictions;
3. Improving finite sample estimation and inference using more information about the DGP;

Nonparametric shape restrictions

(Chetverikov, Santos, Shaikh, 2018)

1. Sharper identification: getting point identification and improving partial identification;
2. Testable implications based on shape restrictions;
3. Improving finite sample estimation and inference using more information about the DGP;
4. Shape restrictions are arguably more economically meaningful than smoothness restrictions: one derivative vs two derivatives.

Monotone RDDs

Study	Outcome(s)	Treatment(s)	Running variable
Lee (2008)	Votes share in next election	Incumbency	Initial votes share
Duflo, Dupas and Kremer (2011)	Endline scores	Higher-achieving peers	Intitial attainment
Abdulkadiroğlu, Angrist and Pathak (2014)	Standardized test scores	Attending elite school	Admission scores
Lucas and Mbiti (2014)	Probability of graduation	Attending elite secondary school	Admission scores
Hoekstra (2009)	Earnings	Attending flagship state university	SAT score
Clark (2010)	Test scores, university enrollment	Attending selective high school	Assignment test
Kaniel and Parham (2017)	Net capital flow	Appearance in the WSJ ranking	Returns
Schmieder, Von Wachter and Bender (2012)	Unemployment duration	Unemployment benefits	Age
Card, Dobkin, and Maestas (2008)	Health care utilization	Coverage under Medicare	Age
Shigeoka (2014)	Outpatient visits	Cost-sharing policy	Age
Carpenter and Dobkin (2009)	Alcohol-related mortality	Ability to drink legally	Age
Jacob and Lefgren (2004)	Academic achievements	Summer school, grade retention	Test scores
Baum-Snow and Marion (2009)	Income, property value	Tax credit program	Fraction of eligible
Buettner (2006)	Business tax rate	Fiscal equalization transfers	Tax base
Card, Chetty, and Weber (2007)	Job finding hazard	Severance pay	Job tenure
Chiang (2009)	Medium run test scores	Sanctions threat	School performance
Ferreira (2010)	Probability to move to a new house	Ability to transfer tax benefits	Age
Lalive (2007)	Unemployment duration	Unemployment benefits	Age
Litschig and Morrison (2013)	Education, literacy, poverty	Government transfers	Size of municipality
Ludwig and Miller (2007)	Mortality, educational attainment	Head Start funding	County poverty rate
Matsudaira (2008)	Test scores	Summer school	Test scores
Chay and Greenstone (2005)	Housing prices	Regulatory status	Pollution levels
Greenstone and Gallagher (2012)	Housing prices	Superfund clean-up status	Ranking of hazard

Figure: Examples of monotone designs

Sharp designs

1. Potential outcomes framework (Hahn, Todd, and Van der Klaauw, 2001)
$$Y = DY_1 + (1-D)Y_0,$$
where
   - $D = \mathbb{1}\{X \geq c\}$ is the treatment indicator;
   - $X \in \mathbf{R}$ is the running variable and c is the cut-off;
   - $Y_1, Y_0 \in \mathbf{R}$ are potential outcomes for treated and untreated.

Sharp designs

1. Potential outcomes framework (Hahn, Todd, and Van der Klaauw, 2001)

$$Y = DY_1 + (1 - D)Y_0,$$

   where
   - $D = \mathbb{1}\{X \geq c\}$ is the treatment indicator;
   - $X \in \mathbf{R}$ is the running variable and c is the cut-off;
   - $Y_1, Y_0 \in \mathbf{R}$ are potential outcomes for treated and untreated.

2. Causal effect

$$\theta = \mathbb{E}[Y_1 - Y_0 | X = c].$$

Sharp designs

1. Potential outcomes framework (Hahn, Todd, and Van der Klaauw, 2001)

$$Y = DY_1 + (1 - D)Y_0,$$

   where

   - $D = \mathbb{1}\{X \geq c\}$ is the treatment indicator;
   - $X \in \mathbf{R}$ is the running variable and c is the cut-off;
   - $Y_1, Y_0 \in \mathbf{R}$ are potential outcomes for treated and untreated.

2. Causal effect

$$\theta = \mathbb{E}[Y_1 - Y_0 | X = c].$$

3. (Y, D, X) are observed while (Y_1, Y_0) are not: fundamental problem of causal inference.

Identification: assumptions

(OC) One sided continuity: $x \mapsto \mathbb{E}[Y_1|X = x]$ is right-continuous and $x \mapsto \mathbb{E}[Y_0|X = x]$ is left-continuous at the cut-off.

(M1) Monotonicity 1: $x \mapsto \mathbb{E}[Y_1|X = x]$ and $x \mapsto \mathbb{E}[Y_0|X = x]$ are monotone in some neighborhood of the cut-off.

(M2) Monotonicity 2: $\mathbb{E}[Y_1|X = c] \geq \mathbb{E}[Y_0|X = c]$ in the non-decreasing case or $\mathbb{E}[Y_1|X = c] \leq \mathbb{E}[Y_0|X = c]$ in the non-increasing case

Identification

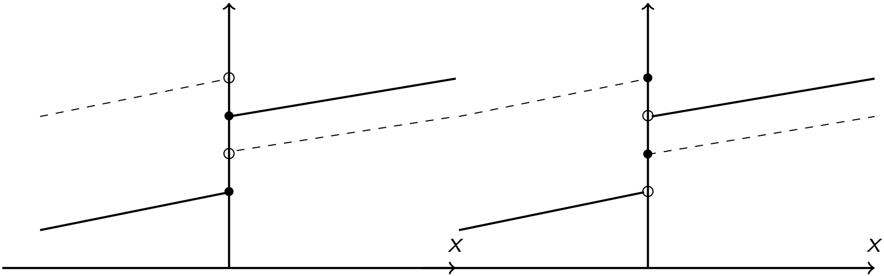
Theorem

Suppose that (OC) and (M1) assumptions are satisfied. Then

$$\lim_{x\downarrow c} \mathbb{E}[Y|X = x] - \lim_{x\uparrow c} \mathbb{E}[Y|X = x] \quad (1)$$

exists and equals to θ. Moreover, under (M1) and (M2) if θ equals to the expression in Eq. 1, then the (OC) conditions holds.

Illustration



(a) Identification under (OC) (b) Counterexample when (OC) fails

Figure: Identification in the sharp RDD. The thick line represents $\mathbb{E}[Y_0|X=x], x<0$ and $\mathbb{E}[Y_1|X=x], x\geq 0$ while the dashed line represents $\mathbb{E}[Y_1|X=x], x<0$ and $\mathbb{E}[Y_0|X=x], x\geq 0$. The thick line coincides with $x\mapsto \mathbb{E}[Y|X=x]$.

Comments

1. Relax continuity to the one-sided continuity for sharp designs (is this well-known?).
2. Under two monotonicity conditions, the one-sided continuity is the weakest possible identifying assumption.
3. Manipulation in the running variable seems to be related to failure of the one-sided continuity of $x \mapsto \mathbb{E}[Y_0|X = x]$ (testable implications?).

Estimation

Causal effect

$$\theta = \lim_{x \downarrow c} \mathbb{E}[Y|X = x] - \lim_{x \uparrow c} \mathbb{E}[Y|X = x].$$

1. Empirical practice: estimate conditional mean functions before and after the cut-off using nonparametric local polynomial estimators and compute the difference.

Estimation

Causal effect

$$\theta = \lim_{x \downarrow c} \mathbb{E}[Y|X = x] - \lim_{x \uparrow c} \mathbb{E}[Y|X = x].$$

1. Empirical practice: estimate conditional mean functions before and after the cut-off using nonparametric local polynomial estimators and compute the difference.
2. Asymptotic properties are well-known, see (Fan and Gijbels, 1992).

Estimation

Causal effect

$$\theta = \lim_{x \downarrow c} \mathbb{E}[Y|X = x] - \lim_{x \uparrow c} \mathbb{E}[Y|X = x].$$

1. Empirical practice: estimate conditional mean functions before and after the cut-off using nonparametric local polynomial estimators and compute the difference.
2. Asymptotic properties are well-known, see (Fan and Gijbels, 1992).
3. Need to select the kernel function and the bandwidth parameter. The bandwidth is typically estimated from the data and the theory is developed for the deterministic bandwidth.

Contributions of this paper

1. New approach to monotone sharp and fuzzy RDD based on the isotonic regression.

Contributions of this paper

1. New approach to monotone sharp and fuzzy RDD based on the isotonic regression.
2. Aim to avoid estimating tuning parameters.

Contributions of this paper

1. New approach to monotone sharp and fuzzy RDD based on the isotonic regression.
2. Aim to avoid estimating tuning parameters.
3. First treatment of the isotonic regression at the boundary of its support based on a new tightness result, cf. (Kulikov and Lopuhaä, 2006) and the KMT approximation for the Grenander estimator.

Contributions of this paper

1. New approach to monotone sharp and fuzzy RDD based on the isotonic regression.
2. Aim to avoid estimating tuning parameters.
3. First treatment of the isotonic regression at the boundary of its support based on a new tightness result, cf. (Kulikov and Lopuhaä, 2006) and the KMT approximation for the Grenander estimator.
4. New bootstrap methodology for a non-standard inference problem.

Isotonic regression

1. Nonparametric regression

$$Y = m(X) + \varepsilon, \qquad \mathbb{E}[\varepsilon|X] = 0,$$

where m is a monotone on $[0, 1]$.

Isotonic regression

1. Nonparametric regression

$$Y = m(X) + \varepsilon, \qquad \mathbb{E}[\varepsilon|X] = 0,$$

where m is a monotone on $[0,1]$.

2. Isotonic regression estimator: nonparametric least-squares over the set of non-decreasing functions

$$\hat{m}(.) = \underset{m \in \mathcal{M}[0,1]}{\arg\min} \sum_{i=1}^{n} (Y_i - m(X_i))^2.$$

Isotonic regression

1. Nonparametric regression

$$Y = m(X) + \varepsilon, \qquad \mathbb{E}[\varepsilon|X] = 0,$$

where m is a monotone on $[0,1]$.

2. Isotonic regression estimator: nonparametric least-squares over the set of non-decreasing functions

$$\hat{m}(.) = \arg\min_{m \in \mathcal{M}[0,1]} \sum_{i=1}^{n} (Y_i - m(X_i))^2.$$

3. Can be computed, e.g., using the pool adjacent violators algorithm: scales up similarly to the OLS estimator.

Isotonic regression: graphical representation

(W.T. Reid , 1955): the estimator $\hat{m}(x)$ is the left derivative of the greatest convex minorant of the cumulative sum diagram

$$t \mapsto (F_n(t), M_n(t)), \qquad t \in [0,1]$$

at $t = x$ with

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X_i \leq t\} \qquad \text{and} \qquad M_n(t) = \frac{1}{n}\sum_{i=1}^{n} Y_i \mathbb{1}\{X_i \leq t\}.$$

Isotonic regression: graphical representation

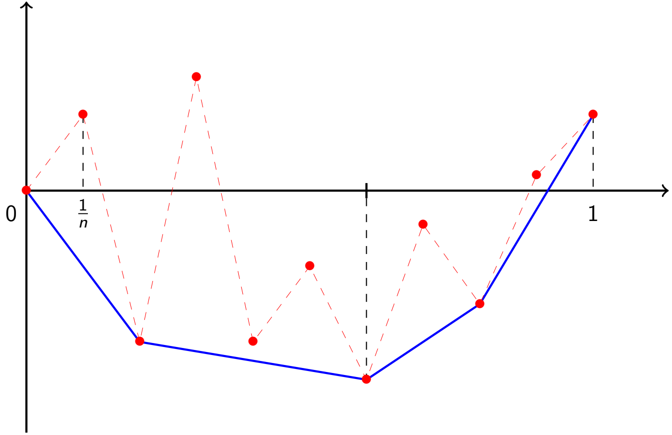


Figure: $\hat{m}(x)$ is the left derivative of the greatest convex minorant (broken blue line) of the cumulative sum diagram $t \mapsto (F_n(t), M_n(t))$ (red dots) with $t \in [0, 1]$.

Isotonic regression at the boundary: closed-form expression

Estimator of the boundary point $m(0) = \lim_{x\downarrow 0} m(x)$ is the slope of the first segment of the cumulative sum diagram

$$\hat{m}(X_{(1)}) = \min_{1\leq i\leq n} \frac{1}{i}\sum_{j=1}^{i} Y_{(j)},$$

where $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ is the order statistics and $(Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)})$ is the induced order statistics.

Isotonic regression at the boundary: inconsistency

Theorem

Suppose that $x \mapsto \Pr(Y \leq y|X = x)$ *is continuous for every* y *and that* $F_{\varepsilon|X=0}(-\epsilon) > 0$ *for some* $\epsilon > 0$. *Then*

$$\liminf_{n\to\infty} \Pr(|\hat{m}(X_{(1)}) - m(0)| > \epsilon) > 0.$$

Isotonic regression at the boundary: inconsistency

Proof.

For any $\epsilon > 0$

$$\begin{aligned}\Pr(|\hat{m}(X_{(1)}) - m(0)| > \epsilon) &\geq \Pr\left(\min_{1\leq i\leq n}\frac{1}{i}\sum_{j=1}^{i} Y_{(j)} < m(0) - \epsilon\right)\\ &\geq \Pr(Y_{(1)} < m(0) - \epsilon)\\ &= \int \Pr(Y \leq m(0) - \epsilon | X = x)\mathrm{d}F_{X_{(1)}}(x)\\ &\to \Pr(Y \leq m(0) - \epsilon | X = 0)\\ &= F_{\varepsilon|X=0}(-\epsilon),\end{aligned}$$

where we use the fact that $X_{(1)} \xrightarrow{d} 0$. □

Non-standard asymptotics

Theorem

Boundary corrected estimators $\hat{m}(cn^{-a})$ *with* $c > 0$ *and* $a \in (0, 1)$

(i) *For* $a \in (0, 1/3)$

$$n^{\frac{1}{3}} \left(\hat{m}\left(cn^{-a} \right) - m(0) \right) \xrightarrow{d} \left| \frac{4m'(0)\sigma^2(0)}{f(0)} \right|^{1/3} \operatorname{argmax}_{t \in \mathbf{R}} \{ W_t - t^2 \}.$$

(ii) *For* $a \in [1/3, 1)$

$$n^{\frac{1-a}{2}} \left(\hat{m}\left(cn^{-a} \right) - m(0) \right) \xrightarrow{d} D^L_{[0,\infty)} \left(\sqrt{\frac{\sigma^2(0)}{cf(0)}} W_t + \frac{t^2 c}{2} m'(0) \mathbb{1}_{a=1/3} \right) (1$$

where $(W_t)_{t \in \mathbf{R}}$ *is the two-sided Brownian motion,* $\sigma^2(x) = \operatorname{Var}(Y|X = x)$, $f(x)$ *is the density of* X*, and* $D^L_A(g)(x)$ *is the left derivative of the greatest convex minorant of* $g : A \to \mathbf{R}$ *at a point* $x \in A \subset \mathbf{R}$.

Ingredients of the proof

1. Switching relation (Groeneboom, 1985): for every $x \in (0,1)$ and $a \in \mathbf{R}$

$$\hat{m}(x) \leq a \iff \mathrm{argmax}_{s\in[0,1]} \{aF_n(s) - M_n(s)\} \geq x.$$

Ingredients of the proof

1. Switching relation (Groeneboom, 1985): for every $x \in (0,1)$ and $a \in \mathbf{R}$

$$\hat{m}(x) \leq a \iff \mathrm{argmax}_{s\in[0,1]}\{aF_n(s) - M_n(s)\} \geq x.$$

2. Argmax continuous mapping theorem of (Kim and Pollard, 1990): if $Z_n \xrightarrow{d} Z$ uniformly on compact sets and
   (i) $(Z(t))_{t\in\mathbf{R}}$ is a continuous stochastic process with a unique maximizer;
   (ii) $\lim_{|t|\to\infty} Z(t) = -\infty$;
   (iii) Tightness: $\mathrm{argmax}_{t\in\mathbf{R}} Z_n(t) = O_P(1)$.

   Then $\mathrm{argmax}_{t\in\mathbf{R}} Z_n(t) \xrightarrow{d} \mathrm{argmax}_{t\in\mathbf{R}} Z(t)$

Ingredients of the proof

1. Switching relation (Groeneboom, 1985): for every $x \in (0,1)$ and $a \in \mathbf{R}$

$$\hat{m}(x) \leq a \iff \mathrm{argmax}_{s\in[0,1]} \{aF_n(s) - M_n(s)\} \geq x.$$

2. Argmax continuous mapping theorem of (Kim and Pollard, 1990): if $Z_n \xrightarrow{d} Z$ uniformly on compact sets and
   (i) $(Z(t))_{t\in\mathbf{R}}$ is a continuous stochastic process with a unique maximizer;
   (ii) $\lim_{|t|\to\infty} Z(t) = -\infty$;
   (iii) Tightness: $\mathrm{argmax}_{t\in\mathbf{R}} Z_n(t) = O_P(1)$.

   Then $\mathrm{argmax}_{t\in\mathbf{R}} Z_n(t) \xrightarrow{d} \mathrm{argmax}_{t\in\mathbf{R}} Z(t)$

3. Old and new tightness results for the boundary point: (Kim and Pollard, 1990) and (van der Vaart and Wellner, 2000) results do not always apply.

Ingredients of the proof

1. Switching relation (Groeneboom, 1985): for every $x \in (0,1)$ and $a \in \mathbf{R}$

$$\hat{m}(x) \leq a \iff \mathrm{argmax}_{s\in[0,1]}\{aF_n(s) - M_n(s)\} \geq x.$$

2. Argmax continuous mapping theorem of (Kim and Pollard, 1990): if $Z_n \xrightarrow{d} Z$ uniformly on compact sets and
   (i) $(Z(t))_{t\in\mathbf{R}}$ is a continuous stochastic process with a unique maximizer;
   (ii) $\lim_{|t|\to\infty} Z(t) = -\infty$;
   (iii) Tightness: $\mathrm{argmax}_{t\in\mathbf{R}} Z_n(t) = O_P(1)$.

   Then $\mathrm{argmax}_{t\in\mathbf{R}} Z_n(t) \xrightarrow{d} \mathrm{argmax}_{t\in\mathbf{R}} Z(t)$
3. Old and new tightness results for the boundary point: (Kim and Pollard, 1990) and (van der Vaart and Wellner, 2000) results do not always apply.
4. Do not rely on the strong approximation, cf., (Kulikov and Lopuhaä, 2006).

Comments

1. "Slow" corrections: cn^{-a} with $a \in (0, 1/3)$ lead to the asymptotic distribution similar to the one at the interior point, cf., (Wright, 1981).

Comments

1. "Slow" corrections: cn^{-a} with $a \in (0, 1/3)$ lead to the asymptotic distribution similar to the one at the interior point, cf., (Wright, 1981).
2. "Slow" corrections have large finite-sample bias converging at a slower than cube-root rate $\implies$ not recommended to use in practice.

Comments

1. "Slow" corrections: cn^{-a} with $a \in (0, 1/3)$ lead to the asymptotic distribution similar to the one at the interior point, cf., (Wright, 1981).
2. "Slow" corrections have large finite-sample bias converging at a slower than cube-root rate $\implies$ not recommended to use in practice.
3. "Fast" corrections: cn^{-a} with $a \in [1/3, 1)$ generate one-sided counterpart to the distribution at the interior point.

Comments

1. "Slow" corrections: cn^{-a} with $a \in (0, 1/3)$ lead to the asymptotic distribution similar to the one at the interior point, cf., (Wright, 1981).
2. "Slow" corrections have large finite-sample bias converging at a slower than cube-root rate $\implies$ not recommended to use in practice.
3. "Fast" corrections: cn^{-a} with $a \in [1/3, 1)$ generate one-sided counterpart to the distribution at the interior point.
4. The fastest cube-root convergence rate is achieved when $a = 1/3$

$$n^{1/3}\left(\hat{m}(cn^{-1/3}) - m(0)\right) \xrightarrow{d} D^{L}_{[0,\infty)}\left(\sqrt{\frac{\sigma^2(0)}{cf(0)}}W_t + \frac{t^2 c}{2}m'(0)\right)(1).$$

Comments

1. "Slow" corrections: cn^{-a} with $a \in (0, 1/3)$ lead to the asymptotic distribution similar to the one at the interior point, cf., (Wright, 1981).
2. "Slow" corrections have large finite-sample bias converging at a slower than cube-root rate $\implies$ not recommended to use in practice.
3. "Fast" corrections: cn^{-a} with $a \in [1/3, 1)$ generate one-sided counterpart to the distribution at the interior point.
4. The fastest cube-root convergence rate is achieved when $a = 1/3$

$$n^{1/3}\left(\hat{m}(cn^{-1/3}) - m(0)\right) \xrightarrow{d} D^L_{[0,\infty)}\left(\sqrt{\frac{\sigma^2(0)}{cf(0)}}W_t + \frac{t^2 c}{2}m'(0)\right) (1).$$

5. Minimax optimal convergence rate under the assumption m' exists.

Comments

1. "Slow" corrections: cn^{-a} with $a \in (0, 1/3)$ lead to the asymptotic distribution similar to the one at the interior point, cf., (Wright, 1981).
2. "Slow" corrections have large finite-sample bias converging at a slower than cube-root rate $\implies$ not recommended to use in practice.
3. "Fast" corrections: cn^{-a} with $a \in [1/3, 1)$ generate one-sided counterpart to the distribution at the interior point.
4. The fastest cube-root convergence rate is achieved when $a = 1/3$

$$n^{1/3}\left(\hat{m}(cn^{-1/3}) - m(0)\right) \xrightarrow{d} D^L_{[0,\infty)}\left(\sqrt{\frac{\sigma^2(0)}{cf(0)}}W_t + \frac{t^2 c}{2}m'(0)\right) (1).$$

5. Minimax optimal convergence rate under the assumption m' exists.
6. The distribution is not pivotal.

What about the constant c?

1. Interior point $x \in (0, 1)$

$$n^{1/3}\left(\hat{m}(x) - m(x)\right) \xrightarrow{d} D^{L}_{(-\infty,\infty)}\left(\sqrt{\frac{\sigma^2(0)}{f(0)}}W_t + \frac{t^2}{2}m'(0)\right)(1).$$

What about the constant c?

1. Interior point $x \in (0,1)$

$$n^{1/3}\left(\hat{m}(x) - m(x)\right) \xrightarrow{d} D^L_{(-\infty,\infty)}\left(\sqrt{\frac{\sigma^2(0)}{f(0)}}W_t + \frac{t^2}{2}m'(0)\right)(1).$$

2. Boundary point $x = 0$

$$n^{1/3}\left(\hat{m}(cn^{-1/3}) - m(x)\right) \xrightarrow{d} D^L_{[0,\infty)}\left(\sqrt{\frac{\sigma^2(0)}{cf(0)}}W_t + \frac{t^2 c}{2}m'(0)\right)(1).$$

What about the constant c?

1. Interior point $x \in (0,1)$

$$n^{1/3}\left(\hat{m}(x) - m(x)\right) \xrightarrow{d} D^L_{(-\infty,\infty)}\left(\sqrt{\frac{\sigma^2(0)}{f(0)}}W_t + \frac{t^2}{2}m'(0)\right)(1).$$

2. Boundary point $x = 0$

$$n^{1/3}\left(\hat{m}(cn^{-1/3}) - m(x)\right) \xrightarrow{d} D^L_{[0,\infty)}\left(\sqrt{\frac{\sigma^2(0)}{cf(0)}}W_t + \frac{t^2c}{2}m'(0)\right)(1).$$

3. We get $c = 1$ automatically for the tuning-free isotonic regression at the interior point.

What about the constant c?

1. Interior point $x \in (0,1)$

$$n^{1/3}\left(\hat{m}(x) - m(x)\right) \xrightarrow{d} D^L_{(-\infty,\infty)}\left(\sqrt{\frac{\sigma^2(0)}{f(0)}}W_t + \frac{t^2}{2}m'(0)\right)(1).$$

2. Boundary point $x = 0$

$$n^{1/3}\left(\hat{m}(cn^{-1/3}) - m(x)\right) \xrightarrow{d} D^L_{[0,\infty)}\left(\sqrt{\frac{\sigma^2(0)}{cf(0)}}W_t + \frac{t^2 c}{2}m'(0)\right)(1).$$

3. We get $c = 1$ automatically for the tuning-free isotonic regression at the interior point.
4. (not recommended) Alternative is to estimate the constant:
   - Increasing the variance with a hope to reduce the bias and the asymptotic MSE: the finite-sample MSE increases in our MC experiments, see also (Kulikov and Lopuhaä, 2006) for the Grenander estimator;
   - Inference after the model selection problem?

Non-standard inferences: the bootstrap?

$$n^{1/3}\left(\hat{m}(n^{-1/3}) - m(0)\right) \xrightarrow{d} D^{L}_{[0,\infty)}\left(\sqrt{\frac{\sigma^2(0)}{f(0)}}W_t + \frac{t^2}{2}m'(0)\right) \quad (1)$$

1. The distribution is not pivotal: estimating σ^2, f, m' and discretizing the time?

Non-standard inferences: the bootstrap?

$$n^{1/3}\left(\hat{m}(n^{-1/3}) - m(0)\right) \xrightarrow{d} D^L_{[0,\infty)}\left(\sqrt{\frac{\sigma^2(0)}{f(0)}}W_t + \frac{t^2}{2}m'(0)\right) \quad (1)$$

1. The distribution is not pivotal: estimating σ^2, f, m' and discretizing the time?
2. The bootstrap fails for cube-root consistent estimators as they are not smooth functions of the data: Isotonic regression, Manski's maximum score, Grenander estimator, current status model...

Non-standard inferences: the bootstrap?

$$n^{1/3}\left(\hat{m}(n^{-1/3}) - m(0)\right) \xrightarrow{d} D^L_{[0,\infty)}\left(\sqrt{\frac{\sigma^2(0)}{f(0)}}W_t + \frac{t^2}{2}m'(0)\right) \quad (1)$$

1. The distribution is not pivotal: estimating σ^2, f, m' and discretizing the time?
2. The bootstrap fails for cube-root consistent estimators as they are not smooth functions of the data: Isotonic regression, Manski's maximum score, Grenander estimator, current status model...
3. The bootstrap does not estimate consistently m' for the isotonic regression.

Non-standard inferences: the bootstrap?

$$n^{1/3}\left(\hat{m}(n^{-1/3}) - m(0)\right) \xrightarrow{d} D^{L}_{[0,\infty)}\left(\sqrt{\frac{\sigma^2(0)}{f(0)}}W_t + \frac{t^2}{2}m'(0)\right) \quad (1)$$

1. The distribution is not pivotal: estimating σ^2, f, m' and discretizing the time?
2. The bootstrap fails for cube-root consistent estimators as they are not smooth functions of the data: Isotonic regression, Manski's maximum score, Grenander estimator, current status model...
3. The bootstrap does not estimate consistently m' for the isotonic regression.
4. Available solutions: subsampling, smoothed bootstrap, reshaping the objective function (Cattaneo, Jansson, and Nagasawa, 2019).

New solution

1. Using $\hat{m}(n^{-1/2})$ instead of $\hat{m}(n^{-1/3})$ and killing the drift term with m'.

New solution

1. Using $\hat{m}(n^{-1/2})$ instead of $\hat{m}(n^{-1/3})$ and killing the drift term with m'.
2. $n^{-1/2}$ balances the convergence rate of the estimator and the rate at which the drift vanishes.

Trimmed wild bootstrap

1. Simulate wild bootstrap samples

$$Y_i^* = \tilde{m}(X_i) + \eta_i^* \tilde{\varepsilon}_i, \qquad i = 1, \dots, n;$$

2. $(\eta_i^*)_{i=1}^n$ are i.i.d., independent from the data;
3. Trimmed isotonic regression estimator

$$\tilde{m}(x) = \begin{cases} \hat{m}(x), & x \in (n^{-1/2}, 1) \\ \hat{m}(n^{-1/2}), & x \in [0, n^{-1/2}] \end{cases}$$

   and $(\tilde{\varepsilon}_i)_{i=1}^n$ are corresponding residuals.
4. Under some regularity conditions, the trimmed wild bootstrap is consistent in probability.

Isotonic regression discontinuity design estimators

Sharp iRDD estimator

$$\hat{\theta} = \hat{m}_+(n^{-a}) - \hat{m}_-(n^{-a}),$$

where we run two isotonic regressions

$$\hat{m}_-(.) = \operatorname*{arg\,min}_{m \in \mathcal{M}[-1,0)} \sum_{i \in I_-} (Y_i - m(X_i))^2, \quad \hat{m}_+(.) = \operatorname*{arg\,min}_{m \in \mathcal{M}[0,1]} \sum_{i \in I_+} (Y_i - m(X_i))^2.$$

and

- $a = 1/3$ for point estimation;
- $a = 1/2$ for inferences;

Asymptotic distribution

Theorem

Under some regularity conditions

$$n^{1/3}(\hat{\theta} - \theta) \xrightarrow{d} \xi_+ - \xi_-,$$

where

$$\xi_+ = D^L_{[0,\infty)}\left(\sqrt{\frac{\sigma_+^2}{f_+}}W_t^+ + \frac{t^2}{2}m'_+\right)(1)$$

$$\xi_- = D^L_{(-\infty,0]}\left(\sqrt{\frac{\sigma_-^2}{f_-}}W_t^- + \frac{t^2}{2}m'_-\right)(-1).$$

and W_t^+ and W_t^- are two independent standard Brownian motions originating from zero and running in opposite directions.

Bootstrap consistency

Theorem

Under some regularity conditions

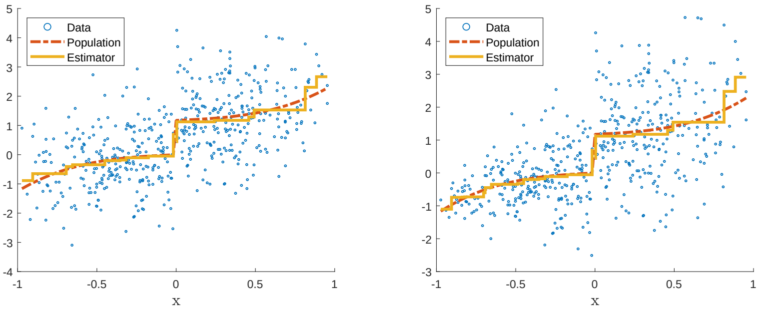
$$\left|\Pr^*\left(n^{1/4}(\hat{\theta}^* - \hat{\theta}) \leq u\right) - \Pr\left(n^{1/4}(\hat{\theta} - \theta) \leq u\right)\right| \xrightarrow{P} 0,$$

where $\Pr^*(.) = \Pr(.|(X_i, Y_i)_{i=1}^{\infty})$.

MC experiments: design

1. DGP:
$$Y = m(X) + \theta \mathbb{1}_{[0,1]}(X) + \sigma(X)\varepsilon,$$
where $\varepsilon \sim N(0,1)$ and $\varepsilon \perp\!\!\!\perp X$.
2. Two specifications: $m(x) = x^3 + 0.25x$ (DGP3) or $m(x) = \exp(0.25x)$ (DGP2).
3. Homoskedasticity ($\sigma(x) = 1$) and heteroskedasticity ($\sigma(x) = \sqrt{x+1}$).
4. $X \sim 2 \times \mathrm{Beta}(\alpha, \beta) - 1$ with low density near the cut-off ($\alpha = \beta = 0.5$, DGP2) and high density near the cut-off ($\alpha = \beta = 2$, DGP1,3).
5. Causal effect $\theta = 1$.
6. 5,000 replications.

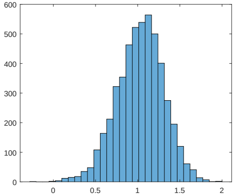
MC experiments: single run



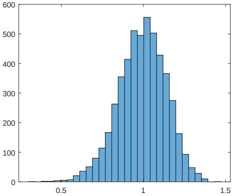
(a) Homoskedasticity (b) Heteroskedasticity

Figure: Single MC experiment, $n = 500$.

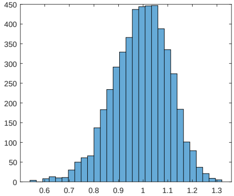
MC experiments: finite sample distribution



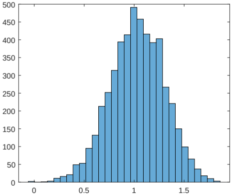
(a) $n = 100$



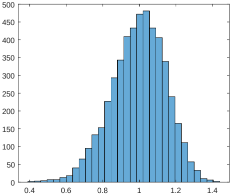
(b) $n = 500$



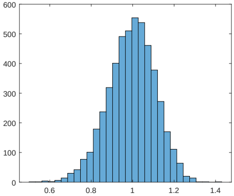
(c) $n = 1000$



(d) $n = 100$



(e) $n = 500$



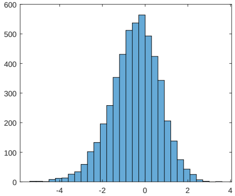
(f) $n = 1000$

Figure: Homoskedasticity in (a)-(c) and heteroskedasticity in (d)-(f)

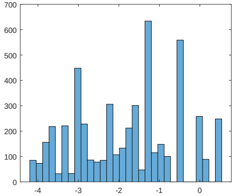
MC experiments: finite sample distribution

	n	Homoskedasticity			Heteroskedasticity		
		Bias	Var	MSE	Bias	Var	MSE
DGP 1							
	100	0.020	0.077	0.077	0.027	0.077	0.078
	500	-0.008	0.022	0.022	-0.006	0.022	0.022
	1000	-0.006	0.013	0.013	-0.005	0.013	0.013
DGP 2							
	100	-0.153	0.137	0.160	-0.138	0.141	0.160
	500	-0.081	0.044	0.050	-0.077	0.045	0.050
	1000	-0.063	0.027	0.031	-0.060	0.027	0.031
DGP 3							
	100	0.093	0.089	0.097	0.098	0.090	0.099
	500	0.017	0.024	0.024	0.018	0.024	0.024
	1000	0.006	0.015	0.015	0.007	0.015	0.015

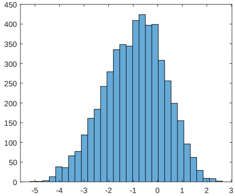
MC experiments: exact distribution vs the bootstrap



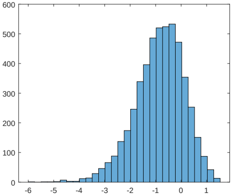
(a) Exact distribution



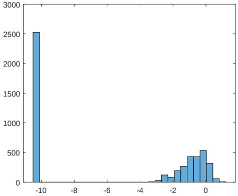
(b) Naive bootstrap



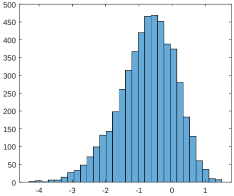
(c) Trimmed bootstrap



(d) Exact distribution



(e) Naive bootstrap



(f) Trimmed bootstrap

Figure: Sample size: $n = 100$ in panels (a)-(c) and $n = 1,000$ in panels (d)-(f).

Incumbency advantage (Lee, 2008)

1. Causal effect of incumbency on electoral outcomes: incumbents by definition are more successful politicians.

Incumbency advantage (Lee, 2008)

1. Causal effect of incumbency on electoral outcomes: incumbents by definition are more successful politicians.
2. (Lee, 2008): 7.7% incumbency advantage for the U.S. Congressional elections.

Incumbency advantage (Lee, 2008)

1. Causal effect of incumbency on electoral outcomes: incumbents by definition are more successful politicians.
2. (Lee, 2008): 7.7% incumbency advantage for the U.S. Congressional elections.
3. Monotonicity is plausible: candidates with a larger margin have on average larger vote share on the next election.

Incumbency advantage (Lee, 2008)

1. Causal effect of incumbency on electoral outcomes: incumbents by definition are more successful politicians.
2. (Lee, 2008): 7.7% incumbency advantage for the U.S. Congressional elections.
3. Monotonicity is plausible: candidates with a larger margin have on average larger vote share on the next election.
4. iRDD gives point estimates 13.8% with 95% confidence interval $[6.6\%, 26.5\%]$.

Incumbency advantage (Lee, 2008)

1. Causal effect of incumbency on electoral outcomes: incumbents by definition are more successful politicians.
2. (Lee, 2008): 7.7% incumbency advantage for the U.S. Congressional elections.
3. Monotonicity is plausible: candidates with a larger margin have on average larger vote share on the next election.
4. iRDD gives point estimates 13.8% with 95% confidence interval $[6.6\%, 26.5\%]$.
5. Without boundary corrections (iRDD is inconsistent) the point estimate is 6.6%.

Incumbency advantages

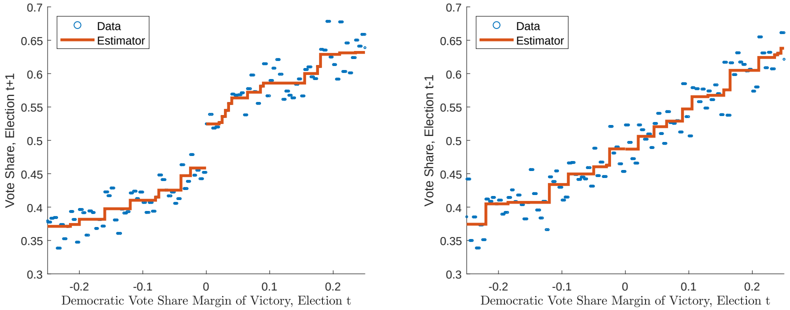


Figure: Incumbency advantage. Sample size: 6,559 observations with 3,819 observations below the cut-off.

Conclusions

1. New approach to nonparametric monotone RD designs;
2. Theory for the isotonic regression estimator at the boundary of its support based on new tightness results;
3. New wild bootstrap method that works without additional nonparametric smoothing (or subsampling);
4. Inference with valid standard errors.

Thank you!

email: babii.andrii@gmail.com