Isotonic regression discontinuity designs

Andrii Babii* Rohit Kumar†
UNC Chapel Hill *ISI, Delhi*

September 14, 2020

Abstract

This paper studies the estimation and inference in the isotonic regression at the boundary of its support, an object that is particularly interesting and required in the analysis of monotone regression discontinuity designs. We show that the isotonic regression is inconsistent at the boundary and that consistency can be restored with a suitable boundary correction. The large sample distribution at the boundary is driven by the one-sided Brownian motion. Since the distribution is not pivotal, we also introduce the trimmed wild bootstrap and show its consistency without subsampling or additional nonparametric smoothing. The results are illustrated for sharp and fuzzy monotone regression discontinuity designs. We find in Monte Carlo experiments that shape restrictions can improve dramatically the finite-sample performance of unrestricted estimators. An empirical analysis of the incumbency effect in the U.S. House elections is provided.

Keywords: isotonic regression, boundary point, wild bootstrap, regression discontinuity design, monotonicity, shape constraints.

JEL classification: C14, C31.

1 Introduction

Regression discontinuity designs, see Thistlethwaite and Campbell (1960), are widely recognized as one of the most credible quasi-experimental strategies for identifying

*Department of Economics, University of North Carolina at Chapel Hill - Gardner Hall, CB 3305 Chapel Hill, NC 27599-3305. Corresponding author. Email: babii.andrii@gmail.com.

†Indian Statistical Institute, Delhi, 7 S.J.S. Sansanwal Marg, ISI Delhi Centre, New Delhi, Delhi 110016.

and estimating causal effects. In a nutshell, such designs exploit the discontinuity in the treatment assignment probability around a certain cutoff value of some covariate. The discontinuous treatment assignment probability frequently occurs due to laws and regulations governing economic and political life. A comprehensive list of empirical applications using regression discontinuity designs can be found in Lee and Lemieux (2010); see also Imbens and Lemieux (2008), and Cattaneo, Idrobo, and Titiunik (2019) for the methodological review and Imbens and Wooldridge (2009), Abadie and Cattaneo (2018) for their place in the program evaluation literature. On the methodological side, in the seminal paper Hahn, Todd, and Van der Klaauw (2001) translate regression discontinuity designs into the potential outcomes framework and introduce the local polynomial nonparametric estimator to estimate causal effects in sharp and fuzzy designs.

Regression discontinuity designs encountered in the empirical practice are frequently monotone. Indeed, development and educational programs are often prescribed based on poverty or achievement scores that are monotonically related to average outcomes. For instance, when evaluating the effect of subsidies for fertilizers on yields, the yield per acre is expected to be non-decreasing in the size of the farm due to the increasing returns to scale. Alternatively, when evaluating the effectiveness of the cash transfers program on the households food expenditures, we expect that more affluent households spend, on average, more on food, since food is a normal good; see Appendix A.5 for a sample of other examples drawn from the empirical research.

Despite this prevalence in the empirical practice, little is known about how monotonicity can be incorporated in the estimation procedure.[1] Estimation and inference in regression discontinuity designs require estimating the conditional mean functions at the boundary of the support. In this paper, we focus on the isotonic regression estimator and establish its formal statistical properties for the boundary point.[2] The isotonic regression estimator originates from the work of Ayer, Brunk, Ewing, Reid, and Silverman (1955), Brunk (1956), and van Eeden (1958). Brunk (1970) derives its asymptotic distribution at the *interior point* under restrictive assumptions that the regressor is deterministic and regression errors are homoskedastic. His treatment builds upon the ideas of Rao (1969), who derived the asymptotic distribution of the

[1] Shape restrictions in regression discontinuity designs is a relatively unexplored area. To the best of our knowledge, Armstrong (2015) is the only existing work that focuses on the k-nearest neighbors estimator.

[2] This estimator is relatively unknown in econometrics and to the best of our knowledge has not been previously considered in the RDD setting; see an excellent review paper Chetverikov, Santos, and Shaikh (2018) for a comprehensive discussion of this estimator and further references.

monotone density estimator, also known as the Grenander estimator; see Grenander (1956). Wright (1981) provides the final characterization of the large sample distribution for the interior point when the regressor is random and regression errors are heteroskedastic.

However, to the best of our knowledge, little is known about the behavior of the isotonic regression at the *boundary point*, which is a building block of our isotonic regression discontinuity (iRDD) estimators. This situation contrasts strikingly with the local polynomial estimator, whose boundary behavior is well-understood; see Fan and Gijbels (1992). Most of the existing results for isotonic estimators at the boundary are available only for the Grenander estimator; see Woodroofe and Sun (1993), Kulikov and Lopuhaä (2006), and Balabdaoui, Jankowski, Pavlides, Seregin, and Wellner (2011). More precisely, we know that the Grenander estimator is inconsistent at the boundary of the support and that the consistent estimator can be obtained with additional boundary correction or penalization. At the same time, some isotonic estimators, e.g., in the current status model, are consistent at the boundary without corrections; see Durot and Lopuhaä (2018). Anevski and Hössjer (2002) discuss the inconsistency of the isotonic regression at the discontinuity point with deterministic equally spaced covariate and homoskedasticity. However, Anevski and Hössjer (2002) do not discuss whether the isotonic regression with random covariate is consistent at the boundary of its support and do not provide a consistent estimator even in the restrictive equally spaced fixed design case.

In this paper, we aim to understand the behavior of the isotonic regression with a random regressor at the boundary of its support. We show formally that the isotonic regression estimator is inconsistent and tends to underestimate the value of the regression at the left boundary. The inconsistency is related to the extreme-value behavior of the closest to the boundary observation. We introduce boundary-corrected estimators and derive large sample approximations to corresponding distributions. The major technical difficulty when deriving asymptotic distributions in this setting is to establish the tightness of the maximizer of a certain empirical process. This condition is typically needed to apply the argmax continuous mapping theorem of Kim and Pollard (1990). The difficulty stems from the fact that conventional tightness results of Kim and Pollard (1990) and van der Vaart and Wellner (2000) do not always apply. For the Grenander estimator, Kulikov and Lopuhaä (2006) suggest a solution to this problem based on the Komlós-Major-Tusnády strong approximation. In our setting, this approach entails the strong approximation to the general empirical process, see Koltchinskii (1994) and Chernozhukov, Newey, and Santos (2015), which is more problematic to apply due to slower convergence rates. Consequently, we provide the alternative generic proof which does not rely on the strong approximation and which might be applied to other boundary-corrected shape-constrained estimators.

Since the asymptotic distribution is not pivotal, we introduce a novel trimmed wild bootstrap procedure and establish its consistency. The procedure consists of trimming values of the estimated regression function that are very close to the boundary when simulating wild bootstrap samples. Somewhat unexpectedly, we discover that the trimming and the appropriate boundary correction restores the consistency of the wild bootstrap without additional nonparametric smoothing or subsampling, which is typically needed in such settings.[3]

The paper is organized as follows. In Section 2, we develop the asymptotic theory for the isotonic regression estimator at the boundary point. The reader interested more in the regression discontinuity designs can skip this section and go directly to Section 3, where we discuss isotonic sharp and fuzzy designs. In Section 4, we study the finite sample performance of the iRDD estimator with Monte Carlo experiments. Section 5 estimates the effect of incumbency using the sharp iRDD on the data of Lee (2008). Section 6 concludes. Additional results, proofs, and Monte Carlo experiments appear in the Appendix.

2 Isotonic regression at the boundary

In this section, we study the isotonic regression estimator at the boundary point. Readers interested in regression discontinuity designs only may skip this section and go directly to Section 3.

2.1 Estimator

We focus on the generic nonparametric regression

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

where the conditional mean function $m(x) = \mathbb{E}[Y|X = x]$ is assumed to be montone. For simplicity of presentation, we normalize the support to $[0, 1]$ and assume that m belongs to the set of non-decreasing functions on $[0, 1]$

$$\mathcal{M}[0, 1] = \{\phi : [0, 1] \to \mathbf{R} : \ \phi(x_1) \leq \phi(x_2), \ \forall x_1 \leq x_2\} .$$

[3] In contrast, the bootstrap typically fails at the interior point; see Kosorok (2008a), Sen, Banerjee, and Woodroofe (2010), Guntuboyina and Sen (2018), and Patra, Seijo, and Sen (2018) for the discussion of this problem and various case-specific remedies, and Cattaneo, Jansson, and Nagasawa (2020) for generic solutions that apply to a class of cube-root consistent estimators.

For a sample $(Y_i, X_i)_{i=1}^n$, the isotonic regression estimator, denoted $\hat{m}$, solves the constrained least-squares problem

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

see Brunk (1970). Note that the estimator is uniquely determined at data points and is conventionally interpolated as a piecewise constant function elsewhere. Let $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ be the ordered values of the covariate and let $(Y_{(1)}, Y_{(2)} \ldots, Y_{(n)})$ be the corresponding observations of the outcome variable. It is easy to see that the isotonic regression is alternatively characterized as a solution to

$$\min_{\phi_1 \leq \phi_2 \leq \cdots \leq \phi_n} \sum_{i=1}^{n} (Y_{(i)} - \phi_i)^2.$$

While, the isotonic regression is a solution to the convex optimization problem and can be obtained with standard constrained convex optimization packages, an efficient way to compute the estimator is via the pool adjacent violators algorithm; see Ayer, Brunk, Ewing, Reid, and Silverman (1955). Although the isotonic regression features the number of estimated parameters of the same magnitude as the sample size, the pool adjacent violators algorithm is typically computationally cheaper than nonparametric kernel estimators, and its computational complexity is closer to that of the OLS estimator.

2.2 Large sample distribution

We are interested in estimating the value of the regression function at the boundaries of its support $[0, 1]$. More precisely, we focus on the regression at the left boundary, denoted $m(0) = \lim_{x \downarrow 0} m(x)$. The natural estimator of $m(0)$ is $\hat{m}(X_{(1)})$. Unfortunately, it follows from Theorem A.1 in the appendix that this estimator is inconsistent. The inconsistency occurs because $X_{(1)}$ converges to zero too fast according to laws of the extreme value theory. More precisely, the estimator $\hat{m}(X_{(1)})$ underestimates $m(0)$ and $\hat{m}(X_{(n)})$ overestimates $m(1)$ in small samples; see Section A.1 in the Appendix for more details.[4]

To estimate $m(0)$ consistently, we the consider boundary-corrected estimators $\hat{m}(cn^{-a})$ for some $c, a > 0$ and study the large sample behavior of this estimator for

[4] We are grateful to Matt Masten for this observation. More generally, in the isotonic regression discontinuity design, the isotonic regression estimator underestimates the causal effect at the cutoff, which can be used as a tuning-free lower bound on the estimated causal effect.

a range of values $c, a > 0$. The following assumption imposes a mild restriction on the distribution of the data.

Assumption 2.1. $(Y_i, X_i)_{i=1}^n$ *is an i.i.d. sample of* (Y, X) *such that (i)* $\mathbb{E}[|\varepsilon|^{2+\delta}|X] \leq C < \infty$ *for some* $\delta > 0$ *and* m *is uniformly bounded; (ii) the distribution of* X *has a Lebesgue density* f*, uniformly bounded away from zero and infinity, and* $f(0) = \lim_{x\downarrow 0} f(x)$ *exists; (iii) the conditional variance* $\sigma^2(x) = \mathrm{Var}(Y|X = x)$ *is uniformly bounded and* $\sigma^2(0) = \lim_{x\downarrow 0} \sigma^2(x)$ *exists; (iv)* m *is continuously differentiable in the neighborhood of zero with* $m'(0) = \lim_{x\downarrow 0} m'(x) > 0$*; (v)* $m \in \mathcal{M}[0, 1]$*.*

For a stochastic process $\{Z_t : t \in A \subset \mathbf{R}\}$, let $D_A^L(Z_t)(s)$ denote the left derivative of the greatest convex minorant at $s \in A$.[5] We say that m is γ-Hölder continuous in the neighborhood of zero if there exists a constant $C < \infty$ such that for all $x > 0$ in this neighborhood

$$|m(x) - m(0)| \leq C|x|^{\gamma}.$$

Our first result describes the large sample behavior of boundary corrected isotonic regression estimators for a range of boundary corrections.

Theorem 2.1. *Suppose that Assumption 2.1 is satisfied and that* $c > 0$*. Then*

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

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

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

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

where $(W_t)_{t\in\mathbf{R}}$ *is a two-sided Brownian motion,*[6] *and for* $a \in (1/3, 1)$*, we can replace Assumption 2.1 (iv) by the* γ*-Höder continuity with* $\gamma > (1 - a)/2a$*.*

The proof of this result can be found in the appendix. The most challenging part of the proof is establishing tightness when $a \in (1/3, 1)$. The difficulty comes

[5] The greatest convex minorant of a function $g : A \to \mathbf{R}$ is defined as the maximal convex function h such that $h(x) \leq g(x), \forall x \in A$.

[6] The two-sided Brownian motion is defined as two independent Brownian motions originating from zero and moving in the opposite directions.

from the fact that the quadratic term vanishes in this asymptotic regime and the standard tightness results for isotonic estimators do not apply.[7] For the Grenander estimator, Kulikov and Lopuhaä (2006) suggest a solution to this problem based on the Komlós-Major-Tusnády strong approximation. In our case, we would need to rely on the strong approximation to the generic empirical process, see Koltchinskii (1994), which leads to suboptimal results due to slower convergence rates and uniform boundedness restrictions. Our proof does not rely on the strong approximation and is based on a more standard partitioning argument.

Theorem 2.1 shows that for "slow" boundary corrections with $a \in (0, 1/3)$, the asymptotic distribution coincides with the one at the interior point, cf., Wright (1981). However, since the estimator is centered at $m(cn^{-a})$, it is inconsistent for $m(0)$ because the "bias" $n^{1/3}(m(cn^{-a}) - m(0))$ does not vanish for all $a \in (0, 1/3)$.

The choice $a = 1/3$ leads to the cube-root consistent estimator similar to the tuning-free isotonic regression estimator at the interior point

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

For the interior point $x \in (0, 1)$, Wright (1981) shows that

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

This suggests that the boundary correction with $a = 1/3$ and $c = 1$ should behave similarly to the tuning-free isotonic regression estimator at the interior point. Alternatively, one can try to improve the point estimates with a data-driven MSE-optimal choice which we investigate in Section 4.

For the "fast" boundary corrections with $a \in (1/3, 1)$, the convergence rate is slower than $O(n^{-1/3})$, which comes with a benefit of relaxing the smoothness requirement for m. For instance, for $a = 1/2$, we only need the γ-Hölder smoothness with $\gamma > 1/2$ and do not require that m' exists, in which case, the convergence rate is $O(n^{-1/4})$. Due to the slower than the cube-root convergence rate, we do not recommend the "fast" boundary corrections for the point estimation, unless one expects that m is not differentiable in the neighborhood of zero. However, the "fast" corrections are useful when it comes to inference, because as we shall show in the following section, the bootstrap works without additional subsampling or nonparametric smoothing.

[7] See Kim and Pollard (1990) and van der Vaart and Wellner (2000), Theorem 3.2.5.

Remark 2.1. *One can show that for a non-decreasing function* $m : [-1, 0] \to \mathbf{R}$ *and* $a \in (0, 1/3)$ *the asymptotic distribution of* $\hat{m}(-cn^{-a})$ *is the same as in Theorem 3.1, while for* $a \in [1/3, 1)$

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

where $f(0), m(0)$*, and* $\sigma^2(0)$ *are defined as the left limits.*

2.3 Trimmed wild bootstrap

It is well-known that the bootstrap fails for various isotonic estimators at the interior point.[8] Several resampling schemes are available in the literature, including the smoothed nonparametric or m-out-of-n bootstrap, see Sen, Banerjee, and Woodroofe (2010) and Patra, Seijo, and Sen (2018); reshaping the objective function, see Cattaneo, Jansson, and Nagasawa (2020); and smoothed residual bootstrap, see Guntuboyina and Sen (2018). Interestingly, as we shall show below, for the boundary point, the "fast" boundary corrections restore the consistency of the bootstrap. Consequently, we focus on more conventional bootstrap inferences.

The wild bootstrap, see Wu (1986) and Liu (1988), is arguably the most natural resampling scheme for the nonparametric regression. Unlike the naive nonparametric bootstrap, the wild bootstrap imposes the structure of the nonparametric regression model in the bootstrap world, so we may expect it to work better in finite samples than resampling methods that do not incorporate such information. At the same time, unlike the residual bootstrap, it does not rule out the heteroskedasticity.

The bootstrap procedure is as follows. First, we obtain the isotonic regression estimator $\hat{m}$, construct the trimmed estimator[9]

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

and compute residuals $\tilde{\varepsilon}_i = Y_i - \tilde{m}(X_i)$ for $1 \leq i \leq n$. Second, we construct the wild bootstrap samples as follows:

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

[8]See Kosorok (2008a) and Sen, Banerjee, and Woodroofe (2010) for a formal statement and Delgado, Rodríguez-Poo, and Wolf (2001), Léger and MacGibbon (2006), an Abrevaya and Huang (2005) for earlier evidences.

[9]Note that the estimator $\tilde{m}$ depends on c and a and that for the sake of simplicity of presentation we suppress this dependence.

where $(\eta_i^*)_{i=1}^n$ are i.i.d. random variables, independent of $(Y_i, X_i)_{i=1}^n$, and such that $\mathbb{E}\eta_i^* = 0$, $\mathrm{Var}(\eta_i^*) = 1$, and $\mathbb{E}|\eta_i^*|^{2+\delta} < \infty$.

Let $\Pr^*(.) = \Pr(.|(Y_i, X_i)_{i=1}^\infty)$ denote the bootstrap probability conditionally on the data, and let $\hat{m}^*$ be the isotonic regression estimator computed from the bootstrapped sample $(Y_i^*, X_i)_{i=1}^n$. The following result establishes the consistency of the trimmed wild bootstrap procedure.

Theorem 2.2. *Suppose that Assumption 2.1 is satisfied. Suppose additionally that (i) $\mathbb{E}[\varepsilon^4|X] \leq C$ for some $C < \infty$ and m is continuous; (iv) m is γ-Hölder continuous with $\gamma > (1-a)/2a$. Then for every $u \in \mathbf{R}$ and $a \in (1/3, 1)$*

$$\left|\Pr^*\left(n^{\frac{1-a}{2}}\left(\hat{m}^*(cn^{-a}) - \hat{m}(cn^{-a})\right) \leq u\right) - \Pr^*\left(n^{\frac{1-a}{2}}\left(\hat{m}(cn^{-a}) - m(0)\right) \leq u\right)\right| \xrightarrow{P} 0.$$

Theorem 2.2 shows that the boundary correction alone restores the consistency of the bootstrap. The reason is that for "fast" corrections the quadratic term $m'(0)$ disappears from the asymptotic distribution and this is precisely the term that is not consistently estimated by the bootstrap. Note that the bootstrap is consistent for every $a \in (1/3, 1)$. However, there is a trade-off between the consistency rate of the bootstrapped estimator $O(n^{(a-1)/2})$ and the rate $O(n^{(1-3a)/2})$ at which the quadratic term disappears; see the proof of Theorem 2.1. This trade-off should translate into the trade-off between the rates at which the confidence intervals shrink and the rate at which coverage errors vanish. Therefore, it seems reasonable to take $a = 1/2$ as a rule-of-thumb, which leads to the same $O(n^{-1/4})$ rate balancing the two rates.[10] Note that the slower $O(n^{-1/4})$ convergence rate is the price to pay for making the bootstrap work. However, the slower rate also comes with a benefit since this result does not require the existence of m' and relies only on the γ-Hölder continuity with smoothness index $\gamma > 1/2$.

Remark 2.2. *It is possible to show that for the non-decreasing function $m : [-1, 0] \to \mathbf{R}$ and $a \in (1/3, 1)$, the bootstrap is consistent in probability.*

3 Isotonic regression discontinuity designs

In this section, we apply our results on the isotonic regression at the boundary to monotone regression discontinuity designs.

[10] The inference-optimal choice of c is more difficult. One can still use the rule-of-thumb $c = 1$. We find in Monte Carlo experiments that there is some scope for improvements over this default choice.

3.1 Setup

Following Hahn, Todd, and Van der Klaauw (2001), we focus on the potential outcomes framework

$$Y = Y_1 D + Y_0(1 - D),$$

where $D \in \{0,1\}$ is a binary treatment indicator (1 if treated and 0 otherwise), $Y_1, Y_0 \in \mathbf{R}$ are unobserved potential outcomes of treated and untreated units, and Y is the actual observed outcome.

The causal parameter of interest is the average treatment effect at the cutoff $x_0 \in \mathbf{R}$ of some running variable $X \in \mathbf{R}$, denoted

$$\theta = \mathbb{E}[Y_1 - Y_0 | X = x_0].$$

Without further assumptions, θ is not identified in the sense that it depends on the distribution of unobserved potential outcomes (Y_0, Y_1); see Holland (1986).[11]

Hahn, Todd, and Van der Klaauw (2001) shows formally that the causal effect is identified as

$$\theta = \lim_{x \downarrow x_0} \mathbb{E}[Y|X = x] - \lim_{x \uparrow x_0} \mathbb{E}[Y|X = x], \tag{1}$$

while for fuzzy designs, the causal effect is identified as[12]

$$\theta = \frac{\lim_{x \downarrow x_0} \mathbb{E}[Y|X = x] - \lim_{x \uparrow x_0} \mathbb{E}[Y|X = x]}{\lim_{x \downarrow x_0} \Pr(D = 1|X = x) - \lim_{x \uparrow x_0} \Pr(D = 1|X = x)}; \tag{2}$$

see Proposition A.4.1 in the Appendix for a statement of this result under very weak one-sided continuity of conditional means of potential outcomes.

3.2 Estimators

We are interested in estimating the average causal effect θ of a binary treatment D on some outcome Y. For a unit i, with $1 \leq i \leq n$, we observe (Y_i, D_i, X_i). Assuming that θ is identified from the distribution of (Y, D, X) according to Eq. 1, the estimator of θ is obtained by plugging-in the corresponding boundary corrected isotonic regression estimators before and after the cutoff.

[11] While regression discontinuity designs identify local effects, there exist several approaches that aim to extrapolate away from the cutoff; see Bertanha (2020), Angrist and Rokkanen (2015), and other references therein. Note that if the causal effect changes monotonically in the running variable, one can obtain bounds on the causal effect between the two points.

[12] Identification in regression discontinuity designs might be testable; see McCrary (2008), Cattaneo, Jansson, and Ma (2019), Bugni and Canay (2020), Canay and Kamat (2017).

Isotonic regression discontinuity designs (iRDD) exploit the monotonicity of the expected outcome $m(x) = \mathbb{E}[Y|X = x]$ and the treatment assignment probability $p(x) = \Pr(D = 1|X = x)$. For concreteness, assume that both functions are non-decreasing. We also assume that X has compact support $[-1, 1]$ and normalize the cutoff to $x_0 = 0$. This restriction is without loss of generality up to the monotone transformation. Let $\hat{m}_-$ and $\hat{m}_+$ be solutions to

$$\min_{\phi \in \mathcal{M}[-1,0)} \sum_{i \in I_-} (Y_i - \phi(X_i))^2 \qquad \text{and} \qquad \min_{\phi \in \mathcal{M}[0,1]} \sum_{i \in I_+} (Y_i - \phi(X_i))^2$$

respectively, where I_- and I_+ are sets of indices corresponding to negative and positive values of observations of the running variable. Therefore, $\hat{m}_-$ and $\hat{m}_+$ are the isotonic regression estimators before and after the cutoff. Note that the isotonic regression is a solution to the constrained least-squares problem over the closed convex cone of monotone functions and should have the projection interpretation when the monotonicity is violated. We might expect that the deviations from monotonicity far from the cutoff should be less harmful as opposed to the deviations at the cutoff.

Choice of c and a. Following the discussion in Section 2, we recommend setting $c = 1$[13] and $a = 1/3$ for the point estimation and $a = 1/2$ for inference. For the point estimation, this choice delivers the one-sided counterpart to the asymptotic distribution at the interior, while for inference, this choice leads to the consistent bootstrap estimates of the asymptotic distribution, balancing the accuracy of the asymptotic approximation and the convergence rate of the estimator. The choice $a = 1/3$ for point estimation can also be justified under the quadratic loss function as the one that balances the rates of the bias and the variance. On the other hand, we expect that the choice $a = 1/2$ for inference should balance the rate of convergence of confidence intervals and the rate of the accuracy of bootstrap approximations. The inference-optimal choice of c is a more difficult problem that we leave for future research. One possible solution is to develop the loss-function based approach that pays attention to the coverage errors and confidence interval length; see Calonico, Cattaneo, and Farrell (2020) or Lazarus, Lewis, Stock, and Watson (2018) for this type of analysis.[14]

Therefore, to obtain point estimates for sharp designs we focus on the following

[13] We also consider the MSE-optimal choice of c in Monte Carlo experiments and find that it tends to increase the MSE in small samples.

[14] Recently, Han and Kato (2019) develop Berry-Esseen inequalities that might be leveraged on to characterize the coverage error.

boundary-corrected sharp iRDD estimator

$$\hat{\theta} = \hat{m}_+\left(cn^{-1/3}\right) - \hat{m}_-\left(-cn^{-1/3}\right).$$

For fuzzy designs, we also need to estimate treatment assignment probabilities before and after the cutoff, denoted $\hat{p}_-$ and $\hat{p}_+$, solving

$$\min_{\pi\in\mathcal{M}[-1,0)}\sum_{i\in I_-}(D_i-\pi(X_i))^2 \qquad \text{and} \qquad \min_{\pi\in\mathcal{M}[0,1]}\sum_{i\in I_+}(D_i-\pi(X_i))^2.$$

The fuzzy iRDD estimator is computed as

$$\hat{\theta}^F = \frac{\hat{m}_+\left(cn^{-1/3}\right) - \hat{m}_-\left(-cn^{-1/3}\right)}{\hat{p}_+\left(cn^{-1/3}\right) - \hat{p}_-\left(-cn^{-1/3}\right)}.$$

3.3 Large sample distribution

Put $p(x) = \Pr(D = 1|X = x)$ and for a function $g : [-1,1] \to \mathbf{R}$, define $g_+ = \lim_{x\downarrow 0} g(x)$ and $g_- = \lim_{x\uparrow 0} g(x)$ with some abuse of notation. The following assumption imposes several mild restrictions on the distribution of the data.

Assumption 3.1. *$(Y_i, D_i, X_i)_{i=1}^n$ is an i.i.d. sample of (Y, D, X) such that (i) $\mathbb{E}[|\varepsilon|^{2+\delta}|X] \leq C < \infty$ for some $\delta > 0$ and m is uniformly bounded; (ii) the distribution of X has Lebesgue density f, uniformly bounded away from zero and infinity on the support of X, and such that f_- and f_+ exist; (iii) σ^2 is uniformly bounded on $[-1,1]$ and σ^2_+ and σ^2_- exist; (iv) m is continuously differentiable in the right and left neighborhoods of zero with $m'_-, m'_+ > 0$; (v) $m \in \mathcal{M}[-1,1]$.*

Assumption 3.1 is comparable to assumptions typically used in the RDD literature, e.g., see Hahn, Todd, and Van der Klaauw (2001), Theorem 4. Note that we are agnostic about the smoothness of the marginal density of X, and only assume the existence of one-sided derivatives of conditional means. The following results describes the large sample approximation to the distribution of the sharp iRDD estimator.

Theorem 3.1. *Suppose that Assumption 3.1 is satisfied. Then*

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

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

The proof of this result can be found in the appendix. Theorem 3.1 shows that the large sample distribution of the sharp iRDD estimator can be approximated by the difference of slopes of the greatest convex minorants of two scaled independent Brownian motions plus the parabola originating from zero and running in opposite directions. To describe the large sample distribution for the fuzzy iRDD, define additionally the conditional covariance function $\rho(x) = \mathbb{E}[\varepsilon(D - p(X))|X = x]$.

Theorem 3.2. *Suppose that Assumption 3.1 is satisfied. Suppose additionally that $p \in \mathcal{M}[-1, 1]$ is continuously differentiable in the right and left neighborhoods of zero with $p'_-, p'_+ > 0$, and $p_+, p_- \in (0, 1)$. Then*

$$n^{1/3}(\hat{\theta}^F - \theta) \xrightarrow{\mathrm{d}} \frac{1}{p_+ - p_-}\xi_1 - \frac{m_+ - m_-}{(p_+ - p_-)^2}\xi_2$$

with

$$\xi_1 = D^L_{[0,\infty)}\left(\sqrt{\frac{\sigma_+^2}{cf_+}}W_t^+ + \frac{t^2c}{2}m'_+\right)(1) - D^L_{(-\infty,0]}\left(\sqrt{\frac{\sigma_-^2}{cf_-}}W_t^- + \frac{t^2c}{2}m'_-\right)(-1)$$

$$\xi_2 = D^L_{[0,\infty)}\left(\sqrt{\frac{p_+(1-p_+)}{cf_+}}B_t^+ + \frac{t^2c}{2}p'_+\right)(1) - D^L_{(-\infty,0]}\left(\sqrt{\frac{p_-(1-p_-)}{cf_-}}B_t^- + \frac{t^2c}{2}p'_-\right)(-1),$$

where W_t^+, W_t^-, B_t^+, and B_t^- are standard Brownian motions such that any two processes with different signs are independent, and

$$\mathrm{Cov}(W_t^+, B_s^+) = \frac{\rho_+(t \wedge s)}{\sqrt{\sigma_+^2 p_+(1-p_+)}}, \qquad \mathrm{Cov}(W_t^-, B_s^-) = \frac{\rho_-(t \wedge s)}{\sqrt{\sigma_-^2 p_-(1-p_-)}}.$$

Both results follow from the CLT for the boundary-corrected isotonic regression estimator obtained in Theorem 2.1; see appendix for the formal proof. A consequence of Theorems 3.1 and 3.2 is that the boundary-corrected iRDD estimators $\hat{\theta}$ and $\hat{\theta}^F$ are consistent in probability for the causal effect parameter θ and provide valid point estimates.

3.4 Trimmed wild bootstrap

In this section, we study the trimmed wild bootstrap. The resampling procedure is as follows. First, we construct the trimmed estimator

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

Second, we simulate the wild bootstrap samples as follows

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

where $(\eta_i^*)_{i=1}^n$ are i.i.d. multipliers, independent of $(Y_i, D_i, X_i)_{i=1}^n$, and $\tilde{\varepsilon}_i = Y_i - \tilde{m}(X_i)$. We call this procedure trimming since it trims the estimator close to boundaries when we generate bootstrap samples. Trimming is needed in addition to the boundary correction of the iRDD estimator

$$\check{\theta} = \hat{m}_+ \left(cn^{-1/2}\right) - \hat{m}_- \left(-cn^{-1/2}\right)$$

and its bootstrapped counterpart

$$\check{\theta}^* = \hat{m}_+^*(cn^{-1/2}) - \hat{m}_-^*(-cn^{-1/2}),$$

where $\hat{m}_+^*$ and $\hat{m}_-^*$ are isotonic estimators computed from the bootstrapped sample $(Y_i^*, D_i, X_i)_{i=1}^n$ similarly to $\hat{m}_-$ and $\hat{m}_+$. The following results establishes the bootstrap consistency for the sharp iRDD.

Theorem 3.3. *Suppose that Assumptions 3.1 are satisfied and that m is γ-Hölder continuous in the left and the right neighborhoods of zero with $\gamma > 1/2$. Suppose also that m is continuous on $[-1, 0)$ and $(0, 1]$, and that $\mathbb{E}[\varepsilon^4|X] \leq C < \infty$. If multipliers $(\eta_i^*)_{i=1}^n$ are such that $\mathbb{E}\eta_i^* = 0$, $\operatorname{Var}(\eta_i^*) = 1$, and $\mathbb{E}|\eta_i^*|^{2+\delta} < \infty$ for some $\delta > 0$, then for every $u \in \mathbf{R}$*

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

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

For fuzzy designs, the bootstrapped estimator is

$$\check{\theta}^{F*} = \frac{\hat{m}_+^*(cn^{-1/2}) - \hat{m}_-^*(-cn^{-1/2})}{\hat{p}_+(cn^{-1/2}) - \hat{p}_-(-cn^{-1/2})}.$$

The proof of the bootstrap consistency is similar to the proof of Theorems 3.2 and 3.3, so we omit it.

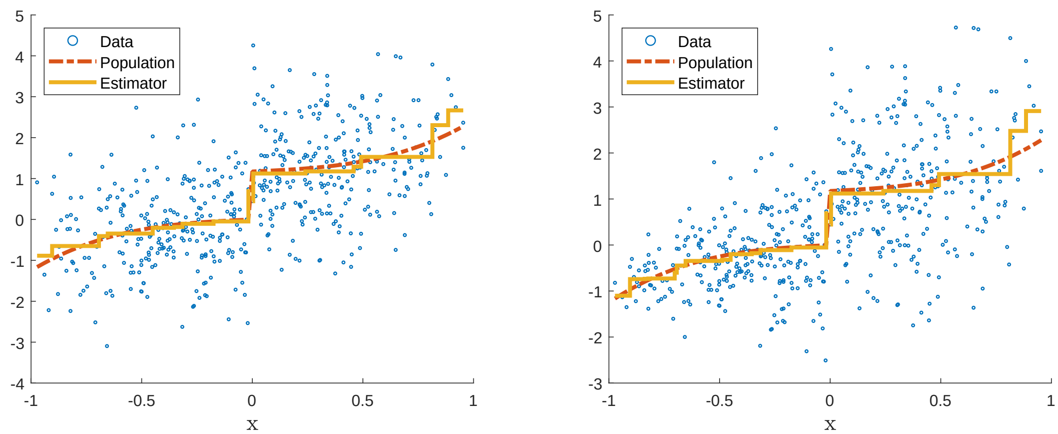
4 Monte Carlo experiments

In this section, we study the finite-sample performance of our iRDD estimator. We simulate $5,000$ samples of size $n \in \{200, 500, 1000\}$ as follows:

$$Y = m(X) + \theta \mathbb{1}_{[0,1]}(X) + \sigma(X)\varepsilon,$$

where $\varepsilon \sim N(0,1)$ and $\varepsilon \perp\!\!\!\perp X$.

In the baseline DGP, we set $m(x) = x^3 + 0.25x$, $\theta = 1$, $X \sim 2 \times \text{Beta}(2,2) - 1$, and $\sigma(x) = 1$ (homoskedasticity), and $\sigma(x) = \sqrt{x+1}$ (heteroskedasticity). We compute the boundary-corrected estimator using the pool adjacent violators algorithm. Figure 1 shows the piecewise-constant interpolation of fitted isotonic regression functions before and after the cutoff, the population regression function, and the data. The data-generating process (DGP) features a low signal-to-noise ratio, and the magnitude of the jump discontinuity is not detectable from visual inspection of the scatter plot.



(a) Homoskedasticity

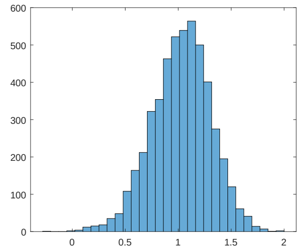
(b) Heteroskedasticity

Figure 1: Single MC experiment, $n = 500$.

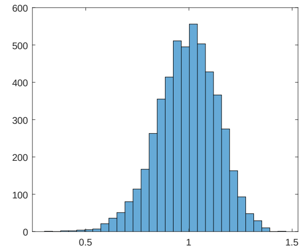
Figures 2 illustrates the finite sample distribution of the boundary-corrected iRDD estimator for samples of different sizes. The exact finite-sample distribution is centered around the population value of the parameter and concentrates around the population parameter as the sample size increases.

Tables 1 and 2 report results of more comprehensive Monte Carlo experiments for several data-generating processes and shows the exact finite-sample bias, variance, and MSE of our iRDD estimator. We consider the following variations of the baseline DGP with two different functional forms and different amount of density near the cutoff:

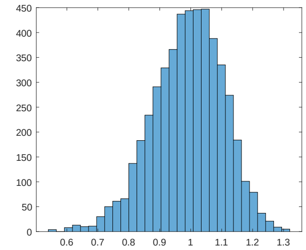
1. DGP 1 sets $X \sim 2 \times \text{Beta}(2,2) - 1$ and $m(x) = \exp(0.25x)$;
2. DGP 2 sets $X \sim 2 \times \text{Beta}(0.5,0.5) - 1$ (low density) and $m(x) = \exp(0.25x)$;



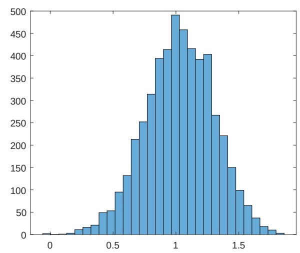
(a) $n = 200$



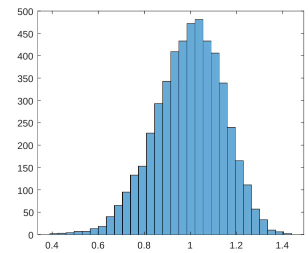
(b) $n = 500$



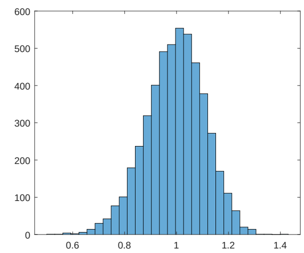
(c) $n = 1000$



(d) $n = 200$



(e) $n = 500$



(f) $n = 1000$

Figure 2: Finite sample distribution of the iRDD estimator: homoskedastic design in (a)-(c) and heteroskedastic in (d)-(f). Results based on 5,000 MC experiments.

3. DGP 3 sets $X \sim 2 \times \text{Beta}(2,2) - 1$ and $m(x) = x^3 + 0.25x$;

4. DGP 4 sets $X \sim 2 \times \text{Beta}(0.5, 0.5) - 1$ (low density) and $m(x) = x^3 + 0.25x$;

To see the benefits of shape restrictions in small samples, we benchmark our iRDD estimators with the unrestricted local polynomial[15] and the k-nearest neighbors estimators.[16]

Results of experiments are consistent with the asymptotic theory. The bias, the variance, and the MSE decrease with the sample size. As expected, the MSE is higher when the density near the cutoff is lower. The results for heteroskedastic designs in Table 2 are similar. The shape-constrained iRDD estimator outperforms the

[15] We rely on the implementation available in the `rdrobust` R package; see Calonico, Cattaneo, and Titiunik (2014), Calonico, Cattaneo, Farrell, and Titiunik (2019), and Calonico, Cattaneo, and Farrell (2018) for details.

[16] Note that the iRDD estimator with the boundary correction $\lfloor n^{1/3} \rfloor$ might resemble the k-nearest neighbors estimator with $k_n = \lfloor n^{1/3} \rfloor$ neighbors. However, such comparison is deceptive because the k-nearest neighbors estimator with $\lfloor n^{1/3} \rfloor$ neighbors is restricted to take a constant value in the *entire* neighborhood of the boundary, while the isotonic regression estimator is *unrestricted*.

Table 1: MC experiments: homoskedasticity

	n	iRDD Bias	iRDD Var	iRDD MSE	LP Bias	LP Var	LP MSE	k-NN Bias	k-NN Var	k-NN MSE
DGP 1										
	200	0.001	0.043	0.043	0.008	0.319	0.319	0.010	0.500	0.500
	500	-0.009	0.022	0.022	-0.007	0.110	0.110	0.010	0.280	0.280
	1000	-0.008	0.013	0.013	-0.002	0.056	0.056	0.000	0.220	0.220
DGP 2										
	200	-0.111	0.080	0.092	0.003	0.747	0.747	0.040	0.500	0.500
	500	-0.079	0.042	0.049	-0.002	0.225	0.225	0.010	0.290	0.290
	1000	-0.063	0.027	0.031	0.001	0.102	0.102	0.000	0.220	0.220
DGP 3										
	200	-0.131	0.041	0.057	-0.002	0.322	0.322	0.030	0.510	0.510
	500	-0.114	0.019	0.031	-0.005	0.109	0.109	0.010	0.290	0.290
	1000	-0.098	0.011	0.020	0.001	0.056	0.056	0.010	0.220	0.220
DGP 4										
	200	-0.337	0.075	0.158	-0.019	0.773	0.774	0.050	0.490	0.490
	500	-0.261	0.037	0.087	-0.021	0.213	0.213	0.010	0.290	0.290
	1000	-0.213	0.021	0.055	-0.007	0.104	0.104	0.010	0.220	0.220

Exact finite-sample bias, variance, and MSE of iRDD, local polynomial (LP), and k-NN estimators. 5000 experiments. Local linear estimator with kernel=‘triangular’ and bwselect=‘mserd’.

unrestricted estimators. In many cases, we achieve more than 5-10 fold reduction of the MSE. In the Appendix, Section A.3, we also consider DGPs, where the regression function violates the monotonicity constraint and regressions that are steep at the cutoff. The results are still favorable towards the iRDD estimator.

We shall acknowledge that the local linear and the k-NN estimators are applicable in more general settings as they do not rely on the monotonicity assumption and we do not recommend using the iRDD estimator for non-monotone designs.[17] However, under the monotonicity, our estimator adapts well to the boundary and seems to outperforms the existing methods. We shall also note that the case of the steep regression function is difficult for all existing approaches and that it becomes difficult to distinguish between the discontinuous and the steep continuous regression functions; see Bertanha (2020) for related impossibility results.

[17] The monotonicity of regression is a testable; see Chetverikov (2019) and references therein.

Table 2: MC experiments: heteroskedasticity

	n	iRDD Bias	iRDD Var	iRDD MSE	LP Bias	LP Var	LP MSE	k-NN Bias	k-NN Var	k-NN MSE
DGP 1										
	200	0.005	0.043	0.043	0.009	0.319	0.319	0.010	0.490	0.490
	500	-0.007	0.022	0.022	-0.008	0.111	0.111	0.010	0.280	0.280
	1000	-0.007	0.013	0.013	-0.001	0.056	0.056	0.000	0.220	0.220
DGP 2										
	200	-0.101	0.083	0.093	0.006	0.791	0.791	0.090	0.200	0.210
	500	-0.075	0.043	0.049	0.000	0.224	0.224	0.040	0.090	0.090
	1000	-0.061	0.027	0.031	-0.000	0.103	0.103	0.030	0.070	0.070
DGP 3										
	200	-0.127	0.041	0.057	-0.004	0.323	0.323	0.020	0.510	0.510
	500	-0.111	0.019	0.031	-0.005	0.109	0.109	0.020	0.280	0.280
	1000	-0.096	0.011	0.020	0.000	0.057	0.057	0.000	0.230	0.230
DGP 4										
	200	-0.286	0.073	0.155	-0.019	0.753	0.754	0.020	0.480	0.490
	500	-0.223	0.035	0.084	-0.019	0.213	0.213	0.000	0.280	0.280
	1000	-0.181	0.019	0.052	-0.009	0.104	0.104	0.010	0.210	0.210

Exact finite-sample bias, variance, and MSE of iRDD, local polynomial (LP), and k-NN estimators. 5000 experiments. Local linear estimator with kernel='triangular' and bwselect='mserd'.

Table 3: MC experiments for the iRDD estimator with the MSE-optimal boundary correction.

	n	Homoskedasticity Bias	Var	MSE	Heteroskedasticity Bias	Var	MSE
DGP 1							
	200	0.163	0.070	0.096	0.172	0.070	0.100
	500	0.060	0.018	0.021	0.063	0.018	0.022
	1000	0.043	0.011	0.013	0.045	0.011	0.013
DGP 2							
	200	0.024	0.097	0.097	0.045	0.100	0.102
	500	0.016	0.031	0.031	0.022	0.032	0.032
	1000	0.013	0.020	0.020	0.016	0.020	0.020
DGP 3							
	200	0.280	0.085	0.163	0.286	0.085	0.167
	500	0.104	0.021	0.032	0.106	0.021	0.032
	1000	0.070	0.012	0.017	0.071	0.012	0.017
DGP 4							
	200	0.241	0.149	0.207	0.249	0.149	0.211
	500	0.082	0.041	0.048	0.085	0.042	0.049
	1000	0.050	0.025	0.027	0.052	0.025	0.028

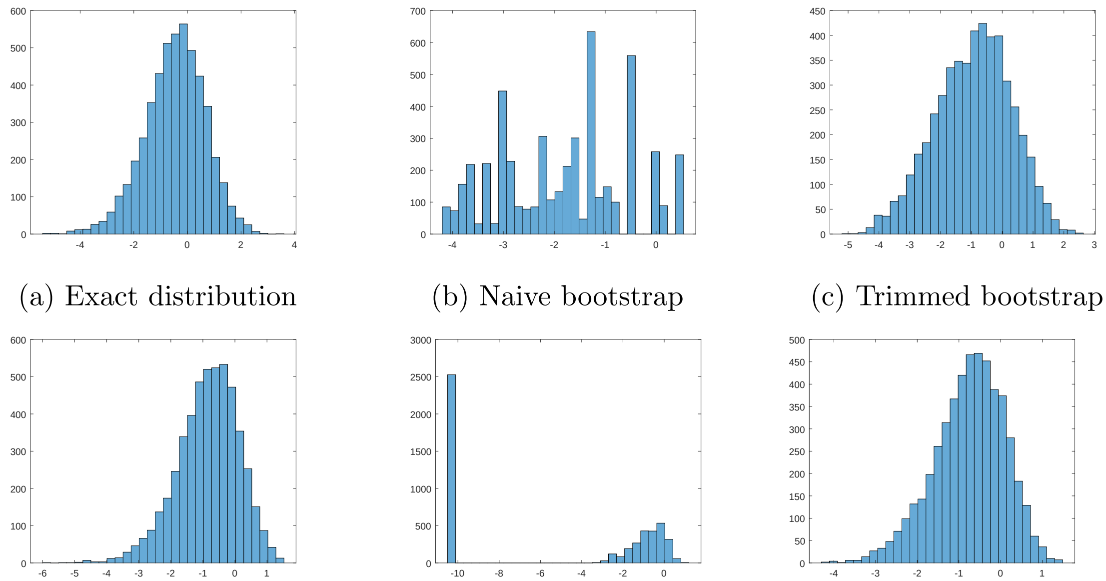
Finite-sample bias, variance, and MSE of $\hat{\theta}$. 5,000 experiments.

Next, we consider whether the data-driven choice of c can improve the rule-of-thumb $c = 1$. According to Proposition A.2.2, the MSE-optimal estimator is

$$\hat{m}^*(0) = \hat{m}(c^* \hat{A} n^{-1/3}),$$

where $\hat{A} = \left(\frac{2}{\hat{m}'(0)}\sqrt{\frac{\hat{\sigma}^2(0)}{\hat{f}(0)}}\right)^{2/3}$, $\hat{m}', \hat{\sigma}^2, \hat{f}$ are consistent estimators of m', σ^2, f, and c^* minimizes $\mathbb{E}|D^L_{[0,\infty)}(W_t - t^2)(c)|^2$. Kulikov and Lopuhaä (2006) find in simulations that $c^* \approx 0.345$. To have an idea of how much the MSE-optimal estimator performs in small samples, we consider the oracle choice of m', σ^2, and f. Results of these Monte Carlo experiments are presented in Table 3. We find that the MSE-optimal estimator is inferior in small samples compared to the rule-of-thumb choice $c = 1$, cf. Tables 1 and 2. The increase in the MSE might come from the fact that asymptotic MSE might be a poor approximation to the finite-sample MSE, which is consistent with simulation results for the Grenander estimator in Kulikov and Lopuhaä (2006).

We asses the validity of the trimmed wild bootstrap approximations, In Figure 3, we plot the exact distribution $n^{1/4}(\hat{\theta}-\theta)$ and the bootstrap distribution $n^{1/4}(\hat{\theta}^{*}-\hat{\theta})$ for samples of size $n \in \{200, 1000\}$. Both estimators are computed using $a = 1/2$ and $c = 1$. As we can see from panels (b) and (e), the naive wild bootstrap without trimming and boundary correction does not work. On the other hand, the trimmed wild bootstrap mimics the finite-sample distribution. In our simulations, we use Rademacher multipliers for the bootstrap in all our experiments, i.e., $\eta_i \in \{-1, 1\}$ with equal probabilities.



(a) Exact distribution (b) Naive bootstrap (c) Trimmed bootstrap

(d) Exact distribution (e) Naive bootstrap (f) Trimmed bootstrap

Figure 3: Finite sample distribution and the bootstrap distribution. Sample size: $n = 200$ in panels (a)-(c) and $n = 1,000$ in panels (d)-(f). 5,000 MC experiments.

Finally, we report the empirical coverage and the length of wild bootstrap confidence intervals in Table 4. Since the inference-optimal choice of c is less obvious, we report results for a range of value of c. We find that for $c = 1$, the coverage is approximately 90% for a nominal coverage 95% and that the coverage gets closer to the nominal 95% coverage for larger values of c as a sample size increases. Designing inference-optimal data-driven methods to select c is an important point left for the future research; see e.g., Calonico, Cattaneo, and Farrell (2018) and Lazarus, Lewis, Stock, and Watson (2018) for example of such analysis. Note also that the confidence intervals shrink as the sample size increases.

	c	$n=200$ Coverage	Length	$n=500$ Coverage	Length	$n=1000$ Coverage	Length
DGP 1	1	0.892	1.108	0.910	0.897	0.909	0.763
	2	0.920	0.915	0.925	0.722	0.926	0.603
	3	0.928	0.851	0.938	0.647	0.929	0.536
	4	0.937	0.822	0.946	0.608	0.946	0.497
	5	0.938	0.814	0.930	0.585	0.933	0.472
DGP 2	1	0.852	1.460	0.889	1.242	0.882	1.078
	2	0.888	1.186	0.904	0.980	0.907	0.843
	3	0.897	1.049	0.903	0.854	0.912	0.736
	4	0.891	0.963	0.903	0.782	0.921	0.669
	5	0.886	0.914	0.899	0.731	0.912	0.625
DGP 3	1	0.892	1.139	0.903	0.929	0.913	0.786
	2	0.911	0.965	0.917	0.760	0.924	0.633
	3	0.915	0.897	0.925	0.693	0.932	0.570
	4	0.907	0.875	0.937	0.657	0.940	0.536
	5	0.869	0.872	0.922	0.642	0.937	0.515
DGP 4	1	0.853	1.550	0.886	1.310	0.895	1.134
	2	0.885	1.303	0.915	1.071	0.913	0.906
	3	0.899	1.186	0.913	0.962	0.926	0.808
	4	0.878	1.121	0.920	0.898	0.939	0.754
	5	0.846	1.080	0.917	0.863	0.935	0.718

Table 4: Coverage and length. 5000 experiments.

5 Empirical illustration

Do incumbents have any electoral advantage? An extensive literature, going back at least to Cummings Jr. (1967), aims to answer this question. Estimating the causal effect is elusive because incumbents, by definition, are more successful politicians. Using the regression discontinuity design, Lee (2008), documents that for the U.S. Congressional elections during 1946-1998, the incumbency advantage is 7.7% of the votes share on the next election. The design is plausibly monotone, since we do not expect that candidates with a larger margin should have a smaller vote share on the next election, at least on average. The unconstrained regression estimates presented by Lee (2008) also support the monotonicity empirically.

We revise the main finding of Lee (2008) with our sharp iRDD estimator. The dataset is publicly available as a companion for the book by Angrist and Pischke (2008). We use the rule-of-thumb choices of tuning parameters, $c = 1$, $a = 1/3$ (point estimation), and $a = 1/2$ (inference). Figure 4 presents the isotonic regression estimates[18] of the average vote share for the Democratic party at next elections as a function of the vote share margin at the previous election (left panel). There is a pronounced jump in average outcomes for Democrats who barely win the election, compared to the results for the penultimate election (right panel). We find the point estimate of 13.8% with the 95% confidence interval $[6.6\%, 26.5\%]$. While we reject the hypothesis that the incumbency advantage did not exist, our confidence intervals give a wider range of estimates. Our confidence interval may be conservative if the underlying regression is two times differentiable, however, it is robust to the failure of this assumption as well as to the inference after the model selection issues.

Of course, different approaches work differently in finite samples and it is hard to have a definite comparison. Lee (2008) estimates the causal effect by fitting parametric regressions with the global fourth-degree polynomial, which might be unstable at the boundary; see Gelman and Imbens (2019). We, on the other hand, rely on the nonparametric boundary-corrected isotonic regression. Interestingly, the local linear estimator with the data-driven bandwidth parameter as implemented in the STATA package `rdrobust` estimates the causal effect of 6.6%.

We also compute iRDD estimates using isotonic regressions without the boundary correction, and evaluating the regression function at the most extreme to the boundary observations. With this approach, we obtain point estimates of 6.6%. As a consequence of the monotonicity, the isotonic regression estimator is biased upwards to the left of the cutoff and downwards to the right of the cutoff; see the

[18] We use the piecewise-constant interpolation, but a higher-order polynomial interpolation is another alternative that would produce visually more appealing estimates.

Appendix, Section A.1. Therefore, our iRDD estimators without the boundary correction provide a lower bound on the estimated causal effect, which might be of significant interest given that it is obtained completely tuning-free.

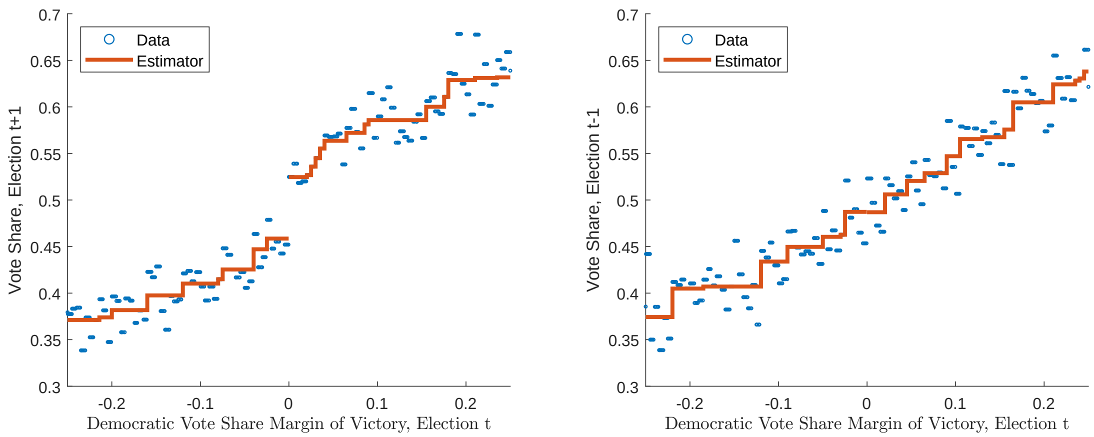


Figure 4: Incumbency advantage. Sample size: 6,559 observations with 3,819 observations below the cutoff.

6 Conclusion

This paper studies the isotonic regression estimator at the boundary of its support, an object that is particularly interesting and required in the analysis of monotone regression discontinuity designs. To the best of our knowledge, our paper is the first to develop the asymptotic theory for the isotonic regression at the boundary point. Building on these results, we offer a new perspective on monotone regression discontinuity designs and contribute more broadly to the growing literature on nonparametric identification, estimation, and inference under shape restrictions.

While the present paper focuses on the isotonic regression discontinuity designs in the classical sharp and fuzzy setting, it also opens several directions for future research. First, since various functional relations in economics are expected to be monotone, the isotonic regression might be of interest in other applications beyond the regression discontinuity designs. Second, it could be interesting to investigate how the monotonicity and other shape restrictions can be incorporated in other regression discontinuity designs; see Escanciano and Cattaneo (2017) for a review. Designing the coverage-optimal choices of the boundary correction is another important point that should be addressed in the future research. Finally, in the large-sample approximations that we use for the inference, we do not assume the existence of derivatives and rely instead on the weaker one-sided Hölder continuity condition. This indicates that our results might be honest to the relevant Hölder class, but additional study is needed to confirm this conjecture.

Acknowledgements

We greatly benefited from comments made by the Editor, the Associate Editor and valuable referees. We are also grateful to Alex Belloni, Federico Bugni, Matias Cattaneo, Xiaohong Chen, Peter Hansen, Jonathan Hill, Jia Li, Matt Masten, Andrew Patton, Andres Santos, Valentin Verdier, Jon Wellner, and other participants of various seminars and conferences for helpful comments and conversations. A special thanks goes to Kenichi Nagasawa for his insightful comments on the earlier draft. All remaining errors are ours.

References

Abadie, A., and M. D. Cattaneo (2018): "Econometric methods for program evaluation," *Annual Review of Economics*, 10, 465–503.

Abrevaya, J., and J. Huang (2005): "On the bootstrap of the maximum score estimator," *Econometrica*, 73(4), 1175–1204.

Anevski, D., and O. Hössjer (2002): "Monotone regression and density function estimation at a point of discontinuity," *Journal of Nonparametric Statistics*, 14(3), 279–294.

Angrist, J. D., and J.-S. Pischke (2008): *Mostly harmless econometrics: an empiricist's companion*. Princeton University Press.

Angrist, J. D., and M. Rokkanen (2015): "Wanna get away? Regression discontinuity estimation of exam school effects away from the cutoff," *Journal of the American Statistical Association*, 110(512), 1331–1344.

Armstrong, T. (2015): "Adaptive testing on a regression function at a point," *Annals of Statistics*, 43(5), 2086–2101.

Ayer, M., H. D. Brunk, G. Ewing, T. Reid, and E. Silverman (1955): "An empirical distribution function for sampling with incomplete information," *Annals of Mathematical Statistics*, 26(4), 641–647.

Balabdaoui, F., H. Jankowski, M. Pavlides, A. Seregin, and J. Wellner (2011): "On the Grenander estimator at zero," *Statistica Sinica*, 21(2), 873–899.

Barlow, R. E., D. Bartholomew, J. Bremner, and H. Brunk (1972): *Statistical inference under order restrictions: The theory and application of isotonic regression*. Wiley New York.

Bertanha, M. (2020): "Regression discontinuity design with many thresholds," *Journal of Econometrics*.

Brunk, H. (1956): "On an inequality for convex functions," *Proceedings of the American Mathematical Society*, 7(5), 817–824.

——— (1970): "Estimation of isotonic regression," in *Nonparametric Techniques in Statistical Inference*, pp. 177–195. Cambridge University Press.

Bugni, F. A., and I. A. Canay (2020): "Testing Continuity of a Density via g-order statistics in the Regression Discontinuity Design," *Journal of Econometrics (forthcoming)*.

Calonico, S., M. D. Cattaneo, and M. H. Farrell (2018): "Optimal bandwidth choice for robust bias corrected inference in regression discontinuity designs," *arXiv preprint arXiv:1809.00236*.

——— (2020): "Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs," *The Econometrics Journal*, 23(2), 192–210.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2019): "Regression discontinuity designs using covariates," *Review of Economics and Statistics*, 101(3), 442–451.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014): "Robust nonparametric confidence intervals for regression-discontinuity designs," *Econometrica*, 82(6), 2295–2326.

Canay, I. A., and V. Kamat (2017): "Approximate permutation tests and induced order statistics in the regression discontinuity design," *Review of Economic Studies*, 85(3), 1577–1608.

CATTANEO, M. D., N. IDROBO, AND R. TITIUNIK (2019): *A Practical Introduction to Regression Discontinuity Designs: Foundations*. Cambridge University Press.

CATTANEO, M. D., M. JANSSON, AND X. MA (2019): "Simple local polynomial density estimators," *Journal of the American Statistical Association*, pp. 1–7.

CATTANEO, M. D., M. JANSSON, AND K. NAGASAWA (2020): "Bootstrap-based inference for cube root consistent estimators," *Econometrica (forthcoming)*.

CHERNOZHUKOV, V., W. K. NEWEY, AND A. SANTOS (2015): "Constrained conditional moment restriction models," *arXiv preprint arXiv:1509.06311*.

CHETVERIKOV, D. (2019): "Testing regression monotonicity in econometric models," *Econometric Theory*, 35(4), 729–776.

CHETVERIKOV, D., A. SANTOS, AND A. M. SHAIKH (2018): "The econometrics of shape restrictions," *Annual Review of Economics*, 10, 31–63.

CUMMINGS JR., M. C. (1967): *Congressmen and the Electorate: Elections for the U. S. House and the President, 1920–1964*. New York: The Free Press.

DELGADO, M. A., J. M. RODRÍGUEZ-POO, AND M. WOLF (2001): "Subsampling inference in cube root asymptotics with an application to Manski's maximum score estimator," *Economics Letters*, 73(2), 241–250.

DUROT, C., AND H. P. LOPUHAÄ (2018): "Limit theory in monotone function estimation," *Statistical Science*, 33(4), 547–567.

ESCANCIANO, J. C., AND M. D. CATTANEO (2017): *Regression discontinuity designs: theory and applications (Advances in Econometrics, volume 38)*. Emerald Group Publishing.

FAN, J., AND I. GIJBELS (1992): "Variable bandwidth and local linear regression smoothers," *Annals of Statistics*, 20(4), 2008–2036.

GELMAN, A., AND G. IMBENS (2019): "Why high-order polynomials should not be used in regression discontinuity designs," *Journal of Business & Economic Statistics*, 37(3), 447–456.

GRENANDER, U. (1956): "On the theory of mortality measurement. Part II," *Scandinavian Actuarial Journal*, 1956(2), 125–153.

Groeneboom, P., and G. Jongbloed (2014): *Nonparametric estimation under shape constraints*. Cambridge University Press.

Guntuboyina, A., and B. Sen (2018): "Nonparametric shape-restricted regression," *Statistical Science*, 33(4), 568–594.

Hahn, J., P. Todd, and W. Van der Klaauw (2001): "Identification and estimation of treatment effects with a regression-discontinuity design," *Econometrica*, 69(1), 201–209.

Han, Q., and K. Kato (2019): "Berry-Esseen bounds for Chernoff-type non-standard asymptotics in isotonic regression," *arXiv preprint arXiv:1910.09662*.

Holland, P. W. (1986): "Statistics and causal inference," *Journal of the American Statistical Association*, 81(396), 945–960.

Imbens, G. W., and T. Lemieux (2008): "Regression discontinuity designs: A guide to practice," *Journal of Econometrics*, 142(2), 615–635.

Imbens, G. W., and J. M. Wooldridge (2009): "Recent developments in the econometrics of program evaluation," *Journal of Economic Literature*, 47(1), 5–86.

Kim, J., and D. Pollard (1990): "Cube root asymptotics," *Annals of Statistics*, 18(1), 191–219.

Koltchinskii, V. I. (1994): "Komlós-Major-Tusnády approximation for the general empirical process and Haar expansions of classes of functions," *Journal of Theoretical Probability*, 7(1), 73–118.

Kosorok, M. R. (2008a): "Bootstrapping the Grenander estimator," in *Beyond parametrics in interdisciplinary research: Festschrift in honor of Professor Pranab K. Sen*, vol. 1, pp. 282–292. Institute of Mathematical Statistics.

——— (2008b): *Introduction to empirical processes and semiparametric inference*. Springer.

Kulikov, V. N., and H. P. Lopuhaä (2006): "The behavior of the NPMLE of a decreasing density near the boundaries of the support," *Annals of Statistics*, 34(2), 742–768.

Lazarus, E., D. J. Lewis, J. H. Stock, and M. W. Watson (2018): "HAR inference: Recommendations for practice," *Journal of Business & Economic Statistics*, 36(4), 541–559.

Lee, D. S. (2008): "Randomized experiments from non-random selection in U.S. House elections," *Journal of Econometrics*, 142(2), 675–697.

Lee, D. S., and T. Lemieux (2010): "Regression discontinuity designs in economics," *Journal of Economic Literature*, 48(2), 281–355.

Léger, C., and B. MacGibbon (2006): "On the bootstrap in cube root asymptotics," *Canadian Journal of Statistics*, 34(1), 29–44.

Liu, R. Y. (1988): "Bootstrap procedures under some non-i.i.d. models," *Annals of Statistics*, 16(4), 1696–1708.

McCrary, J. (2008): "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of econometrics*, 142(2), 698–714.

Patra, R. K., E. Seijo, and B. Sen (2018): "A consistent bootstrap procedure for the maximum score estimator," *Journal of Econometrics*, 205(2), 488–507.

Rao, B. P. (1969): "Estimation of a unimodal density," *Sankhyā: The Indian Journal of Statistics, Series A*, 31(1), 23–36.

Rudin, W. (1976): *Principles of mathematical analysis*. McGraw-Hill New York.

Sen, B., M. Banerjee, and M. Woodroofe (2010): "Inconsistency of bootstrap: The Grenander estimator," *Annals of Statistics*, 38(4), 1953–1977.

Thistlethwaite, D. L., and D. T. Campbell (1960): "Regression-discontinuity analysis: An alternative to the ex post facto experiment," *Journal of Educational Psychology*, 51(6), 309–317.

van der Vaart, A. W., and J. A. Wellner (2000): *Weak convergence and empirical processes: with applications to statistics*. Springer.

van Eeden, C. (1958): "Testing and estimating ordered parameters of probability distribution," *Doctoral dissertation, University of Amsterdam.*

Woodroofe, M., and J. Sun (1993): "A penalized maximum likelihood estimate of f(0+) when f is non-increasing," *Statistica Sinica*, 3(2), 501–515.

Wright, F. (1981): "The asymptotic behavior of monotone regression estimates," *Annals of Statistics*, 9(2), 443–448.

Wu, C. (1986): "Jackknife, bootstrap and other resampling methods in regression analysis," *Annals of Statistics*, 14(4), 1261–1295.

APPENDIX

A.1 Inconsistency at the boundary

Put

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

By Barlow, Bartholomew, Bremner, and Brunk (1972), Theorem 1.1, $\hat{m}(x)$ 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 $t = x$; see Figure A.1. The natural estimator of $m(0)$ is $\hat{m}(X_{(1)})$, which corresponds to the slope of the first-segment

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

where $X_{(j)}$ is the j^{th} order statistics and $Y_{(j)}$ is the corresponding observation of the dependent variable.

The following result shows the isotonic regression estimator is can be inconsistent at the boundary.[19] We believe that the exact asymptotic behavior of $\hat{m}(X_{(1)})$ is an interesting problem to study on its own; see Balabdaoui, Jankowski, Pavlides, Seregin, and Wellner (2011) for such analysis in the case of the Grenander estimator.

Theorem A.1. *Suppose that the conditional distribution $\varepsilon | X = 0$ is not degenerate and that $x \mapsto \Pr(Y \leq y | X = x)$ is continuous for every y. Then there exists $\epsilon > 0$ such that*

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

Proof. Let $F_{\varepsilon|X}$ be the conditional CDF of $\varepsilon|X$. Note that $\int_{-\infty}^{0} e \mathrm{d}F_{\varepsilon|X}(e|0) < 0$ since $\mathbb{E}[\varepsilon|X] = 0$ and the distribution $\varepsilon|X = 0$ is not degenerate. Then, by continuity

[19] We are grateful to the anonymous referee for suggesting how to improve the statement of this result.

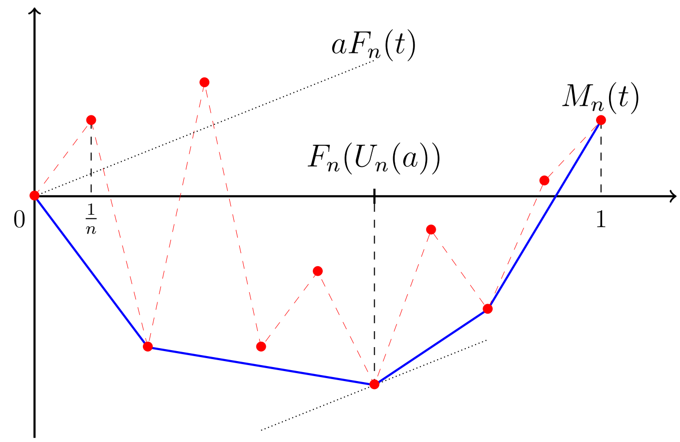


Figure A.1: If $\hat{m}(x) \leq a$, then the largest distance between the line with slope a going through the origin and the greatest convex minorant (broken blue line) of the cumulative sum diagram $t \mapsto (F_n(t), M_n(t))$ (red dots) will be achieved at some point to the right of $F_n(x)$.

$\exists \epsilon > 0$ such that $\int_{-\infty}^{-\epsilon} e \mathrm{d}F_{\varepsilon|X}(e|0) < 0$. This shows that $F_{\varepsilon|X}(-\epsilon|0) > 0$. Next, for any $\epsilon > 0$

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

where we use the fact that $X_{(1)} \xrightarrow{d} 0$. Therefore, there exists $\epsilon > 0$ such that

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

□

The exact asymptotic behavior of $\hat{m}(X_{(1)})$ is not clear, but we know that it will underestimate $m(0)$ in general. To see this, note that

$$\begin{aligned}\Pr\left(\hat{m}(X_{(1)}) \leq m(0)\right) &\geq \Pr(\hat{m}(cn^{-1/3}) \leq m(0)) \\ &\to 1,\end{aligned}$$

where the first inequality follows from $X_{(1)} \leq cn^{-1/3}$ for some $c > 0$ and monotonicity and the second by the consistency of $\hat{m}(cn^{-1/3})$; see Theorem 2.1 (ii). Therefore,

$$\lim_{n\to\infty} \Pr\left(\hat{m}(X_{(1)}) \leq m(0)\right) = 1$$

and for large n, $\hat{m}(X_{(1)})$ will be smaller than $m(0)$ with high probability. Similarly, one can show that $\hat{m}(X_{(n)})$ overestimates 1.

A.2 Proofs of main results

Proof of Theorem 2.1. By Barlow, Bartholomew, Bremner, and Brunk (1972), Theorem 1.1, $\hat{m}(x)$ 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 $t = x$. This corresponds to the piecewise-constant left-continuous interpolation. Put

$$U_n(a) = \arg\max_{s\in[0,1]} \left\{aF_n(s) - M_n(s)\right\}.$$

Then for any[20] $x \in (0,1)$ and $a \in \mathbf{R}$

$$\begin{aligned}\hat{m}(x) \leq a &\iff F_n(U_n(a)) \geq F_n(x) \\ &\iff U_n(a) \geq x,\end{aligned} \tag{A.1}$$

as can be seen from Figure A.1, see also Groeneboom and Jongbloed (2014), Lemma 3.2.

[20] For a continuous function $\Phi : [0,1] \to \mathbf{R}$, we define $\arg\max_{t\in[0,1]} \Phi(t) = \max\{s \in [0,1] : \Phi(s) = \max_{t\in[0,1]} \Phi(t)\}$ to accomodate non-unique maximizers. Recall that continuous function attains its maximum on compact intervals and its argmax is a closed set with a well-defined maximal element.

Case (i): $a \in (0, 1/3)$. For every $u \in \mathbf{R}$

$$
\begin{aligned}
&\Pr\left(n^{1/3}\left(\hat{m}\left(cn^{-a}\right) - m(cn^{-a})\right) \leq u\right) \\
&= \Pr\left(\hat{m}\left(cn^{-a}\right) \leq n^{-1/3}u + m(cn^{-a})\right) \\
&= \Pr\left(\operatorname*{arg\,max}_{s\in[0,1]}\left\{\left(n^{-1/3}u + m(cn^{-a})\right)F_n(s) - M_n(s)\right\} \geq cn^{-a}\right) \\
&= \Pr\left(\operatorname*{arg\,max}_{t\in[-cn^{1/3-a},(1-cn^{-a})n^{1/3}]}\left\{\left(n^{-1/3}u + m(cn^{-a})\right)F_n(tn^{-1/3}+cn^{-a}) - M_n(tn^{-1/3}+cn^{-a})\right\} \geq 0\right),
\end{aligned}
$$

where the second equality follows by the switching relation in Eq. A.1 and the last by the change of variables $s \mapsto tn^{-1/3} + cn^{-a}$.

The location of the argmax is the same as the location of the argmax of the following process

$$
Z_{n1}(t) \triangleq I_{n1}(t) + II_{n1}(t) + III_{n1}(t)
$$

due to scale and shift invariance with

$$
\begin{aligned}
I_{n1}(t) &= \sqrt{n}(P_n - P)g_{n,t}, \qquad g_{n,t} \in \mathcal{G}_{n1} \\
II_{n1}(t) &= n^{2/3}\mathbb{E}\left[(m(cn^{-a}) - Y)\left(\mathbb{1}_{[0,tn^{-1/3}+cn^{-a}]}(X) - \mathbb{1}_{[0,cn^{-a}]}(X)\right)\right] \\
III_{n1}(t) &= n^{1/3}u[F_n(tn^{-1/3}+cn^{-a}) - F_n(cn^{-a})],
\end{aligned}
$$

where

$$
\mathcal{G}_{n1} = \left\{g_{n,t}(y,x) = n^{1/6}(m(cn^{-a}) - y)\left(\mathbb{1}_{[0,tn^{-1/3}+cn^{-a}]}(x) - \mathbb{1}_{[0,cn^{-a}]}(x)\right) :\ t \in [-K, K]\right\}.
$$

We will show that the process Z_{n1} converges weakly to a non-degenerate Gaussian process in $\ell^\infty[-K, K]$ for every $K < \infty$.

Under Assumption 2.1 (ii)-(iii) the covariance structure of the process I_{n1} converges pointwise to the one of the two-sided scaled Brownian motion (two independent Brownian motions starting from zero and running in the opposite directions). Indeed, when $s, t \geq 0$

$$
\begin{aligned}
\operatorname{Cov}(g_{n,t}, g_{n,s}) &= n^{1/3}\mathbb{E}\left[|Y - m(cn^{-a})|^2\mathbb{1}_{[cn^{-a},cn^{-a}+n^{-1/3}(t\wedge s)]}(X)\right] + O(n^{-1/3}) \\
&= n^{1/3}\mathbb{E}\left[(\varepsilon^2 + |m(X) - m(cn^{-a})|^2)\mathbb{1}_{[cn^{-a},cn^{-a}+n^{-1/3}(t\wedge s)]}(X)\right] + o(1) \\
&= n^{1/3}\int_{cn^{-a}}^{cn^{-a}+n^{-1/3}(t\wedge s)}\left(\sigma^2(x) + |m(x) - m(cn^{-a})|^2\right)f(x)\mathrm{d}x + o(1) \\
&= \left(\sigma^2(\xi_n) + |m(cn^{-a}) - m(\xi_n)|^2\right)f(\xi_n)(t\wedge s) + o(1) \\
&= \sigma^2(0)f(0)(s\wedge t) + o(1),
\end{aligned}
$$

where we use the mean-value theorem for some ξ_n between cn^{-a} and $cn^{-a}+n^{-1/3}(t\wedge s)$. Similarly, it can be shown that

$$\mathrm{Cov}(g_{n,t},g_{n,s})=\begin{cases}\sigma^2(0)f(0)(|s|\wedge|t|)+o(1) & s,t\leq 0\\ o(1) & \mathrm{sign}(s)\neq\mathrm{sign}(t).\end{cases}$$

The class $\mathcal{G}_{n1}$ is VC subgraph with VC index 2 and the envelop

$$G_{n1}(y,x)=n^{1/6}|y-m(cn^{-a})|\mathbb{1}_{[cn^{-a}-n^{-1/3}K,cn^{-a}+n^{-1/3}K]}(x),$$

which is square integrable

$$\begin{aligned}\mathbb{E}G_{n1}^2(Y,X)&=n^{1/3}\mathbb{E}[|Y-m(cn^{-a})|^2\mathbb{1}_{[cn^{-a}-n^{-1/3}K,cn^{-a}+n^{-1/3}K]}(X)]\\&=n^{1/3}\mathbb{E}[\left(\varepsilon^2+|m(X)-m(cn^{-a})|^2\right)\mathbb{1}_{[cn^{-a}-n^{-1/3}K,cn^{-a}+n^{-1/3}K]}(X)]\\&=n^{1/3}\int_{cn^{-a}-n^{-1/3}K}^{cn^{-a}+n^{-1/3}K}\left(\sigma^2(x)+|m(x)-m(cn^{-a})|^2\right)f(x)\mathrm{d}x\\&=O(1).\end{aligned}$$

Next, we verify the Lindeberg's condition under Assumption 2.1 (i)

$$\begin{aligned}\mathbb{E}G_{n1}^2\mathbb{1}\{G_n>\eta\sqrt{n}\}&\leq\frac{\mathbb{E}G_n^{2+\delta}}{\eta^\delta n^{\delta/2}}\\&=\frac{n^{(2+\delta)/6}}{\eta^\delta n^{\delta/2}}\mathbb{E}\left[|Y-m(cn^{-a})|^{2+\delta}\mathbb{1}_{[cn^{-a}-n^{-1/3}K,cn^{-a}+n^{-1/3}K]}(X)\right]\\&=\frac{n^{(2+\delta)/6}}{\eta^\delta n^{\delta/2}}O(n^{-1/3})\\&=o(1).\end{aligned}$$

Lastly, under Assumption 2.1 (iii), for every $\delta_n\to 0$

$$\begin{aligned}\sup_{|t-s|\leq\delta_n}\mathbb{E}|g_{n,t}-g_{n,s}|^2&=n^{1/3}\sup_{|t-s|\leq\delta_n}\mathbb{E}\left[\left|Y-m(cn^{-a})\right|^2\mathbb{1}_{[cn^{-a}+n^{-1/3}t,cn^{-a}+n^{-1/3}s]}(X)\right]\\&=n^{1/3}\sup_{|t-s|\leq\delta_n}\mathbb{E}\left[\left(\varepsilon^2+|m(X)-m(cn^{-a})|^2\right)\mathbb{1}_{[cn^{-a}+n^{-1/3}t,cn^{-a}+n^{-1/3}s]}(X)\right]\\&=O(\delta_n)\\&=o(1).\end{aligned}$$

Therefore, by van der Vaart and Wellner (2000), Theorem 2.11.22

$$I_{n1}(t)\rightsquigarrow\sqrt{\sigma^2(0)f(0)}W_t\qquad\text{in}\qquad\ell^\infty[-K,K].$$

Under Assumption 2.1 (ii) and (iv) by Taylor's theorem

$$\begin{aligned}
II_{n1}(t) &= n^{2/3}\mathbb{E}\left[(m(cn^{-a}) - Y)\left(\mathbb{1}_{[0,cn^{-a}+tn^{-1/3}]}(X) - \mathbb{1}_{[0,cn^{-a}]}(X)\right)\right] \\
&= n^{2/3}\int_{F(cn^{-a})}^{F(cn^{-a}+tn^{-1/3})} \left(m(cn^{-a}) - m(F^{-1}(u))\right) \mathrm{d}u \\
&= -n^{2/3}\frac{m'(0)}{2f(0)}(1+o(1))[F(cn^{-a}+tn^{-1/3}) - F(cn^{-a})]^2 \\
&= -\frac{t^2 m'(0)}{2f(0)}[f(0)]^2(1+o(1)) \\
&= -\frac{t^2}{2}m'(0)f(0) + o(1)
\end{aligned}$$

uniformly over $[-K,K]$. Lastly, by the uniform law of the iterated logarithm

$$\begin{aligned}
III_{n1}(t) &= n^{1/3}u[F(tn^{-1/3}+cn^{-a}) - F(cn^{-a})] + o(1) \\
&= utf(0) + o(1)
\end{aligned}$$

uniformly over $t \in [-K,K]$. Therefore, for every $K < \infty$

$$Z_{n1}(t) \rightsquigarrow utf(0) - \sqrt{\sigma^2(0)f(0)}W_t - \frac{t^2}{2}m'(0)f(0) \triangleq Z_1(t), \quad \text{in} \quad \ell^\infty[-K,K]. \tag{A.2}$$

Next, we verify conditions of the argmax continuous mapping theorem Kim and Pollard (1990), Theorem 2.7. First, note that since

$$\mathrm{Var}(Z_1(s) - Z_1(t)) = \sigma^2(0)f(0)|t-s| \neq 0, \qquad \forall t \neq s,$$

by Kim and Pollard (1990), Lemma 2.6, the process $t \mapsto Z_1(t)$ achieves its maximum a.s. at a unique point. Second, by law of iterated logarithm for the Brownian motion

$$\limsup_{t\to\infty} \frac{W_t}{\sqrt{2t\log\log t}} = 1, \qquad \text{a.s.}$$

which shows that the quadratic term dominates asymptotically, i.e., $Z_1(t) \to -\infty$ as $|t| \to \infty$. It follows that the maximizer of $t \mapsto Z_1(t)$ is tight. Lastly, by Lemma A.2.1 the argmax of $t \mapsto Z_{n1}(t)$ is uniformly tight. Therefore, by the argmax continuous mapping theorem, see Kim and Pollard (1990), Theorem 2.7

$$\begin{aligned}
&\Pr\left(n^{1/3}\left(\hat{m}\left(cn^{-a}\right) - m(0)\right) \leq u\right) \\
&\to \Pr\left(\arg\max_{t\in\mathbf{R}} Z_1(t) \geq 0\right) \\
&= \Pr\left(\arg\max_{t\in\mathbf{R}}\left\{ut - \sqrt{\frac{\sigma^2(0)}{f(0)}}W_t - \frac{t^2}{2}m'(0)\right\} \geq 0\right).
\end{aligned}$$

By the change of variables $t \mapsto \left(\frac{a}{b}\right)^{2/3} s + \frac{c}{2b}$, scale invariance of the Brownian motion $W_{\sigma^2 t - \mu} = \sigma W_t - W_\mu$, and scale and shift invariance of the argmax

$$\arg\max_{t\in\mathbf{R}} \left\{aW_t - bt^2 + ct\right\} = \left(\frac{a}{b}\right)^{2/3} \arg\max_{s\in\mathbf{R}}\{W_s - s^2\} + \frac{c}{2b}.$$

This allows us to simplify the limiting distribution as

$$\begin{aligned}\Pr\left(n^{1/3}\left(\hat{m}\left(cn^{-a}\right) - m(0)\right) \leq u\right) &\to \Pr\left(\arg\max_{t\in\mathbf{R}}\left\{ut - \sqrt{\frac{\sigma^2(0)}{f(0)}}W_t - \frac{t^2}{2}m'(0)\right\} \geq 0\right) \\ &= \Pr\left(\left|\frac{4m'(0)\sigma^2(0)}{f(0)}\right|^{1/3} \arg\max_{s\in\mathbf{R}}\{W_s - s^2\} \leq u\right),\end{aligned}$$

where we the use symmetry of the objective function.

Case (ii): $a \in [1/3, 1)$**.** For every $u \in \mathbf{R}$

$$\begin{aligned}&\Pr\left(n^{(1-a)/2}\left(\hat{m}\left(cn^{-a}\right) - m(0)\right) \leq u\right) \\ &= \Pr\left(\hat{m}\left(cn^{-a}\right) \leq m(0) + n^{(a-1)/2}u\right) \\ &= \Pr\left(\arg\max_{s\in[0,1]}\left\{\left(n^{(a-1)/2}u + m(0)\right)F_n(s) - M_n(s)\right\} \geq cn^{-a}\right) \\ &= \Pr\left(\arg\max_{t\in[0,n^a/c]}\left\{(n^{(a-1)/2}u + m(0))F_n(cn^{-a}t) - M_n(cn^{-a}t)\right\} \geq 1\right),\end{aligned}$$

where the second equality follows by the switching relation in Eq. A.1 and the last by the change of variables $s \mapsto cn^{-a}t$.

The location of the argmax is the same as the location of the argmax of the following process

$$Z_{n2}(t) \triangleq I_{n2}(t) + II_{n2}(t) + III_{n2}(t) + IV_{n2}(t)$$

with

$$\begin{aligned}I_{n2}(t) &= \sqrt{n}(P_n - P)g_{n,t}, \quad g_{n,t} \in \mathcal{G}_{n2} \\ II_{n2}(t) &= n^{(a+1)/2}\mathbb{E}[(m(0) - Y)\mathbb{1}_{[0,cn^{-a}t]}(X)] \\ III_{n2}(t) &= n^a u(F_n(cn^{-a}t) - F(cn^{-a}t)) \\ IV_{n2}(t) &= n^a uF(cn^{-a}t),\end{aligned}$$

where

$$\mathcal{G}_{n2} = \left\{ g_{n,t}(y,x) = n^{a/2}(m(0)-y)\mathbb{1}_{[0,cn^{-a}t]}(x) : \quad t \in [0,K] \right\}.$$

We will show that the process Z_{n2} converges weakly to a non-degenerate Gaussian process in $\ell^\infty[0,K]$ for every $K < \infty$.

Under Assumption 2.1 (ii)-(iii) the covariance structure of the process I_{n2} converges pointwise to the one of the scaled Brownian motion

$$\begin{aligned}
\mathrm{Cov}(g_{n,t}, g_{n,s}) &= n^a \mathbb{E}\left[|Y - m(0)|^2 \mathbb{1}_{[0,cn^{-a}(t\wedge s)]}(X)\right] + o(1) \\
&= n^a \mathbb{E}[\varepsilon^2 \mathbb{1}_{[0,cn^{-a}(t\wedge s)]}(X)] + n^a \mathbb{E}[|m(X) - m(0)|^2 \mathbb{1}_{[0,cn^{-a}(t\wedge s)]}(X)] + o(1) \\
&= n^a \int_0^{cn^{-a}(t\wedge s)} \sigma^2(x)\mathrm{d}F(x) + n^a \int_0^{cn^{-a}(t\wedge s)} |m(x) - m(0)|^2 \mathrm{d}F(x) + o(1) \\
&= \sigma^2(0) f(0) c(s \wedge t) + o(1).
\end{aligned}$$

The class $\mathcal{G}_{n2}$ is VC subgraph with VC index 2 and envelop

$$G_{n2}(y,x) = n^{a/2}|y - m(0)|\mathbb{1}_{[0,cn^{-a}K]}(x),$$

which is square integrable

$$\begin{aligned}
PG_{n2}^2 &= n^a \mathbb{E}[|Y - m(0)|^2 \mathbb{1}_{[0,cn^{-a}K]}(X)] \\
&= n^a \mathbb{E}[\varepsilon^2 \mathbb{1}_{[0,cn^{-a}K]}(X)] + n^a \mathbb{E}[|m(X) - m(0)|^2 \mathbb{1}_{[0,cn^{-a}K]}(X)] \\
&= n^a \int_0^{n^{-a}K} \sigma^2(x)\mathrm{d}F(x) + o(1) \\
&= O(1).
\end{aligned}$$

Next, we verify the Lindeberg's condition under Assumption 2.1 (i)

$$\begin{aligned}
\mathbb{E}G_{n2}^2 \mathbb{1}\{G_n > \eta\sqrt{n}\} &\leq \frac{\mathbb{E}G_{n2}^{2+\delta}}{\eta^\delta n^{\delta/2}} \\
&= \frac{n^{\frac{(2+\delta)a}{2}}}{\eta^\delta n^{\delta/2}} \mathbb{E}[|Y - m(0)|^{2+\delta} \mathbb{1}_{[0,cn^{-a}K]}(X)] \\
&= \frac{n^{\frac{(2+\delta)a}{2}}}{\eta^\delta n^{\delta/2}} O(n^{-a}) = O(n^{-\delta(1-a)/2}) \\
&= o(1).
\end{aligned}$$

Lastly, under Assumption 2.1 (iii), for every $\delta_n \to 0$

$$\begin{aligned}
\sup_{|t-s|\leq\delta_n} \mathbb{E}|g_{n,t} - g_{n,s}|^2 &= n^a \sup_{|t-s|\leq\delta_n} \mathbb{E}\left[|Y - m(0)|^2 \mathbb{1}_{[cn^{-a}t, cn^{-a}s]}(X)\right] \\
&= n^a \sup_{|t-s|\leq\delta_n} \mathbb{E}\left[\varepsilon^2 \mathbb{1}_{[cn^{-a}t, cn^{-a}s]}(X)\right] + \mathbb{E}\left[|m(X) - m(0)|^2 \mathbb{1}_{[cn^{-a}t, cn^{-a}s]}(X)\right] \\
&= O(\delta_n) \\
&= o(1).
\end{aligned}$$

Therefore, by van der Vaart and Wellner (2000), Theorem 2.11.22

$$I_{n2}(t) \rightsquigarrow \sqrt{\sigma^2(0) f(0) c} W_t \qquad \text{in} \qquad \ell^\infty[0, K].$$

Next,

$$II_{n2}(t) = n^{(a+1)/2} \int_0^{F(cn^{-a}t)} (m(0) - m(F^{-1}(u)))\mathrm{d}u.$$

For $a = 1/3$, under Assumption 2.1 (iv), by Taylor's theorem uniformly over t on compact sets

$$\begin{aligned}
II_{n2}(t) &= -n^{(a+1)/2} \frac{1}{2} \frac{m'(F^{-1}(0))}{f(F(0))} [F(cn^{-a}t)]^2 (1 + o(1)) \\
&= -n^{(1-3a)/2} \frac{t^2 c^2}{2} m'(0) f(0)(1 + o(1)) \\
&= -\frac{t^2 c^2}{2} m'(0) f(0) + o(1),
\end{aligned}$$

while for $a \in (1/3, 1)$ under γ-Hölder continuity of m since $\gamma > (1-a)/2a$

$$\begin{aligned}
II_{n2}(t) &= n^{(a+1)/2} \int_0^{cn^{-a}t} (m(0) - m(x)) f(x) \mathrm{d}x \\
&\lesssim n^{(a+1)/2} \int_0^{cn^{-a}t} |x|^\gamma \mathrm{d}x \\
&= O\left(n^{\frac{1-2a\gamma-a}{2}}\right) \\
&= o(1).
\end{aligned} \tag{A.3}$$

Next, by the maximal inequality Kim and Pollard (1990), p.199,

$$\mathbb{E}\left[\sup_{t\in[0,K]} |F_n(cn^{-a}t) - F(cn^{-a}t)|\right] \leq n^{-1/2} J(1) \sqrt{PG_n^2}, \tag{A.4}$$

where $J(1) < \infty$ is the uniform entropy integral and

$$\begin{aligned} PG_n^2 &= F(cn^{-a}t) \\ &= f(0)(1+o(1))cn^{-a}. \end{aligned}$$

Since $a < 1$

$$III_{n2}(t) = O_P(n^{(a-1)/2}) = o_P(1).$$

uniformly over $[0, K]$.

Lastly,

$$\begin{aligned} IV_{n2}(t) &= n^a uF(cn^{-a}t) \\ &= utf(0)c + o(1) \end{aligned}$$

uniformly over $t \in [0, K]$. Therefore, for every $K < \infty$

$$Z_{n2}(t) \rightsquigarrow utf(0)c + \sqrt{\sigma^2(0)f(0)c}W_t - \frac{t^2c^2}{2}m'(0)f(0)\mathbb{1}_{a=1/3} \triangleq Z_2(t) \quad \text{in} \quad \ell^\infty[0, K]. \tag{A.5}$$

Next, we extend processes Z_{n2} and Z_2 to the entire real line as follows

$$\tilde{Z}_{n2}(t) = \begin{cases} Z_{n2}(t), & t \geq 0 \\ t, & t < 0 \end{cases}, \qquad \tilde{Z}_2(t) = \begin{cases} Z_2(t), & t \geq 0, \\ t, & t < 0. \end{cases}$$

We verify conditions of the argmax continuous mapping theorem Kim and Pollard (1990), Theorem 2.7. First, by the similar argument as before, the argmax of $t \mapsto \tilde{Z}_2(t)$ is unique and tight. Second, by Lemma A.2.1, the argmax of $\tilde{Z}_{n2}$ is uniformly tight for every $u \in \mathbf{R}$ when $a = 1/3$ and for every $u < 0$ when $a \in (1/3, 1)$. Therefore, by the argmax continuous mapping theorem, see Kim and Pollard (1990), Theorem 2.7,

$$\begin{aligned} &\Pr\left(n^{(1-a)/2}\left(\hat{m}\left(cn^{-a}\right) - m(0)\right) \leq u\right) \\ &\to \Pr\left(\operatorname*{arg\,max}_{t\in[0,\infty)} Z_2(t) \geq 1\right) \\ &= \Pr\left(\operatorname*{arg\,max}_{t\in[0,\infty)}\left\{ut - \sqrt{\frac{\sigma^2(0)}{cf(0)}}W_t - \frac{t^2c}{2}m'(0)\mathbb{1}_{a=1/3}\right\} \geq 1\right) \\ &= \Pr\left(D^L_{[0,\infty)}\left(\sqrt{\frac{\sigma^2(0)}{cf(0)}}W_t + \frac{t^2c}{2}m'(0)\mathbb{1}_{a=1/3}\right)(1) \leq u\right) \end{aligned}$$

where the last line follows by the switching relation similar to the one in Eq. A.1.

To conclude, it remains to show that when $a \in (1/3, 1)$

$$\Pr\left(n^{(1-a)/2}\left(\hat{m}\left(cn^{-a}\right) - m(0)\right) \geq 0\right) \to 0.$$

This follows since for every $\epsilon > 0$

$$\begin{aligned}
\Pr\left(n^{(1-a)/2}\left(\hat{m}\left(cn^{-a}\right) - m(0)\right) \geq 0\right) &\leq \Pr\left(n^{(1-a)/2}\left(\hat{m}\left(cn^{-a}\right) - m(0)\right) \geq -\epsilon\right) \\
&\to \Pr\left(\operatorname*{arg\,max}_{t\in[0,\infty)}\left\{-\epsilon t - \sqrt{\frac{\sigma^2(0)}{cf(0)}}W_t\right\} \leq c\right) \\
&= \Pr\left(\operatorname*{arg\,max}_{t\in[0,\infty)}\left\{W_t - \sqrt{\frac{cf(0)}{\sigma^2(0)}}\epsilon t\right\} \leq c\right) \\
&= \Pr\left(\operatorname*{arg\,max}_{t\in[0,\infty)}\{W_t - t\} \leq \frac{c^2 f(0)}{\sigma^2(0)}\epsilon\right),
\end{aligned}$$

which tends to zero as $\epsilon \downarrow 0$ as can be seen from

$$\limsup_{t\downarrow 0} \frac{W_t}{\sqrt{2t\log\log(1/t)}} = 1, \qquad \text{a.s.}$$

$\square$

The proof of Theorem 2.1 is based on the argmax continuous mapping theorem, Kim and Pollard (1990), one of the conditions of which is that the argmax is a uniformly tight sequence of random variables. In our setting it is sufficient to show that the argmax of

$$\mathbb{M}_{n1}(s) \triangleq \left(n^{-1/3}u + m(0)\right)\left[F_n(s+g) - F_n(g)\right] - \left[M_n(s+g) - M_n(g)\right], \qquad s \in [0,1]$$

is $O_P(n^{-1/3})$ for $a \in (0, 1/3)$, where $g > 0$ is arbitrary small, and that the argmax of

$$\mathbb{M}_{n2}(s) \triangleq (n^{(a-1)/2}u + m(0))F_n(s) - M_n(s), \qquad s \in [0,1]$$

is $O_P(n^{-a})$ for $a \in [1/3, 1)$. The following lemma serves this purpose.

Lemma A.2.1. *Suppose that Assumption 2.1 is satisfied. Then*

(i) *For $a \in (0, 1/3)$ and $u \in \mathbf{R}$ and every $g > 0$*

$$\operatorname*{arg\,max}_{s\in[-g,1-g]} \mathbb{M}_{n1}(s) = O_P(n^{-1/3}).$$

(ii) For $a = 1/3$ and $u \in \mathbf{R}$

$$\arg\max_{s\in[0,1]} \mathbb{M}_{n2}(s) = O_P(n^{-1/3}).$$

(iii) For $a \in (1/3, 1]$ and $u < 0$

$$\arg\max_{s\in[0,1]} \mathbb{M}_{n2}(s) = O_P(n^{-a}).$$

Proof.

Case (i): $a \in (0, 1/3)$**.** Put $\mathbb{M}_1(s) \triangleq m(g)[F(s+g) - F(g)] - [M(s+g) - M(g)]$ with $M(s) = \int_0^{F(s)} m(F^{-1}(u))\mathrm{d}u$. For $s_0 = 0$, $\mathbb{M}_1(s_0) = \mathbb{M}_{n1}(s_0) = 0$. By Taylor's theorem, there exists $\xi_{1s} \in (F(s), F(g+s))$ such that

$$\begin{aligned}
M(g+s) - M(s) &= \int_{F(g)}^{F(g+s)} m(F^{-1}(z))\mathrm{d}z \\
&= m(g)[F(g+s) - F(s)] + \frac{1}{2}\frac{m'(F^{-1}(\xi_{1s}))}{f(F^{-1}(\xi_{1s}))}(F(g+s) - F(s))^2 \\
&= m(g)[F(g+s) - F(s)] + \frac{1}{2}\frac{m'(F^{-1}(\xi_{1s}))}{f(F^{-1}(\xi_{1s}))} f^2(\xi_{2s})s^2,
\end{aligned}$$

where the second line follows by the mean-value theorem for some $\xi_{2s} \in (0, s)$. Then for every s in the neighborhood of s_0

$$\begin{aligned}
\mathbb{M}_1(s) - \mathbb{M}_1(s_0) &= m(g)[F(g+s) - F(s)] - [M(g+s) - M(s)] \\
&= -\frac{1}{2}\frac{m'(F^{-1}(\xi_{1s}))}{f(F^{-1}(\xi_{1s}))} f^2(\xi_{2s})s^2 \\
&\lesssim -s^2
\end{aligned}$$

since under Assumption 2.1 f is bounded away from zero and infinity and m' is finite in the neighborhood of zero. Next we will bound the modulus of continuity for some

$\delta > 0$

$$\mathbb{E}\sup_{|s|\leq\delta}|\mathbb{M}_{n1}(s)-\mathbb{M}_1(s)| \leq \mathbb{E}\Bigg[\sup_{|s|\leq\delta}|M_n(s+g)-M_n(g)-M(s+g)+M(g)$$
$$- [m(0)+n^{-1/3}u](F_n(s+g)-F_n(g)-F(s+g)+F(g))|\Bigg]$$
$$+ (n^{-1/3}|u|+|m(0)-m(g)|)|F(g+\delta)-F(g)|$$
$$\leq \mathbb{E}\sup_{|s|\leq\delta}|(P_n-P)h_s| + \mathbb{E}\sup_{|s|\leq\delta}|R_n(s)| + O((n^{-1/3}+g)\delta), \tag{A.6}$$

where $h_s \in \mathcal{H}_\delta = \{h_s(y,x) = (y-m(0))[\mathbb{1}_{[0,s+g]}(x) - \mathbb{1}_{[0,g]}(x)] : \ s \in [0,\delta]\}$ and

$$R_n(s) = n^{-1/3}u(F_n(s+g)-F_n(g)-F(s+g)+F(g)).$$

By the maximal inequality, Kim and Pollard (1990), p.199, the first term in the upper bound in Eq. A.6 is

$$\mathbb{E}\sup_{|s|\leq\delta}|(P_n-P)h_s| \leq n^{-1/2}J(1)\sqrt{PH_\delta^2},$$

where $J(1)$ is the uniform entropy integral, which is finite since $\mathcal{H}_\delta$ is a VC-subgraph class of functions with VC index 2, $H_\delta(y,x) = |y-m(0)|\mathbb{1}_{[g,g+\delta]}(x)$ is the envelop of $\mathcal{H}_\delta$, and

$$\begin{aligned} PH_\delta^2 &= \mathbb{E}[(\sigma^2(X)+|m(X)-m(0)|^2)\mathbb{1}_{[g,g+\delta]}(X)] \\ &= \int_g^{g+\delta}(\sigma^2(x)+|m(x)-m(0)|^2)\mathrm{d}F(x) \\ &= O(\delta). \end{aligned}$$

Next, by the maximal inequality

$$\begin{aligned} \mathbb{E}\sup_{|s|\leq\delta}|R_n(s)| &\leq n^{-1/3}|u|\mathbb{E}\sup_{|s|\leq\delta}|F_n(s+g)-F_n(g)-F(s+g)+F(g)| \\ &\leq n^{-1/3}n^{-1/2}J(1)\sqrt{PH_\delta^2}|u| \\ &= O(n^{-5/6}\delta^{1/2}), \end{aligned}$$

where $J(1) < \infty$ is the uniform entropy integral and $H_\delta(x) = \mathbb{1}_{[g,g+\delta]}(x)$ is the envelop.

Next, setting $\delta = 1$, we get

$$\sup_{s\in[-g,1-g]} |\mathbb{M}_{n1}(s) - \mathbb{M}_1(s)| = o_P(1).$$

Since $m(0) < m(x)$ and $f(x) > 0$ for all $x \in (0, 1]$, the function $s \mapsto \mathbb{M}_1(s)$ is strictly decreasing with a maximum achieved at $-g$, whence by van der Vaart and Wellner (2000), Corollary 3.2.3 (i)

$$\arg\max_{t\in[-g,1-g]} \mathbb{M}_{n1}(t) = o_P(1).$$

Then $\phi_n(\delta) = \delta^{1/2} + n^{1/6}\delta$ is a good modulus of continuity function for $a = 3/2$ and $r_n = n^{1/3}$. Indeed, for this choice $\delta \mapsto \phi_n(\delta)/\delta^a$ is decreasing and

$$r_n^2\phi_n(r_n^{-1}) = O(n^{1/2})$$

Therefore, the result follows by van der Vaart and Wellner (2000), Theorem 3.2.5.

Case (ii): $a = 1/3$**.** Put $\mathbb{M}_2(s) \triangleq m(0)F(s) - M(s)$ with $M(s) = \int_0^{F(s)} m(F^{-1}(u))\mathrm{d}u$. For $s_0 = 0$, $\mathbb{M}_2(s_0) = \mathbb{M}_{n2}(s_0) = 0$. By Taylor's theorem, there exists $\xi_{1s} \in (0, F(s))$ such that

$$\begin{aligned}
M(s) &= \int_0^{F(s)} m(F^{-1}(u))\mathrm{d}u \\
&= m(0)F(s) + \frac{1}{2}\frac{m'(F^{-1}(\xi_{1s}))}{f(F^{-1}(\xi_{1s}))}(F(s))^2 \\
&= m(0)F(s) + \frac{1}{2}\frac{m'(F^{-1}(\xi_{1s}))}{f(F^{-1}(\xi_{1s}))}f^2(\xi_{2s})s^2,
\end{aligned}$$

where the second line follows by the mean-value theorem for some $\xi_{2s} \in (0, s)$. Then

$$\begin{aligned}
\mathbb{M}_2(s) - \mathbb{M}_2(s_0) &= m(0)F(s) - M(s) \\
&= -\frac{1}{2}\frac{m'(F^{-1}(\xi_{1s}))}{f(F^{-1}(\xi_{1s}))}f^2(\xi_{2s})s^2 \\
&\lesssim -s^2.
\end{aligned}$$

Next we will bound the modulus of continuity for some $\delta > 0$

$$\begin{aligned}
\mathbb{E}\sup_{|s|\leq\delta} |\mathbb{M}_{n2}(s) - \mathbb{M}_2(s)| &\leq \mathbb{E}\sup_{|s|\leq\delta} \left|M_n(s) - M(s) - (m(0) + n^{(a-1)/2}u)(F_n(s) - F(s))\right| \\
&\quad + n^{(a-1)/2}|u|F(\delta) \\
&\leq \mathbb{E}\sup_{|s|\leq\delta} |(P_n - P)g_s| + \mathbb{E}\sup_{|s|\leq\delta} |R_n(s)| + O(n^{(a-1)/2}\delta),
\end{aligned} \tag{A.7}$$

where $g_s \in \mathcal{G}_\delta = \{g_s(y,x) = (y - m(0))\mathbb{1}_{[0,s]}(x) : \; s \in [0,\delta]\}$ and

$$R_n(s) = n^{(a-1)/2}u(F_n(s) - F(s).$$

By the maximal inequality Kim and Pollard (1990), p.199, the first term in the upper bound in Eq. A.7 is

$$\mathbb{E}\sup_{|s|\leq\delta}|(P_n - P)g_s| \leq n^{-1/2}J(1)\sqrt{PG_\delta^2},$$

where $J(1)$ is the uniform entropy integral, which is finite since $\mathcal{G}_\delta$ is a VC-subgraph class of functions with VC-index 2, $G_\delta(y,x) = |y - m(0)|\mathbb{1}_{[0,\delta]}(x)$ is the envelop of $\mathcal{G}_\delta$, and

$$\begin{aligned} PG_\delta^2 &= \mathbb{E}[\sigma^2(X)\mathbb{1}_{[0,\delta]}(X)] + \mathbb{E}[|m(X) - m(0)|^2\mathbb{1}_{[0,\delta]}(X)] \\ &= \int_0^\delta (\sigma^2(x) + |m(X) - m(0)|^2)\mathrm{d}F(x) \\ &= O(\delta). \end{aligned}$$

Next, by the maximal inequality

$$\begin{aligned} \mathbb{E}\sup_{|s|\leq\delta}|R_n(s)| &\leq n^{(a-1)/2}|u|\mathbb{E}\sup_{|s|\leq\delta}|F_n(s) - F(s)| \\ &\leq n^{(a-1)/2}n^{-1/2}J(1)\sqrt{PH_\delta^2}|u| \\ &= O(n^{(a-2)/2}\delta^{1/2}), \end{aligned}$$

where $J(1) < \infty$ is the uniform entropy integral and $H_\delta(x) = \mathbb{1}_{[0,\delta]}(x)$ is the envelop.

Next, setting $\delta = 1$, we get

$$\sup_{s\in[0,1]}|\mathbb{M}_{n2}(s) - \mathbb{M}_2(s)| = o_P(1).$$

Since $m(0) < m(x)$ and $f(x) > 0$ for all $x \in (0,1]$, the function $s \mapsto \mathbb{M}_2(s)$ is strictly decreasing with maximum achieved at 0, whence by van der Vaart and Wellner (2000), Corollary 3.2.3 (i)

$$\arg\max_{t\in[0,1]}\mathbb{M}_{n2}(t) = o_P(1).$$

Then the modulus of continuity is $\phi_n(\delta) = \delta^{1/2} + n^{a/2}\delta$. This is a good modulus of the continuity function for $\alpha = 3/2$ and $r_n = n^{1/3}$. For this choice $\delta \mapsto \phi_n(\delta)/\delta^\alpha$ is decreasing and $r_n^2\phi_n(r_n^{-1}) = O(n^{1/2})$. Therefore, the result follows by van der Vaart and Wellner (2000), Theorem 3.2.5.

Case (iii): $a \in (1/3, 1]$ Here van der Vaart and Wellner (2000), Theorem 3.2.5 gives the order $O_P(n^{-1/3})$ only, so we will show directly using the "peeling device" that after the change of variables

$$\arg\max_{s \in [0, n^a]} \left\{ n^{(1-a)/2} \left(m(0) F_n(sn^{-a}) - M_n(sn^{-a}) \right) + u F_n(sn^{-a}) \right\} = O_P(1).$$

Denote the process inside of the argmax as

$$Z_{n2}(s) \triangleq n^{(1-a)/2} \left(m(0) F_n(sn^{-a}) - M_n(sn^{-a}) \right) + n^a u F_n(sn^{-a}).$$

Decompose $Z_{n2} = I_{n2} + II_{n2} + III_{n2} + IV_{n2}$ similarly as in the proof of Theorem 2.1 (with $c = 1$). Next, partition the set $[0, \infty)$ into intervals $S_j = \{s : 2^{j-1} < s \leq 2^j\}$ with j ranging over integers. Then if the argmax exceeds 2^K, it will be located in one of the intervals S_j with $j \geq K$ and $2^{j-1} \leq n^a$. Therefore, using the fact that $u < 0, II_{n2} \leq 0$, and $Z_{n2}(0) = 0$

$$\begin{aligned}
\Pr\left(\arg\max_{s \in [0, n^a]} Z_{n2}(s) > 2^K \right) &\leq \sum_{\substack{j \geq K \\ 2^{j-1} \leq n^a}} \Pr\left(\sup_{s \in S_j} Z_{n2}(s) \geq 0 \right) \\
&\leq \sum_{\substack{j \geq K \\ 2^{j-1} \leq n^a}} \Pr\left(\sup_{s \in S_j} |I_{n2}(s) + III_{n2}(s)| \geq -n^a u F(2^j n^{-a}) \right) \\
&\leq \sum_{\substack{j \geq K \\ 2^{j-1} \leq n^a}} \frac{1}{-un^a F(2^j n^{-a})} \mathbb{E}\left[\sup_{s \in S_j} |I_{n2}(s) + III_{n2}(s)| \right] \\
&\lesssim \sum_{\substack{j \geq K \\ 2^{j-1} \leq n^a}} \frac{1}{-un^a F(2^j n^{-a})} \left\{ 2^{j/2} + n^{(a-1)/2} 2^{j/2} \right\} \\
&\lesssim \sum_{j \geq K} 2^{-j/2},
\end{aligned}$$

where the third line follows by Markov's inequality and the fourth by the maximal inequality as in the proof of Theorem 2.1. The last expression can be made arbitrarily small by the choice of K. □

Proof of Remark 2.1. We sketch only the most important differences below:

$$\begin{aligned}
&\Pr\left(n^{(1-a)/2}\left(\hat{m}(-cn^{-a}) - m(0)\right) \leq u\right) \\
&= \Pr\left(\hat{m}(-cn^{-a}) \leq un^{(a-1)/2} + m(0)\right) \\
&= \Pr\left(\operatorname*{arg\,max}_{s\in[-1,0]}\left\{\left(n^{(a-1)/2}u + m(0)\right)F_n(s) - M_n(s)\right\} \geq -cn^{-a}\right) \\
&= \Pr\left(\operatorname*{arg\,max}_{t\in[-n^a/c,0]}\left\{\left(n^{(a-1)/2}u + m(0)\right)F_n(cn^{-a}t) - M_n(cn^{-a}t)\right\} \geq -1\right) \\
&\to \Pr\left(\operatorname*{arg\,max}_{t\in(-\infty,0]}\left\{ut - \sqrt{\frac{\sigma^2(0)}{cf(0)}}W_t - \frac{t^2c}{2}m'(0)\mathbb{1}_{a=1/3}\right\} \geq -1\right) \\
&= \Pr\left(D^L_{(-\infty,0]}\left(\sqrt{\frac{\sigma^2(0)}{cf(0)}}W_t + \frac{t^2c}{2}m'(0)\mathbb{1}_{a=1/3}\right)(-1) \leq u\right)
\end{aligned}$$

□

Proof of Theorem 2.2. Put

$$M_n^*(t) = \frac{1}{n}\sum_{i=1}^{n} Y_i^*\mathbb{1}\{X_i \leq t\}$$

and note that now $\hat{m}^*(x)$ 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 $t = x$. By the argument similar to the one used in the proof of Theorem 2.1 for every $u < 0$

$$\begin{aligned}
&\Pr{}^*\left(n^{(1-a)/2}\left(\hat{m}^*(cn^{-a}) - \hat{m}(cn^{-a})\right) \leq u\right) \\
&= \Pr{}^*\left(\operatorname*{arg\,max}_{t\in[0,n^a/c]}\left\{(n^{(a-1)/2}u + \hat{m}(cn^{-a}))F_n(cn^{-a}t) - M_n^*(cn^{-a}t)\right\} \geq 1\right).
\end{aligned}$$

The location of the argmax is the same as the location of the argmax of the following process

$$Z_n^*(t) \triangleq I_n^*(t) + II_n^*(t) + III_n^*(t) + IV_n^*(t)$$

with

$$
\begin{aligned}
I_n^*(t) &= -n^{(a-1)/2}\sum_{i=1}^n \eta_i^*\varepsilon_i \mathbb{1}_{[0,cn^{-a}t]}(X_i)\\
II_n^*(t) &= n^{(a-1)/2}\sum_{i=1}^n \eta_i^*(\tilde{m}(X_i)-m(X_i))\mathbb{1}_{[0,cn^{-a}t]}(X_i)\\
III_n^*(t) &= n^{(1+a)/2}\int_0^{cn^{-a}t}(\hat{m}(cn^{-a})-\tilde{m}(x))\mathrm{d}F_n(x)\\
IV_n^*(t) &= n^a u F_n(cn^{-a}t)
\end{aligned}
$$

The process I_n^* is the multiplier empirical process indexed by the class of functions

$$
\mathcal{G}_n = \left\{(\epsilon, x)\mapsto -n^{a/2}\epsilon\mathbb{1}_{[0,cn^{-a}t]}(x):\quad t\in[0,K]\right\}.
$$

This class is of VC subgraph type with VC index 2 and envelop $G_n(\epsilon,x) = n^{a/2}|\epsilon|\mathbb{1}_{[0,cn^{-a}K]}(x)$, which is square-integrable

$$
PG_n^2 = n^a\int_0^{cn^{-a}K}\sigma^2(x)\mathrm{d}F(x) = O(1).
$$

This envelop satisfies Lindeberg's condition for every $\eta > 0$

$$
\begin{aligned}
\mathbb{E}G_n^2\{G_n>\eta\sqrt{n}\} &\leq \frac{n^{\frac{a(2+\delta)}{2}}\mathbb{E}\left[|\varepsilon|^{2+\delta}\mathbb{1}_{[0,cn^{-a}K]}(X)\right]}{\eta^\delta n^{\delta/2}}\\
&= O(n^{\delta(a-1)/2})\\
&= o(1)
\end{aligned}
$$

and for every $g_{n,t}, g_{n,s}\in\mathcal{G}_n$ and $\delta_n\to 0$

$$
\begin{aligned}
\sup_{|t-s|\leq\delta_n}\mathbb{E}|g_{n,t}-g_{n,s}|^2 &= n^a\sup_{|t-s|\leq\delta_n}\mathbb{E}\left[\varepsilon^2\mathbb{1}_{[cn^{-a}s,cn^{-a}t]}(X)\right]\\
&= O(\delta_n)\\
&= o(1).
\end{aligned}
$$

Next, we show that the covariance structure is

$$
\begin{aligned}
\mathbb{E}^*[I_n^*(t)I_n^*(s)] &= n^{a-1}\sum_{i=1}^n\varepsilon_i^2\mathbb{1}_{[0,cn^{-a}(t\wedge s)]}(X_i)\\
&= n^a\mathbb{E}[\varepsilon^2\mathbb{1}_{[0,cn^{-a}(t\wedge s)]}(X_i)] + R_n(t,s)
\end{aligned}
$$

with

$$R_n(t,s) = n^{a-1}\sum_{i=1}^n \varepsilon_i^2 \mathbb{1}_{[0,cn^{-a}(t\wedge s)]}(X_i) - n^a \mathbb{E}\left[\varepsilon^2 \mathbb{1}_{[0,cn^{-a}(t\wedge s)]}(X_i)\right]$$

Since $\mathbb{E}[\varepsilon^4|X] \leq C$, the variance of R_n tends to zero

$$\begin{aligned}
\operatorname{Var}(R_n(t,s)) &= n^{2a-1}\operatorname{Var}(\varepsilon^2 \mathbb{1}_{[0,cn^{-a}(t\wedge s)]}(X)) \\
&\leq n^{2a-1}\mathbb{E}[\varepsilon^4 \mathbb{1}_{[0,cn^{-a}(t\wedge s)]}(X))] \\
&\leq Cn^{2a-1}F(cn^{-a}(t\wedge s)) \\
&= O(n^{a-1}) \\
&= o(1),
\end{aligned}$$

whence by Chebyshev's inequality $R_n(t,s) = o_P(1)$. Therefore, the covariance structure converges pointwise to the one of the scaled Brownian motion

$$\begin{aligned}
\mathbb{E}^*[I_n^*(t)I_n^*(s)] &= n^a \int_0^{cn^{-a}(t\wedge s)} \sigma^2(x)\mathrm{d}F(x) + o_P(1) \\
&= \sigma^2(0)f(0)c(t\wedge s) + o_P(1),
\end{aligned}$$

where $\mathbb{E}^*[.] = \mathbb{E}[.|(Y_i,X_i)_{i=1}^n]$. By van der Vaart and Wellner (2000), Theorem 2.11.22, the class $\mathcal{G}_n$ is Donsker, whence by the multiplier central limit theorem, van der Vaart and Wellner (2000), Theorem 2.9.6

$$\sup_{h\in BL_1(\ell^\infty[0,K])} \left|\mathbb{E}^* h(I_n^*) - \mathbb{E}h\left(\sqrt{\sigma^2(0)f(0)c}W_t\right)\right| \xrightarrow{P} 0.$$

Next, II_n^* is a multiplier empirical process indexed by the degenerate class of functions

$$\mathcal{H}_n = \left\{x \mapsto n^{a/2}\left(\tilde{m}(x) - m(x)\right)\mathbb{1}_{[0,n^{-a}t]}(x) :\ t\in[0,K]\right\}.$$

Since this class is of VC subgraph type with VC index 2, by the maximal inequality

$$\begin{aligned}
\mathbb{E}^*\left[\sup_{t\in[0,K]}|II_n^*(t)|\right] &\lesssim \sqrt{n^a\int_0^{cn^{-a}K}|\tilde{m}(x)-m(x)|^2\mathrm{d}F_n(x)} \\
&= \sqrt{n^a\int_0^K |\tilde{m}(cn^{-a}y) - m(cn^{-a}y)|^2\mathrm{d}F_n(cn^{-a}y)} \\
&= \sqrt{o_P(1)n^aF_n(cn^{-a}K)} \\
&= o_P(1),
\end{aligned}$$

where we apply Proposition A.2.1.

Next, changing variables $x \mapsto cn^{-a}y$ and using the fact that $\hat{m}$ is non-decreasing

$$
\begin{aligned}
III_n^*(t) &= n^{(1+a)/2} \int_0^t (\hat{m}(cn^{-a}) - \tilde{m}(cn^{-a}y)) \mathrm{d}F_n(cn^{-a}y) \\
&\leq n^{(1+a)/2} \int_0^1 (\hat{m}(cn^{-a}) - \tilde{m}(cn^{-a}y)) \mathrm{d}F_n(cn^{-a}y) \\
&\leq n^{(1+a)/2} \sup_{y\in[0,1]} |\hat{m}(cn^{-a}) - \tilde{m}(cn^{-a}y)| F_n(cn^{-a}) \\
&= o_P(1) \left(n^{(1+a)/2}(F_n(cn^{-a}) - F(cn^{-a})) + n^{(1+a)/2}F(cn^{-a})\right) \\
&= o_P(1) \left(O_P(1) + O(n^{(1-a)/2})\right) \\
&= o_P(1),
\end{aligned}
$$

where the fourth line follows by Proposition A.2.1 and Theorem 2.1 (ii), and the fifth since the variance of the term inside is $O(1)$.

Next

$$IV_n^*(t) = utf(0)c + o_P(1)$$

in the same way we treat $III_{n2} + IV_{n2}$ in the proof of Theorem 2.1.

Combining all estimates obtained above together

$$
\begin{aligned}
&\sup_{h\in BL_1(\ell^\infty[0,K])} \left|\mathbb{E}^* h(Z_n^*) - \mathbb{E}h\left(utf(0)c + \sqrt{\sigma^2(0)f(0)c}W_t\right)\right| \\
&= \sup_{h\in BL_1(\ell^\infty[0,K])} |\mathbb{E}^* h(Z_n^*) - \mathbb{E}^* h(utf(0)c + I_n^*)| \\
&\quad + \sup_{h\in BL_1(\ell^\infty[0,K])} \left|\mathbb{E}^* h(utf(0)c + I_n^*) - \mathbb{E}h\left(utf(0)c + \sqrt{\sigma^2(0)f(0)c}W_t\right)\right| \\
&\leq \sup_{t\in[0,K]} |II_n^*(t) + III_n^*(t) + IV_n^*(t)| + o_P(1) \\
&= o_P(1).
\end{aligned}
$$

By Lemma A.2.2, the argmax of $Z_n^*(t)$ is uniformly tight, so by Lemma A.2.3

$$\Pr^*\left(n^{(1-a)/2}\left(\hat{m}^*(cn^{-a}) - \hat{m}(cn^{-a})\right) \leq u\right) \xrightarrow{P} \Pr\left(D^L_{[0,\infty)}\left(\sqrt{\frac{\sigma^2(0)}{cf(0)}}W_t\right)(1) \leq u\right).$$

To conclude, it remains to show that when $a \in (1/3, 1)$

$$\Pr^*\left(n^{(1-a)/2}\left(\hat{m}^*(cn^{-a}) - \hat{m}(cn^{-a})\right) \geq 0\right) \to 0.$$

This follows since for every $\epsilon > 0$

$$
\begin{aligned}
\Pr^*\left(n^{(1-a)/2}\left(\hat{m}^*(cn^{-a}) - \hat{m}(cn^{-a})\right) \geq 0\right) &\leq \Pr^*\left(n^{(1-a)/2}\left(\hat{m}^*(cn^{-a}) - \hat{m}(cn^{-a})\right) \geq -\epsilon\right) \\
&\xrightarrow{P} \Pr\left(\operatorname*{arg\,max}_{t\in[0,\infty)}\left\{-\epsilon t - \sqrt{\frac{\sigma^2(0)}{cf(0)}}W_t\right\} \leq c\right) \\
&= \Pr\left(\operatorname*{arg\,max}_{t\in[0,\infty)}\{W_t - t\} \leq c^2\frac{f(0)}{\sigma^2(0)}\epsilon\right),
\end{aligned}
$$

which tends to zero as $\epsilon \downarrow 0$; see the proof of Theorem 2.1. The result follows from Theorem 2.1 (ii) with $a > 1/3$. $\square$

The following lemma shows tightness of the argmax of the bootstrapped process:

$$
\mathbb{M}_n^*(s) \triangleq n^{(1-a)/2}\left(\hat{m}(n^{-a})F_n(sn^{-a}) - M_n^*(sn^{-a})\right) + n^a uF_n(sn^{-a}), \quad s \in [0, n^a].
$$

Lemma A.2.2. *Suppose that assumptions of Theorem 2.2 are satisfied. Then for every $a \in (1/3, 1]$ and $u < 0$*

$$
\operatorname*{arg\,max}_{s\in[0,n^a]}\mathbb{M}_n^*(s) = O_P(1).
$$

Proof. Decompose $\mathbb{M}_n^* = I_n^* + II_n^* + III_n^* + IV_n^*$ similarly to the proof of Theorem 2.2 (with $c = 1$). Next, partition the set $[0,\infty)$ into intervals $S_j = \{s: 2^{j-1} < s \leq 2^j\}$ with j ranging over integers. Let $\|.\|_{S_j}$ be the supremum norm over S_j. Then if the argmax exceeds 2^K, it will be located in one of the intervals S_j with $j \geq K$ and

$2^{j-1} \leq n^a$. Therefore, using the fact that $u < 0, II_{n2} \leq 0$, and $\mathbb{M}_n^*(0) = 0$

$$
\begin{aligned}
&\Pr{}^*\left(\operatorname*{arg\,max}_{s\in[0,n^a]} \mathbb{M}_n^*(s) > 2^K\right) \\
&\leq \sum_{\substack{j\geq K\\ 2^{j-1}\leq n^a}} \Pr{}^*\left(\sup_{s\in S_j} \mathbb{M}_n^*(s) \geq 0\right) \\
&\leq \sum_{\substack{j\geq K\\ 2^{j-1}\leq n^a}} \Pr{}^*\left(\left\|I_n^* + II_n^* + III_n^* + IV_n^* - n^a uF(.n^{-a})\right\|_{S_j} \geq -n^a uF(2^j n^{-a})\right) \\
&\leq \sum_{\substack{j\geq K\\ 2^{j-1}\leq n^a}} \frac{1}{-un^aF(2^jn^{-a})}\mathbb{E}^*\left\|I_n^* + II_n^* + III_n^* + IV_n^* - n^a uF(.n^{-a})\right\|_{S_j} \\
&\lesssim \sum_{\substack{j\geq K\\ 2^{j-1}\leq n^a}} \frac{1}{-un^aF(2^jn^{-a})} 2^{j/2} O_P(1) \\
&\lesssim \sum_{j\geq K} 2^{-j/2} O_P(1),
\end{aligned}
$$

where the third line follows by Markov's inequality and the fourth by computations below. The last expression is $o_P(1)$ for every $K = K_n \to \infty$. To see that all terms above are controlled as was claimed, first note that the process I_n^* is a multiplier empirical process indexed by the class of functions

$$
\mathcal{G}_n = \left\{(\epsilon, x) \mapsto -n^{a/2}\epsilon \mathbb{1}_{[0,n^{-a}t]}(x) : \quad t \in S_j\right\}.
$$

This class is of VC subgraph type with VC index 2, whence by the maximal inequality

$$
\begin{aligned}
\mathbb{E}^*\left[\sup_{s\in S_j} |I_n^*(t)|\right] &\lesssim n^{a/2}\sqrt{\frac{1}{n}\sum_{i=1}^n \varepsilon_i^2 \mathbb{1}_{[0,n^{-a}2^j]}(X_i)} \\
&= \sqrt{n^a \mathbb{E}[\varepsilon^2 \mathbb{1}_{[0,n^{-a}2^j]}(X)] + o_P(1)} \\
&= O_P(2^{j/2}),
\end{aligned}
$$

where the second line follows since $\mathbb{E}[\varepsilon^4|X] \leq C$. Next, it follows from the proof of Theorem 2.2 (replacing K by 2^j) that

$$
\mathbb{E}^*\left[\sup_{s\in S_j} |II_n^*(t)|\right] = o_P(2^{j/2})
$$

and that $\|III_n^*\|_{S_j} = o_P(1)$. Lastly, by the maximal inequality

$$\sup_{s\in S_j} |IV_n^*(s) - n^a uF(n^{-a}s)| = O_P(n^{(a-1)/2}2^{j/2}).$$

$\square$

The following lemma is a conditional argmax continuous mapping theorem for bootstrapped processes.

Lemma A.2.3. *Suppose that for every $K < \infty$*

(i)

$$\sup_{h\in BL_1(\ell^\infty[0,K])} |\mathbb{E}^* h(Z_n^*) - \mathbb{E}h(Z)| \xrightarrow{P} 0,$$

(ii)

$$\limsup_{n\to\infty} \Pr^*\left(\arg\max_{t\in[0,n^a]} Z_n^*(t) > K\right) = o_P(1), \qquad K \to \infty.$$

(iii) $t \mapsto Z(t)$ has unique maximizer on $[0,\infty)$, which is a tight random variable.

Then

$$\Pr^*\left(\arg\max_{t\in[0,n^a]} Z_n^*(t) \geq z\right) \xrightarrow{P} \Pr\left(\arg\max_{t\in[0,\infty)} Z(t) \geq z\right), \qquad \forall z > 0.$$

Proof. For every K

$$\Pr^*\left(\arg\max_{t\in[0,n^a]} Z_n^*(t) \geq z\right) = \Pr^*\left(\arg\max_{t\in[0,K]} Z_n^*(t) \geq z\right) + R_{n,K},$$

where by (ii)

$$\begin{aligned}
\limsup_{n\to\infty} R_{n,K} &= \limsup_{n\to\infty} \Pr^*\left(\arg\max_{t\in[0,K]} Z_n^*(t) < z, \arg\max_{t\in[K,n^a]} Z_n^*(t) \geq z\right) \\
&\leq \limsup_{n\to\infty} \Pr^*\left(\arg\max_{t\in[0,n^a]} Z_n^*(t) > K\right) \\
&= o_P(1), \qquad K \to \infty,
\end{aligned}$$

by (i) and (iii)

$$\begin{aligned}
\Pr^*\left(\arg\max_{t\in[0,K]} Z_n^*(t) \geq z\right) &= \Pr\left(\arg\max_{t\in[0,K]} Z(t) \geq z\right) + o_P(1) \\
&= \Pr\left(\arg\max_{t\in[0,\infty)} Z(t) \geq z\right) + o_P(1), \qquad K\to\infty.
\end{aligned}$$

More precisely, we used the continuous mapping theorem for the bootstrapped process Kosorok (2008b), Proposition 10.7:

$$\begin{aligned}
\Pr^*\left(\arg\max_{t\in[0,K]} Z_n^*(t) \geq z\right) &= \Pr^*\left(\sup_{t\in[z,K]} Z_n^*(t) \geq \sup_{t\in[0,K]} Z_n^*(t)\right) \\
&\xrightarrow{P} \Pr\left(\sup_{t\in[z,K]} Z(t) \geq \sup_{t\in[0,K]} Z(t)\right) \\
&= \Pr\left(\arg\max_{t\in[0,K]} Z(t) \geq z\right),
\end{aligned}$$

where the convergence is actually uniform over z in arbitrary closed subset of the set of continuity points of $z \mapsto \Pr\left(\arg\max_{t\in[0,K]} Z(t) \geq z\right)$; see Kosorok (2008b), Lemma 10.11. □

The following result is is probabilistic statement of the fact that for monotone functions converging pointwise to a continuous limit we also have the uniform convergence.

Proposition A.2.1. *Suppose that assumptions of Theorem 2.1 are satisfied. If m is continuous on $[0,1]$, then*

$$\sup_{y\in[0,1]} |\tilde{m}(cn^{-a}y) - m(0)| \xrightarrow{P} 0.$$

Proof. For every $y \in [0,1]$, by Theorem 2.1

$$|\tilde{m}(cn^{-a}y) - m(0)| \xrightarrow{P} 0.$$

Since m is uniformly continuous, one can find $0 \leq y_1 \leq \cdots \leq y_p \leq 1$ such that $|m(cn^{-a}y_j) - m(cn^{-a}y_{j-1})| < \epsilon/2$ for all $j = 2,\dots,p$. Then on the event

$\{|\tilde{m}(cn^{-a}y_j) - m(cn^{-a}y_j)| \leq \epsilon/2,\ \forall j = 1, \dots, p\}$ by the monotonicity of $\tilde{m}$, for every x, there exists $j = 2, \dots, p$ such that

$$m(cn^{-a}x) - \epsilon \leq \tilde{m}(cn^{-a}y_{j-1}) \leq \tilde{m}(cn^{-a}x) \leq \tilde{m}(cn^{-a}y_j) \leq m(cn^{-a}x) + \epsilon,$$

whence

$$\Pr\left(|\tilde{m}(cn^{-a}y) - m(cn^{-a}y)| \leq \epsilon, \forall y \in [0,1]\right) \geq 1 - \sum_{j=1}^{p} \Pr\left(|\tilde{m}(cn^{-a}y_j) - m(cn^{-a}y_j)| > \epsilon/2\right).$$

Since p is fixed, the sum of probabilities tends to zero by the pointwise consistency of $\tilde{m}$, which gives the result as $\epsilon > 0$ is arbitrary. $\square$

Proof of Proposition A.4.1. Since treatment of non-decreasing and non-increasing cases is similar, we focus only on the former. Under (M1), Rudin (1976), Theorem 4.29, ensures that all limits in Eq. (1) and Eq. (2) exist. In the sharp design

$$\begin{aligned}
\theta &= \mathbb{E}[Y_1 - Y_0 | X = x_0] \\
&= \lim_{x \downarrow x_0} \left(\mathbb{E}[Y_1|X = x] - \mathbb{E}[Y_0|X = -x]\right) \\
&= \lim_{x \downarrow x_0} \mathbb{E}[Y|X = x] - \lim_{x \uparrow x_0} \mathbb{E}[Y|X = x],
\end{aligned}$$

where the second line follows under Assumption (OC), and the third since for any $x > 0$

$$\mathbb{E}[Y|X = x] = \mathbb{E}[Y_1|X = x] \qquad \text{and} \qquad \mathbb{E}[Y|X = -x] = \mathbb{E}[Y_0|X = -x],$$

which itself is a consequence of $Y = DY_1 + (1 - D)Y_0$ and $D = \mathbb{1}\{X \geq x_0\}$.

Now suppose that (M1) and (M2) are satisfied and that θ coincides with the expression in Eq. 1. Then under (M1)

$$\mathbb{E}[Y_1|X = c] \leq \lim_{x \downarrow x_0} \mathbb{E}[Y_1|X = x] = \lim_{x \downarrow x_0} \mathbb{E}[Y|X = x]$$

and

$$\lim_{x \uparrow x_0} \mathbb{E}[Y|X = x] = \lim_{x \uparrow x_0} \mathbb{E}[Y_0|X = x] \leq \mathbb{E}[Y_0|X = x_0].$$

Combining the two inequalities under (M2)

$$\lim_{x \uparrow x_0} \mathbb{E}[Y|X = x] \leq \mathbb{E}[Y_0|X = x_0] \leq \mathbb{E}[Y_1|X = x_0] \leq \lim_{x \downarrow x_0} \mathbb{E}[Y|X = x].$$

Finally, θ is defined as the difference of inner quantities and also equals to the difference of outer quantities by assumption, which is possible only if (OC) holds, i.e,

$$\lim_{x\uparrow x_0}\mathbb{E}[Y|X=x]=\mathbb{E}[Y_0|X=x_0]\leq\mathbb{E}[Y_1|X=x_0]=\lim_{x\downarrow x_0}\mathbb{E}[Y|X=x].$$

The proof for the fuzzy design is similar to the proof of Hahn, Todd, and Van der Klaauw (2001), Theorem 2, with the only difference that monotonicity ensures existence of limits, so their Assumption (RD), (i) can be dropped. □

Proof of Theorem 3.1. Since

$$n^{1/3}(\hat{\theta}-\theta)=n^{1/3}\left(\hat{m}_+(cn^{-1/3})-m_+\right)-n^{1/3}\left(\hat{m}_-(-cn^{-1/3})-m_-\right),$$

the proof is similar to the proof of Theorem 2.1 and Remark 2.1 with $a=1/3$. Strictly speaking, the proof of Theorem 2.1 and Remark 2.1 change a little. Now $F(0)\neq 0$ and we will have $\tilde{F}(x)=F(x)-F(0)$ and $\tilde{F}_n(x)=F_n(x)-F_n(0)$ instead of $F(x)$ and $F_n(x)$ everywhere in the proof of Theorem 2.1, which will allow us to proceed as before. In the proof of Remark 2.1, we will have $\tilde{F}(x)=F(0)-F(x)$ and $\tilde{F}_n(x)=F_n(0)-F_n(x)$ instead of $F(x)$ and $F_n(x)$. The independence of W_t^+ and W_t^- follows from the independence of two samples. □

Proof of Theorem 3.2. Put $\hat{g}=\hat{m}_+\left(cn^{-1/3}\right)-\hat{m}_-\left(-cn^{-1/3}\right)$, $\hat{h}=\hat{p}_+\left(cn^{-1/3}\right)-\hat{p}_-\left(-cn^{-1/3}\right)$, $g=m_+-m_-$, and $h=p_+-p_-$. By a similar argument as in the proof of Theorem 2.1 and Theorem 3.1

$$n^{1/3}(\hat{g}-g)\xrightarrow{\mathrm{d}}\xi_1\qquad\text{and}\qquad n^{1/3}(\hat{h}-h)\xrightarrow{\mathrm{d}}\xi_2.$$

Consequently,

$$\begin{aligned}n^{1/3}(\hat{\theta}^F-\theta)&=n^{1/3}\left(\frac{\hat{g}}{\hat{h}}-\frac{g}{h}\right)\\&=\frac{n^{1/3}(\hat{g}-g)h-n^{1/3}(\hat{h}-h)g}{h\hat{h}}\\&\xrightarrow{\mathrm{d}}\frac{1}{h}\xi_1-\frac{g}{h^2}\xi_2.\end{aligned}$$

The independence between processes with different signs follows from the independence of two samples, while the covariance structure for processes with the same sign follows from the proof of Theorem 2.1 and

$$\begin{aligned}&\mathrm{Cov}\left(n^{a/2}\varepsilon_i\mathbb{1}_{[0,cn^{-a}t]}(X_i),n^{a/2}(D_i-p(X_i))\mathbb{1}_{[0,cn^{-a}s]}(X_i)\right)\\=&n^a\int_0^{cn^{-a}(t\wedge s)}\mathbb{E}\left[\varepsilon(D-p(X))|X=x\right]f(x)\mathrm{d}x\\\to&\rho_+cf_+(t\wedge s)\end{aligned}$$

and similar computations for negative observations. □

Proof of Theorem 3.3. The result follows from Theorem 2.2. □

Proposition A.2.2. *Suppose that Assumptions of Theorem 2.1 are satisfied. Then for arbitrary $c > 0$*

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

where $A = \left(\frac{2}{m'(0)}\sqrt{\frac{\sigma^2(0)}{f(0)}}\right)^{2/3}$.

Proof. By Theorem 2.1 (ii), for $B > 0$

$$\begin{aligned}
&\Pr\left(n^{1/3}(\hat{m}(cAn^{-1/3}) - m(0) \leq uB\right) \to \\
&\to \Pr\left(D^L_{[0,\infty)}\left(\sqrt{\frac{\sigma^2(0)}{cAf(0)}}W_t + \frac{t^2cA}{2}m'(0)\right)(1) \leq uB\right) \\
&= \Pr\left(\underset{t\in[0,\infty)}{\arg\max}\left\{uBt - \sqrt{\frac{\sigma^2(0)}{cAf(0)}}W_t - \frac{t^2cA}{2}m'(0)\right\} \geq 1\right) \\
&= \Pr\left(\underset{t\in[0,\infty)}{\arg\max}\left\{ut - \frac{1}{B}\sqrt{\frac{\sigma^2(0)}{Af(0)}}W_t - \frac{t^2A}{2B}m'(0)\right\} \geq c\right) \\
&= \Pr\left(\underset{t\in[0,\infty)}{\arg\max}\left\{ut - W_t - t^2\right\} \geq c\right) \\
&= \Pr\left(D^L_{[0,\infty)}(W_t - t^2)(c) \leq u\right)
\end{aligned}$$

where the first equality follows by the switching relation, the second by the change of variables Brownian scaling and invariance of argmax to the scaling, the third by plugging-in corresponding value of A and $B = \left|\frac{m'(0)\sigma^2(0)}{2f(0)}\right|^{1/3}$, and the last by another application of the switching relation. □

A.3 Additional Monte Carlo experiments

In this section we report results of Monte Carlo experiments for a larger class of data generating processes. More precisely, we augment the baseline DGPs 1-4 with four additional DGPs featuring 1) non-monotone functions; 2) functions that are steep at the cutoff; 3) functions that change the slope at the cutoff. The simulations designs are as follows

- DGP 5 sets $X \sim 2 \times \text{Beta}(2,2) - 1$ and

$$m(x) = \begin{cases} 1 - \exp(4x) & x \geq 0 \\ x/5 & x < 0; \end{cases}$$

- DGP 6 sets $X \sim 2 \times \text{Beta}(0.5, 0.5) - 1$ and

$$m(x) = \begin{cases} 1 - \exp(4x) & x \geq 0 \\ x/5 & x < 0; \end{cases}$$

- DGP 7 sets $X \sim 2 \times \text{Beta}(2,2) - 1$ and

$$m(x) = \begin{cases} -\exp(-1.25) + \exp(-(x - .5)^2) & x \geq .5 \\ -\exp(-1.25) + \exp(-5(x - .5)^2) & 0 \leq x < .5 \\ x/5 & x < 0; \end{cases}$$

- DGP 8 sets $X \sim 2 \times \text{Beta}(0.5, 0.5) - 1$ and

$$m(x) = \begin{cases} -\exp(-1.25) + \exp(-(x - .5)^2) & x \geq .5 \\ -\exp(-1.25) + \exp(-5(x - .5)^2) & 0 \leq x < .5 \\ x/5 & x < 0; \end{cases}$$

Note that in DGPs 5-6, the regression function is increasing before the cutoff and steeply decreasing after the cutoff. The DGPs 7-8 feature the regression function that is steeply increasing after the cutoff for all $x \in [0, 0.5)$ and decreasing subsequently on $[0.5, 1]$. These DGPs are the most difficult because they violate the monotonicity constraint and at the same time feature steep regression functions.[21]

We compare the performance of our iRDD estimator to the local polynomial estimator in Table A.1 (homoskedasticity) and Table A.2 (heteroskedasticity). The results show 2-10 fold reduction in the MSE across specifications with the most improvement achieved when the sample size is small.

[21] As we explained in Section 3, since the isotonic regression estimator is the constrained least-squares estimator, we may expect the projection interpretation under the misspecification.

Table A.1: MC experiments: homoskedasticity

	n	iRDD Bias	iRDD Var	iRDD MSE	LP Bias	LP Var	LP MSE
DGP 5							
	200	-0.1340	0.0440	0.0620	0.0193	0.3366	0.3369
	500	-0.1190	0.0200	0.0340	0.0446	0.1097	0.1117
	1000	-0.1020	0.0110	0.0220	0.0427	0.0559	0.0577
DGP 6							
	200	-0.3080	0.0790	0.1740	0.0272	0.8498	0.8503
	500	-0.2380	0.0380	0.0950	0.0559	0.2156	0.2186
	1000	-0.1910	0.0220	0.0590	0.0678	0.1041	0.1086
DGP 7							
	200	-0.1319	0.0409	0.0583	-0.0170	0.3194	0.3197
	500	-0.1131	0.0185	0.0313	-0.0114	0.1116	0.1117
	1000	-0.0986	0.0104	0.0202	-0.0026	0.0535	0.0535
DGP 8							
	200	-0.2891	0.0701	0.1537	-0.0061	0.7805	0.7804
	500	-0.2243	0.0344	0.0847	-0.0247	0.2284	0.2290
	1000	-0.1816	0.0197	0.0527	-0.0170	0.1051	0.1054

Exact finite-sample bias, variance, and MSE of $\hat{\theta}$. 5000 experiments. Local linear estimator with kernel='triangular' and bwselect='mserd'.

Table A.2: MC experiments: heteroskedasticity

	n	IRDD Bias	IRDD Var	IRDD MSE	LP Bias	LP Var	LP MSE
DGP 5							
	200	-0.1300	0.0450	0.0620	0.0202	0.3369	0.3373
	500	-0.1170	0.0200	0.0330	0.0431	0.1104	0.1123
	1000	-0.1000	0.0110	0.0210	0.0422	0.0567	0.0585
DGP 6							
	200	-0.3020	0.0800	0.1710	0.0273	0.8629	0.8635
	500	-0.2340	0.0390	0.0930	0.0624	0.2149	0.2187
	1000	-0.1890	0.0220	0.0580	0.0682	0.1025	0.1072
DGP 7							
	200	-0.1280	0.0410	0.0580	-0.0176	0.3194	0.3196
	500	-0.1110	0.0190	0.0310	-0.0111	0.1092	0.1093
	1000	-0.0970	0.0110	0.0200	-0.0021	0.0534	0.0534
DGP 8							
	200	-0.2810	0.0720	0.1510	-0.0038	0.7862	0.7861
	500	-0.2190	0.0350	0.0830	-0.0233	0.2269	0.2274
	1000	-0.1790	0.0200	0.0520	-0.0168	0.1065	0.1068

Exact finite-sample bias, variance, and MSE of $\hat{\theta}$. 5000 experiments. Local linear estimator with kernel='triangular' and bwselect='mserd'.

A.4 Identification

In this section, we revise the known identification results for regression discontinuity designs and show that in sharp designs, the identification can be achieved under weaker one-sided continuity conditions. We also show that under the additional assumption our identifying conditions are both necessary and sufficient.

The regression discontinuity design postulates that the probability of receiving the treatment changes discontinuously at the cutoff. In the iRDD, we also assume that the expected outcome and the probability of receiving the treatment are both monotone in the running variable. We introduce several assumptions below.

Assumption (M1). *The following functions are monotone in some neighborhood of* x_0 *(i)* $x \mapsto \mathbb{E}[Y_1|X=x]$ *and* $x \mapsto \mathbb{E}[Y_0|X=x]$*; (ii)* $x \mapsto \Pr(D=1|X=x)$*.*

Assumption (M2). $\mathbb{E}[Y_1|X=x_0] \geq \mathbb{E}[Y_0|X=x_0]$ *in the non-decreasing case or* $\mathbb{E}[Y_1|X=x_0] \leq \mathbb{E}[Y_0|X=x_0]$ *in the non-increasing case.*

Assumption (RD). *Suppose that*

$$\lim_{x\downarrow x_0} \Pr(D=1|X=x) \neq \lim_{x\uparrow x_0} \Pr(D=1|X=x).$$

Assumption (OC). *Under Assumption (M1), suppose that* $x \mapsto \mathbb{E}[Y_1|X=x]$ *is right-continuous and* $x \mapsto \mathbb{E}[Y_0|X=x]$ *is left-continuous at* x_0*.*

Assumption (M1) ensures that all limits exist at the discontinuity point. For sharp discontinuity design, all individuals with values of the running variable exceeding the cutoff x_0 receive the treatment, while all individuals below the cutoff do not. In other words, $D = \mathbb{1}\{X \geq x_0\}$, and, whence $x \mapsto \Pr(D=1|X=x)$ trivially satisfies (M1), (ii). Assumption (RD) is also trivially satisfied for the sharp design.

(M2) requires the local responsiveness to the treatment at the cutoff. It is not necessary for the identification, but as we shall see below, it allows us to characterize in some sense both necessary and sufficient conditions. (OC) is weaker than the full continuity at the cutoff. For more general fuzzy designs, we need additionally the conditional independence assumption.

Assumption (CI). *Suppose that* $D \perp\!\!\!\perp (Y_1, Y_0)|X=x$ *for all* x *in some neighborhood of* x_0*.*

Proposition A.4.1. *Under Assumptions (M1) and (OC), in the sharp design*

$$\lim_{x\downarrow x_0} \mathbb{E}[Y|X=x] - \lim_{x\uparrow x_0} \mathbb{E}[Y|X=x]$$

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

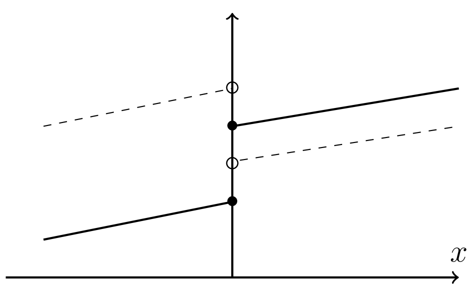
If, additionally, Assumptions (RD) and (CI) are satisfied, and instead of (OC), we assume that $x \mapsto \mathbb{E}[Y_0|X = x]$ and $x \mapsto \mathbb{E}[Y_1 - Y_0|X = x]$ are continuous at x_0, then

$$\frac{\lim_{x\downarrow x_0} \mathbb{E}[Y|X = x] - \lim_{x\uparrow x_0} \mathbb{E}[Y|X = x]}{\lim_{x\downarrow x_0} \Pr(D = 1|X = x) - \lim_{x\uparrow x_0} \Pr(D = 1|X = x)}$$

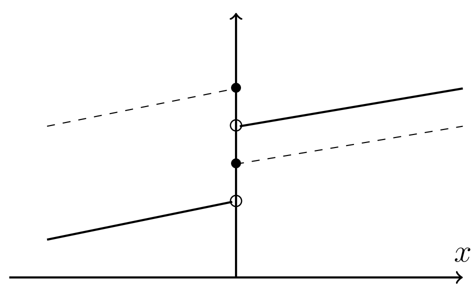
exists and equals to θ.

Proposition A.4.1 shows that for sharp designs, θ is identified for a slightly larger class of distributions than are typically discussed in the literature. It shows that the continuity at the cutoff of both conditional mean functions is not needed and that under monotonicity conditions (M1) and (M2), the one-sided continuity turns out to be both necessary and sufficient for the identification. We illustrate this point in Figure A.2. Panel (a) shows that the causal effect θ can be identified without full continuity. Panel (b) shows that it may happen that the expression in Eq. (1) coincides with θ, yet the two conditional mean functions do not satisfy (OC). Such counterexamples are ruled out by (M2). Inspection of the proof of the Proposition A.4.1 reveals that monotonicity can be easily relaxed if we assume instead that all limits in Eq. (1) and Eq. (2) exist, in which case we recover the result of Hahn, Todd, and Van der Klaauw (2001) under weaker (OC) condition for the sharp design.

It is also worth mentioning that for the fuzzy design, the local monotonicity of the treatment in the running variable allows to identify the causal effect for local compliers; see Hahn, Todd, and Van der Klaauw (2001), Theorem 3.



(a) Identification under (OC)



(b) Counterexample when (OC) fails

Figure A.2: 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 $x \mapsto \mathbb{E}[Y|X = x]$.

A.5 Examples of monotone discontinuity designs

In Table A.3, we collect a list of empirical papers with monotone regression discontinuity designs. We focus only on papers where the global monotonicity is economically plausible and is empirically supported. It is worth stressing that monotonicity restricts only how the average outcome changes with the running variable and that in some references monotonicity appears due to the restricted set of values of the running variable, e.g., elderly people. However, we do not include papers where we might have global piecewise monotonicity with known change points, so the scope of the empirical applicability is probably larger.

References

Abdulkadiroğlu, Atila, Joshua Angrist, and Parag Pathak. 2014. "The elite illusion: Achievement effects at Boston and New York exam schools." *Econometrica*, 82(1): 137–196.

Baum-Snow, Nathaniel, and Justin Marion. 2009. "The effects of low income housing tax credit developments on neighborhoods." *Journal of Public Economics* 93(5-6): 654–666.

Buettner, Thiess. 2006. "The incentive effect of fiscal equalization transfers on tax policy." *Journal of Public Economics,* 90(3): 477–497.

Card, David, Carlos Dobkin, and Nicole Maestas. 2008. "The impact of nearly universal insurance coverage on health care utilization: evidence from Medicare." *American Economic Review,* 98(5): 2242–2258.

Card, David, Raj Chetty, and Andrea Weber. 2007. "Cash-on-hand and competing models of intertemporal behavior: New evidence from the labor market." *The Quarterly Journal of Economics*, 122(4): 1511-1560.

Carpenter, Christopher, and Carlos Dobkin. 2009. "The effect of alcohol consumption on mortality: Regression discontinuity evidence from the minimum drinking age." *American Economic Journal: Applied Economics,* 1(1): 164–182.

Chay, Kenneth Y., and Michael Greenstone. 2005. "Does air quality matter? Evidence from the housing market." *Journal of Political Economy,* 113(2): 376–424.

Table A.3: Monotone regression discontinuity designs

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

Chiang, Hanley. 2009. "How accountability pressure on schools affects student achievement. The threat of sanction on low performing schools can raise test scores in the short run." *Journal of Public Economics,* 93(9-10): 1045–1057.

Clark, Damon. 2010. "Selective schools and academic achievement." *The BE Journal of Economic Analysis & Policy,* 10(1): 1–40.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2011. "Peer effects, teacher incentives, and the impact of tracking: evidence from a randomized evaluation in Kenya." *American Economic Review,* 101(5): 1739–1774.

Ferreira, Fernando. 2010. "You can take it with you: Proposition 13 tax benefits, residential mobility, and willingness to pay for housing amenities." *Journal of Public Economics,* 94(9-10): 661–673.

Hoekstra, Mark. 2009. "The effect of attending the flagship state university on earnings: A discontinuity-based approach." *Review of Economics and Statistics,* 91(4): 717–724.

Greenstone, Michael, and Justin Gallagher. 2012. "Does hazardous waste matter? Evidence from the housing market and the Superfund Program." *The Quarterly Journal of Economics,* 123(3): 951–1003.

Jacob, Brian A., and Lars Lefgren. 2004. "Remedial education and student achievement: A regression discontinuity analysis." *Review of Economics and Statistics,* 86(1): 226–244.

Kaniel, Ron, and Robert Parham. 2009. "WSJ Category Kings – the impact of media attention on consumer and mutual fund investment decisions." *Journal of Financial Economics,* 123(2): 337–356.

Lalive, Rafael. 2007. "Unemployment benefits, unemployment duration and post-unemployment jobs: A regression discontinuity approach." *The American Economic Review*, 97(2): 108–112.

Lee, David S. 2008. "Randomized experiments from non-random selection in U.S. House election." *Journal of Econometrics,* 142(2): 675–697.

Litschig, Stephan, and Kevin M. Morrison. 2013. "The impact of intergovernmental transfers on education outcomes and poverty reduction." *American Economic Journal: Applied Economics,* 5(4): 206–240.

Lucas, Adrienne M., and Isaac M. Mbiti. 2014. "Effects of school quality on student achievement: Discontinuity evidence from Kenya." *American Economic Journal: Applied Economics,* 6(3): 234–263.

Ludwig, Jens, and Douglas L. Miller. 2007. "Does Head Start improve children's life chances? Evidence from regression discontinuity design." *The Quarterly Journal of Economics,* 122(1): 159–208.

Matsudaira, Jordan D. 2008. "Mandatory summer school and student achievement." *Journal of Econometrics,* 142(2): 829–850.

Schmieder, Johannes F., Till von Wachter, and Stefan Bender. 2012. "The effects of extended unemployment insurance over the business cycle: Evidence from regression discontinuity estimates over 20 years." *The Quarterly Journal of Economics,* 127(2): 701–752.

Shigeoka, Hitoshi. 2014. "The effect of patient cost sharing on utilization, health, and risk protection." *American Economic Review,* 104(7): 2152–2184.