Estimation and Inference in Boundary Discontinuity Designs: Distance-Based Methods*

Matias D. Cattaneo† Rocio Titiunik‡ Ruiqi (Rae) Yu§

October 30, 2025

Abstract

We study the statistical properties of nonparametric distance-based (isotropic) local polynomial regression estimators of the boundary average treatment effect curve, a key causal functional parameter capturing heterogeneous treatment effects in boundary discontinuity designs. We present necessary and/or sufficient conditions for identification, estimation, and inference in large samples, both pointwise and uniformly along the boundary. Our theoretical results highlight the crucial role played by the "regularity" of the boundary (a one-dimensional manifold) over which identification, estimation, and inference are conducted. Our methods are illustrated with simulated data. Companion general-purpose software is provided.

Keywords: regression discontinuity, treatment effects estimation, causal inference, isotropic nonparametric regression.

*We thank Alberto Abadie, Boris Hanin, Kosuke Imai, Xinwei Ma, Victor Panaretos, Jörg Stoye, and Jeff Wooldridge for comments and discussions. Cattaneo and Titiunik gratefully acknowledge financial support from the National Science Foundation (SES-2019432, DMS-2210561, and SES-2241575). Cattaneo gratefully acknowledge financial support from the Data-Driven Social Science initiative at Princeton University.

†Department of Operations Research and Financial Engineering, Princeton University.

‡Department of Politics, Princeton University.

§Department of Operations Research and Financial Engineering, Princeton University.

Contents

1 Introduction

Discontinuities in the assignment of a treatment are often used in observational studies to study causal effects. In the standard regression discontinuity (RD) design, each unit has a univariate score and the treatment is assigned according to a threshold rule: units with score equal to or above a known cutoff are assigned to the treatment condition, while units with score below the cutoff are assigned to the control condition. Boundary Discontinuity (BD) designs generalize RD designs to settings where units have a bivariate score, and the treatment is assigned according to a threshold rule based on a one-dimensional boundary curve that partitions the support of the score. A prototypical example is the Geographic RD design and variations thereof [Keele and Titiunik, 2015, Keele et al., 2015, Keele and Titiunik, 2016, Keele et al., 2017, Galiani et al., 2017, Rischard et al., 2021, Diaz and Zubizarreta, 2023], where a geographic boundary splits units into control and treatment areas according to their geolocation. This design is also known as a Multi-Score RD design [Papay et al., 2011, Reardon and Robinson, 2012, Wong et al., 2013]. See Cattaneo and Titiunik [2022] for an overview of the RD design literature, Cattaneo et al. [2024b, Section 5] for a practical introduction to BD designs, and Cattaneo et al. [2026] for a review of the literature on BD designs.

Following a standard causal inference setup [see, e.g., Hernán and Robins, 2020], we let $Y_i(0)$ and $Y_i(1)$ denote the potential outcomes for unit $i = 1, 2, \ldots, n$ under control and treatment assignments, respectively. In BD designs, units have a continuous bivariate score vector $\mathbf{X}_i = (X_{1i}, X_{2i})^\top$ with support $\mathscr{X} \subseteq \mathbb{R}^2$, and they are assigned to either the control group or the treatment group according to their location $\mathbf{X}_i$ relative to a known one-dimensional boundary curve $\mathscr{B}$ that splits the support $\mathscr{X}$ into two disjoint regions: $\mathscr{X} = \mathscr{A}_0 \cup \mathscr{A}_1$, with $\mathscr{A}_0$ and $\mathscr{A}_1$ the control and treatment disjoint (connected) regions, respectively, and $\mathscr{B} = \mathtt{bd}(\mathscr{A}_0) \cap \mathtt{bd}(\mathscr{A}_1)$, where $\mathtt{bd}(\mathscr{A}_t)$ denotes the boundary of the set $\mathscr{A}_t$. Thus, the observed response variable is $Y_i = \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_0) \cdot Y_i(0) + \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_1) \cdot Y_i(1)$. Without loss of generality, we assume that the boundary belongs to the treatment group, that is, $\mathtt{bd}(\mathscr{A}_1) \subset \mathscr{A}_1$ and $\mathscr{B} \cap \mathscr{A}_0 = \emptyset$.

The key building block in BD designs is the Boundary Average Treatment Effect Curve (BATEC), $\tau(\mathbf{x}) = \mathbb{E}[Y_i(1) - Y_i(0) | \mathbf{X}_i = \mathbf{x}]$ for $\mathbf{x} \in \mathscr{B}$, a functional causal parameter that captures heterogeneous treatment effects along the assignment boundary. Following the seminal work of Hahn et al. [2001]

for RD designs with univariate score, Papay et al. [2011], Reardon and Robinson [2012], Wong et al. [2013], Keele and Titiunik [2015], among others, discussed nonparametric identification of the BATEC:

$$\tau(\mathbf{x}) = \mathbb{E}[Y_i(1) - Y_i(0)|\mathbf{X}_i = \mathbf{x}] = \lim_{\mathbf{u}\to\mathbf{x},\mathbf{u}\in\mathscr{A}_1} \mathbb{E}[Y_i|\mathbf{X}_i = \mathbf{u}] - \lim_{\mathbf{u}\to\mathbf{x},\mathbf{u}\in\mathscr{A}_0} \mathbb{E}[Y_i|\mathbf{X}_i = \mathbf{u}],$$

for all $\mathbf{x} \in \mathscr{B}$. Building on this identification result, Cattaneo et al. [2025b] recently developed pointwise (for each $\mathbf{x} \in \mathscr{B}$) and uniform (over $\mathscr{B}$) estimation and inference methods for $\tau(\mathbf{x})$ employing local polynomial regression estimators of $\mathbb{E}[Y_i|\mathbf{X}_i = \mathbf{x}]$ where the two dimensions of the bivariate score $\mathbf{X}_i$ are explicitly incorporated as polynomial bases. These methods are referred to as *location-based* local polynomial regression methods because they directly employ the two dimensions of the bivariate score for estimation. Under assumptions on the geometry of the assignment boundary $\mathscr{B}$, Cattaneo et al. [2025b] show that *location-based* methods for estimation and inference enjoy good statistical properties and are broadly applicable.

In practice, however, scholars often transform the bivariate score according to some univariate distance function, and use this distance measure directly for estimation and inference. For heterogeneity analysis along the boundary, the distance-based polynomial regression thus employ a univariate distance variable to each point on the boundary $\mathbf{x} \in \mathscr{B}$ instead of the bivariate location information encoded in $\mathbf{X}_i$. Reardon and Robinson [2012, Section 4] called this approach distance-based RD, and Keele and Titiunik [2015] discussed this approach in the context of Geographic RD applications. For example, this approach is used because in some practical settings location information may not be available due to confidentiality or other data restrictions, while distance-based information is available and hence it may be the only way to estimate treatment effects. See Cattaneo et al. [2026] for an overview of the empirical literature employing distance in BD designs.

From a nonparametric smoothing perspective, the distance-based approach is connected to isotropic nonparametric regression in statistics and machine learning. We investigate the formal properties of this method, which we call *distance-based* local polynomial regression methods to distinguish them from the location-based local polynomial methods that directly employ the two dimensions of the bivariate score in the polynomial fit. To formalize the approach in the context

of BD designs, we introduce the scalar signed distance-based score for each unit $i = 1, \ldots, n$,

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

where $\mathscr{d}(\cdot,\cdot)$ denotes a distance function. A typical example is the Euclidean distance $\mathscr{d}(\mathbf{X}_i, \mathbf{x}) = \|\mathbf{X}_i - \mathbf{x}\| = \sqrt{(X_{1i} - x_1)^2 + (X_{2i} - x_2)^2}$ for $\mathbf{x} = (x_1, x_2)^\top \in \mathscr{B}$, but other distance measures are possible, particularly in spatial settings [Banerjee, 2005]. For each $\mathbf{x} \in \mathscr{B}$, the distance-based setup is like a standard univariate RD design where $D_i(\mathbf{x}) \in \mathbb{R}$ is the score variable, $c = 0$ is the cutoff, and $D_i(\mathbf{x}) \geq 0$ if unit i is assigned to treatment and $D_i(\mathbf{x}) < 0$ if unit i is assigned to control. Following standard practice in the RD literature [Cattaneo and Titiunik, 2022, Cattaneo et al., 2020, 2024b], the distance-based (univariate) local polynomial treatment effect estimator for each $\mathbf{x} \in \mathscr{B}$ is

$$\widehat{\vartheta}(\mathbf{x}) = \mathbf{e}_1^\top \widehat{\boldsymbol{\gamma}}_1(\mathbf{x}) - \mathbf{e}_1^\top \widehat{\boldsymbol{\gamma}}_0(\mathbf{x}), \qquad \mathbf{x} \in \mathscr{B},$$

where, for $t \in \{0, 1\}$,

$$\widehat{\boldsymbol{\gamma}}_t(\mathbf{x}) = \underset{\boldsymbol{\gamma} \in \mathbb{R}^{p+1}}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} \big(Y_i - \mathbf{r}_p(D_i(\mathbf{x}))^\top \boldsymbol{\gamma}\big)^2 K_h(D_i(\mathbf{x})) \mathbb{1}(D_i(\mathbf{x}) \in \mathscr{I}_t),$$

with $\mathbf{e}_1$ the conformable first unit vector, $\mathbf{r}_p(u) = (1, u, u^2, \cdots, u^p)^\top$ the usual univariate polynomial basis, $K_h(u) = K(u/h)/h^2$ a univariate kernel function $K(\cdot)$ with bandwidth parameter h, and $\mathscr{I}_0 = (-\infty, 0)$ and $\mathscr{I}_1 = [0, \infty)$. The kernel function typically down-weights observations as the distance to $\mathbf{x} \in \mathscr{B}$ increases, while the bandwidth determines the level of localization to each point on the boundary $\mathscr{B}$. In this setup, the observations contribute isotropically in terms of their univariate distance $\mathscr{d}(\mathbf{X}_i, \mathbf{x})$ to the point $\mathbf{x} \in \mathscr{B}$ where estimation occurs.

The distance-based estimator $\widehat{\vartheta}(\mathbf{x})$ is the difference of two distance-based (univariate, isotropic) local polynomial regression estimators along the one-dimensional manifold $\mathscr{B}$. This estimator does not directly target $\tau(\mathbf{x})$; instead, it targets the difference of the expectations $\theta_{1,\mathbf{x}}(r) - \theta_{0,\mathbf{x}}(r)$ at $r = 0$ with $\theta_{t,\mathbf{x}}(r) = \mathbb{E}[Y_i(t) | \mathscr{d}(\mathbf{X}_i, \mathbf{x}) = r]$ for $r \geq 0$ and $t \in \{0, 1\}$, which are univariate conditional expectations induced by transforming the bivariate score using the distance function. We study the statistical properties of using the distance-based local polynomial estimator $\widehat{\vartheta}(\mathbf{x})$

to perform estimation and inference for the BATEC, $\tau(\mathbf{x})$ with $\mathbf{x} \in \mathscr{B}$, presenting necessary and sufficient conditions for identification, estimation and inference in large samples, both pointwise and uniformly along the assignment boundary.

We start by studying identification of $\tau(\mathbf{x})$ via a distance-based approximation. The induced univariate conditional expectations $\theta_{t,\mathbf{x}}(r)$ are of course different from the bivariate conditional expectations $\mu_t(\mathbf{x}) = \mathbb{E}[Y_i(t)|\mathbf{X}_i = \mathbf{x}]$, for $t \in \{0,1\}$. It follows that without restrictions on the data generating process, the assignment boundary, and the distance function, $\lim_{r\to 0} \theta_{t,\mathbf{x}}(r)$ and $\mu_t(\mathbf{x})$ are also different in general. In Theorem 1, we provide sufficient conditions under which these parameters agree, which include restricting the geometry of $\mathscr{B}$ so that it is a rectifiable curve. Theorem 1 contributes to the multi-dimensional RD design literature by establishing a new identification result for distance-based methods [for related results, see Hahn et al., 2001, Papay et al., 2011, Reardon and Robinson, 2012, Wong et al., 2013, Cattaneo et al., 2016, and references therein].

The geometry of the boundary affects not only identification, but also bias approximation. We show two results. First, Theorem 2 shows that "near" irregular points of the assignment boundary that can occur even for a "nice" one-dimensional manifold $\mathscr{B}$ (e.g., a piecewise linear curve with a "kink"), the pth order distance-based estimator $\widehat{\vartheta}(\mathbf{x})$ exhibits an irreducible bias of order h, no matter the polynomial order p used or the amount of smoothness assumed on the location-based underlying conditional expectations $\mu_0(\mathbf{x})$ and $\mu_1(\mathbf{x})$. This drawback occurs because the induced distance-based population regression function $\theta_t(\mathbf{x}) = \mathbb{E}[Y_i(t)|\mathscr{d}(\mathbf{X}_i,\mathbf{x}) = r]$ is at most Lipschitz continuous uniformly near kinks of the boundary $\mathscr{B}$. Second, Theorem 3 shows that when the boundary $\mathscr{B}$ is "nice" and "smooth", the pth order distance-based estimator $\widehat{\vartheta}(\mathbf{x})$ exhibits the usual approximation bias of order h^{p+1} from the nonparametric smoothing literature [see, e.g. Härdle et al., 2004, and references therein]. Our results demonstrate that the distance-based treatment effect estimator $\widehat{\vartheta}(\mathbf{x})$ can exhibit different (potentially large) biases along the boundary $\mathscr{B}$, depending on its local geometry (lack of "smoothness").

These bias approximation results, which appear to be new to the literature, show that standard univariate RD methods for estimation and inference may enjoy different, unexpected properties when deployed directly to a BD design using the univariate distance $D_i(\mathbf{x})$, because the features of the assignment boundary can lead to a large misspecification bias near kinks or other irregularities

of $\mathscr{B}$. The large misspecification bias we highlight when the assignment boundary is non-smooth is particularly worrisome in Geographic RD designs, where the assignment boundary can naturally exhibit many kinks or other irregularities. Our theoretical results thus have concrete implications for bandwidth selection, point estimation, and statistical inference in real-world applications.

We also study estimation and inference for $\tau(\mathbf{x})$ using the distance-based estimator $\widehat{\vartheta}(\mathbf{x})$. Theorem 4 establishes convergence rates and Theorem 5 establishes validity of uncertainty quantification, both pointwise and uniformly over $\mathscr{B}$. These results are presented in terms of a generic misspecification bias quantity in order to accommodate the different cases emerging from the interaction between the distance function $\mathscr{d}(\cdot)$ and geometry of the boundary $\mathscr{B}$. For point estimation, and under appropriate regularity conditions, we show that the estimator $\widehat{\vartheta}(\mathbf{x})$ can achieve the pointwise and uniform (over $\mathscr{B}$) optimal nonparametric convergence rate [see, e.g., Tsybakov, 2008, and references therein], when $\mathscr{B}$ is "nice" and "smooth". For uncertainty quantification, we develop confidence intervals for $\tau(\mathbf{x})$ and confidence bands for the entire treatment effect curve $(\tau(\mathbf{x}) : \mathbf{x} \in \mathscr{B})$. We establish the uniform inference results leveraging a new strong approximation theorem for empirical processes given in the supplemental appendix (Section SA-5). These results appear to be new to the literature, as they pertain pointwise and uniform properties of distance-based (isotropic) local polynomial regression estimators over a one-dimensional manifold.

We also study a more fundamental theoretical question underpinning Theorem 2, which establishes an unimprovable "large" uniform approximation bias for the distance-based estimator $\widehat{\vartheta}(\mathbf{x})$ when $\mathscr{B}$ is not "nice" and "smooth". Theorem 6 establishes the minimax uniform convergence rate for the class of isotropic nonparametric estimators of a smooth, compact supported, bivariate nonparametric conditional expectation over a one-dimensional rectifiable manifold. Our result shows that, regardless of the level of smoothness assumed on the underlying nonparametric conditional expectation, the best univariate isotropic nonparametric estimator in the class can at most achieve the uniform convergence rate $n^{-1/4}$. This convergence rate corresponds to the optimal minimax uniform convergence rate for the class of Lipschitz continuous bivariate conditional expectations when considering all possible nonparametric estimators, up to polylog(n) terms. Therefore, Theorem 6 implies that the convergence rate of the distance-based estimator $\widehat{\vartheta}(\mathbf{x})$ established by combining Theorems 2 and 4 is unimprovable, up to polylog(n) terms, without further restricting the geometry of the assignment boundary $\mathscr{B}$ (the one-dimensional manifold over which uniformity is established),

or enlarging the class of distance-based estimators to explicitly exploit additional information about the data generating process. Theorem 6 appears to be new to the minimax estimation literature [see, e.g., Tsybakov, 2008, and references therein].

Taken together, ours theoretical results characterize identification, estimation, and inference for the BATEC when this causal parameter is analyzed using univariate distance-based local polynomial regression methods in BD designs. As discussed in Section 5, our findings also inform practice, and suggest new "regularized" estimation and inference distance-based methods when specific information about the geometry of the assignment boundary $\mathscr{B}$ is readily available. In addition, Section 7 presents a small simulation study demonstrating the finite-sample properties of the distance-based methods, employing data generating process calibrated using the real-world dataset from Londoño-Vélez et al. [2020], who analyzed the Colombian governmental post-secondary education subsidy, *Ser Pilo Paga* (SPP). These numerical results are obtained using our companion software `R` package `rd2d` (`https://rdpackages.github.io/rd2d`); see Cattaneo et al. [2025d] for more discussion.

Our paper establishes a new bridge between the methodology for boundary discontinuity (BD) designs and the theoretical study of nonparametric smoothing over submanifolds, thereby contributing to two distinct, yet interconnected literatures in statistics, econometrics, machine learning, and data science. From a methodological perspective, we offer the first foundational results for the analysis and interpretation of BD designs when employing distance-based methods, an approach used for treatment effect estimation and causal inference across numerous quantitative disciplines, but currently lacking a foundational understanding. Theoretically, we provide new estimation and inference results for isotropic local polynomial regression over a submanifold, both pointwise and uniformly, thereby advancing the literature on nonparametric smoothing. In particular, concurrent work by Chen and Gao [2025] develop estimation and inference methods for integral functionals on submanifolds using series/sieve estimation. Our theoretical results substantially differ from theirs in several respects. First, we consider one-dimensional distance to each point on the submanifold over which estimation and inference is conducted. Second, we study (isotropic) local polynomial regression smoothing methods, whose distinct structure requires a different analysis [see, e.g., Cattaneo et al., 2025a, for more discussion]. Third, we study pointwise and uniform over the submanifold estimation and inference methods, and thus uncover new issues related to the (uniform) misspecification bias. Finally, we establish a new minimax estimation rate for isotropic nonparametric

regression methods [e.g., Tsybakov, 2008, and references therein].

The rest of the manuscript is organized as follows. Section 2 introduces the setup and assumptions used throughout the paper. Section 3 investigates identification of $\tau(\mathbf{x})$ via distance-based approximation, while Section 4 investigates the bias approximation properties of the distance-based estimator $\widehat{\vartheta}(\mathbf{x})$. Section 5 discusses estimation and inference for $\tau(\mathbf{x})$ using the distance-based local polynomial regression estimator, while Section 6 tackles the fundamental theoretical minimax estimation question underpinning Theorem 2. Section 7 illustrates the numerical performance of the distance-based methods with simulated data, while Section 8 concludes. The supplemental appendix presents generalizations of our theoretical results, reports their proofs, and gives other theoretical results that may be of independent interest. In particular, it presents a new strong approximation result for empirical process with multiplicative-separable structure and bounded polynomial moments, extending recent work by Cattaneo and Yu [2025].

1.1 Notation

We employ standard concepts and notation from empirical process theory [van der Vaart and Wellner, 1996, Giné and Nickl, 2016] and geometric measure theory [Simon et al., 1984, Federer, 2014]. For a random variable $\mathbf{V}_i$, we write $\mathbb{E}_n[g(\mathbf{V}_i)] = n^{-1}\sum_{i=1}^n g(\mathbf{V}_i)$. For a vector $\mathbf{v} \in \mathbb{R}^k$, the Euclidean norm is $\|\mathbf{v}\| = (\sum_{i=1}^k \mathbf{v}_i^2)^{1/2}$. For a matrix $\mathbf{A}$, $\lambda_{\min}(\mathbf{A})$ denotes the minimum eigenvalue. $C^k(\mathcal{X}, \mathcal{Y})$ denotes the class of k-times continuously differentiable functions from $\mathcal{X}$ to $\mathcal{Y}$, and $C^k(\mathcal{X})$ is a shorthand for $C^k(\mathcal{X}, \mathbb{R})$. For a Borel set $\mathcal{S} \subseteq \mathcal{X}$, the De Giorgi perimeter of $\mathcal{S}$ is $\operatorname{perim}(\mathcal{S}) = \sup_{g \in \mathscr{D}_2(\mathcal{X})} \int_{\mathbb{R}^2} \mathbb{1}(\mathbf{x} \in \mathcal{S}) \operatorname{div} g(\mathbf{x}) d\mathbf{x} / \|g\|_\infty$, where div is the divergence operator, and $\mathscr{D}_2(\mathcal{X})$ denotes the space of C^∞ functions with values in $\mathbb{R}^2$ and with compact support included in $\mathcal{X}$. When $\mathcal{S}$ is connected, and the boundary $\mathtt{bd}(\mathcal{S})$ is a smooth simple closed curve, $\operatorname{perim}(\mathcal{S})$ simplifies to the curve length of $\mathtt{bd}(\mathcal{S})$. A curve $\mathscr{B} \subseteq \mathbb{R}^2$ is a *rectifiable curve* if there exists a Lipschitz continuous function $\gamma : [0,1] \mapsto \mathbb{R}^2$ such that $\mathscr{B} = \gamma([0,1])$. For a function $f : \mathbb{R}^2 \mapsto \mathbb{R}$, $\operatorname{Supp}(f)$ denotes closure of the set $\{\mathbf{x} \in \mathbb{R}^2 : f(\mathbf{x}) \neq 0\}$. For real sequences $a_n = o(b_n)$ if $\limsup_{n\to\infty} \frac{|a_n|}{|b_n|} = 0$, $a_n \lesssim b_n$ if there exists some constant C and $N > 0$ such that $n > N$ implies $|a_n| \leq C|b_n|$. For sequences of random variables $a_n = o_{\mathbb{P}}(b_n)$ if $\operatorname{plim}_{n\to\infty} \frac{|a_n|}{|b_n|} = 0, |a_n| \lesssim_{\mathbb{P}} |b_n|$ if $\limsup_{M\to\infty} \limsup_{n\to\infty} \mathbb{P}[|\frac{a_n}{b_n}| \geq M] = 0$. Finally, $\Phi(x)$ denotes the standard Gaussian cumulative distribution function.

2 Setup and Assumptions

The following assumption collects the core conditions imposed on the underlying data generating process.

Assumption 1 (Data Generating Process)**.** *Let $t \in \{0,1\}$, $p \geq 1$, and $v \geq 2$.*

(i) $(Y_1(t), \mathbf{X}_1^\top)^\top, \ldots, (Y_n(t), \mathbf{X}_n^\top)^\top$ *are independent and identically distributed random vectors.*

(ii) *The distribution of $\mathbf{X}_i$ has a Lebesgue density $f_X(\mathbf{x})$ that is continuous and bounded away from zero on its support $\mathcal{X} = [\mathtt{L}, \mathtt{U}]^2$, for $-\infty < \mathtt{L} < \mathtt{U} < \infty$.*

(iii) $\mu_t(\mathbf{x}) = \mathbb{E}[Y_i(t)|\mathbf{X}_i = \mathbf{x}]$ *is $(p+1)$-times continuously differentiable on $\mathcal{X}$.*

(iv) $\sigma_t^2(\mathbf{x}) = \mathbb{V}[Y_i(t)|\mathbf{X}_i = \mathbf{x}]$ *is bounded away from zero and continuous on $\mathcal{X}$.*

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

Assumption 1 is the natural generalization of standard conditions in the univariate RD design [see, e.g., Cattaneo and Titiunik, 2022, and references therein]. We also need to impose additional restrictions specific to distance-based methods in BD designs. Let

$$\mathbf{\Psi}_{t,\mathbf{x}} = \mathbb{E}\Big[\mathbf{r}_p\Big(\frac{D_i(\mathbf{x})}{h}\Big)\mathbf{r}_p\Big(\frac{D_i(\mathbf{x})}{h}\Big)^\top K_h(D_i(\mathbf{x}))\mathbb{1}(D_i(\mathbf{x}) \in \mathcal{I}_t)\Big]$$

be the fixed-h population Gram matrix associated with the distance-based estimator.

Assumption 2 (Kernel, Distance, and Boundary)**.** *Let $t \in \{0,1\}$.*

(i) $\mathscr{B}$ *is a rectifiable curve.*

(ii) $\mathscr{d} : \mathbb{R}^2 \mapsto [0,\infty)$ *satisfies $\|\mathbf{x}_1 - \mathbf{x}_2\| \lesssim \mathscr{d}(\mathbf{x}_1, \mathbf{x}_2) \lesssim \|\mathbf{x}_1 - \mathbf{x}_2\|$ for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$.*

(iii) $K : \mathbb{R} \to [0,\infty)$ *is compact supported and Lipschitz continuous, or $K(u) = \mathbb{1}(u \in [-1,1])$.*

(iv) $\liminf_{h\downarrow 0} \inf_{\mathbf{x}\in\mathscr{B}} \lambda_{\min}(\mathbf{\Psi}_{t,\mathbf{x}}) \gtrsim 1$.

Assumption 2(i) imposes minimal regularity on the assignment boundary $\mathscr{B}$, which is used to compute integrals as well as to establish uniform results over the univariate submanifold. Assumption 2(ii) requires the distance function be equivalent (up to constants) to the Euclidean distance. Assumption 2(iii) imposes standard conditions on the univariate kernel function. Assumption 2(iv) further restricts (implicitly) the geometry of the boundary $\mathscr{B}$ relative to the kernel and distance

functions, ruling out highly irregular boundary shapes leading to regions with too "few" data points. Lemma SA-1 in the supplemental appendix gives primitive conditions for Assumption 2(iv) when $\mathscr{d}(\cdot)$ is the Euclidean norm. The conditions imposed by Assumption 2 are sufficient for uniform (over $\mathscr{B}$) estimation and inference, but some of them can be weakened for pointwise results—see the supplemental appendix.

3 Identification and Interpretation

For each boundary point $\mathbf{x} \in \mathscr{B}$ and corresponding signed distance score $D_i(\mathbf{x})$, the univariate distance-based local polynomial estimator is $\widehat{\vartheta}(\mathbf{x}) = \widehat{\theta}_{1,\mathbf{x}}(0) - \widehat{\theta}_{0,\mathbf{x}}(0)$, where $\widehat{\theta}_{t,\mathbf{x}}(0) = \mathbf{e}_1^\top \widehat{\boldsymbol{\gamma}}_t(\mathbf{x}) = \mathbf{r}_p(0)^\top \widehat{\boldsymbol{\gamma}}_t(\mathbf{x})$. Conceptually, the estimator $\widehat{\theta}_{t,\mathbf{x}}(0)$ targets the estimand $\theta_{t,\mathbf{x}}(0)$ with

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

which is the univariate conditional expectation induced by the distance transformation applied to the bivariate location variable $\mathbf{X}_i$ for each point $\mathbf{x} \in \mathscr{B}$.

The following theorem establishes identification of the causal functional parameter $\tau(\mathbf{x})$ via the distance-based induced conditional expectations.

Theorem 1 (Identification)**.** *Suppose Assumptions 1(i)–(iii) and 2(i)–(ii) hold. Then,*

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

for all $\mathbf{x} \in \mathscr{B}$*.*

This identification theorem is established by representing $\theta_{t,\mathbf{x}}(r)$ as a sequence of abstract integrals over submanifolds near $\mathscr{B}$, that is, over level sets of a shrinking tubular (generated by $\mathscr{d}(\cdot)$ and) covering the assignment boundary [see, e.g., Federer, 2014, for theoretical background]. Without restricting the data generating process, the assignment boundary, and the distance function, the parameters $\lim_{r \to 0} \theta_{t,\mathbf{x}}(r)$ and $\mu_t(\mathbf{x})$ need not agree. This identification result is new to the literature. For related continuity-based identification results in RD designs see Hahn et al. [2001] in one-dimensional score settings, Papay et al. [2011], Reardon and Robinson [2012], and Keele and

Titiunik [2015] in multi-score settings, and Cattaneo et al. [2016] in multi-cutoff settings.

The isotropic nonparametric methods underlying the distance-based methods in BD designs naturally induce nonparametric misspecification for approximating $\tau(\mathbf{x})$. To circumvent this problem, we view the univariate distance-based local polynomial estimator $\widehat{\boldsymbol{\gamma}}_t(\mathbf{x})$ as the sample analog of the coefficients associated with the best (weighted, local) mean square approximation of the conditional expectation $\mathbb{E}[Y_i(t)|D_i(\mathbf{x})]$ based on $\mathbf{r}_p(D_i(\mathbf{x}))$:

$$\boldsymbol{\gamma}_t^*(\mathbf{x}) = \underset{\boldsymbol{\gamma}\in\mathbb{R}^{p+1}}{\arg\min}\, \mathbb{E}\Big[\big(Y_i - \mathbf{r}_p(D_i(\mathbf{x}))^\top\boldsymbol{\gamma}\big)^2 K_h(D_i(\mathbf{x}))\mathbb{1}(D_i(\mathbf{x})\in\mathscr{I}_t)\Big].$$

Therefore, letting $\theta_{t,\mathbf{x}}^*(0) = \mathbf{e}_1^\top\boldsymbol{\gamma}_t^*(\mathbf{x}) = \mathbf{r}_p(0)^\top\boldsymbol{\gamma}_t^*(\mathbf{x})$, we have the standard least squares (best linear approximation) decomposition

$$\widehat{\theta}_{t,\mathbf{x}}(0) - \theta_{t,\mathbf{x}}(0) = \big[\theta_{t,\mathbf{x}}^*(0) - \theta_{t,\mathbf{x}}(0)\big] + \mathbf{e}_1^\top\boldsymbol{\Psi}_{t,\mathbf{x}}^{-1}\mathbf{O}_{t,\mathbf{x}} + \big[\mathbf{e}_1^\top(\widehat{\boldsymbol{\Psi}}_{t,\mathbf{x}}^{-1} - \boldsymbol{\Psi}_{t,\mathbf{x}}^{-1})\mathbf{O}_{t,\mathbf{x}}\big], \tag{1}$$

where

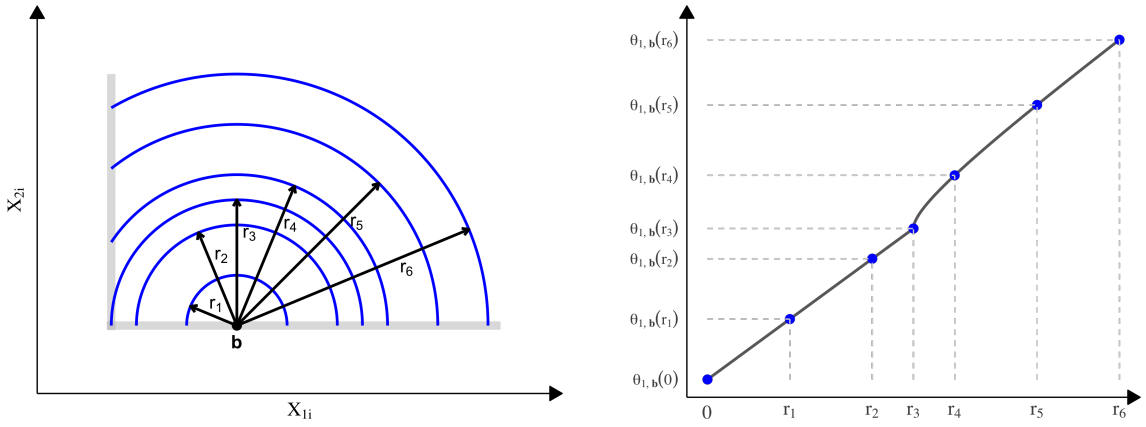
$$\begin{aligned}
\widehat{\boldsymbol{\Psi}}_{t,\mathbf{x}} &= \mathbb{E}_n\Big[\mathbf{r}_p\Big(\frac{D_i(\mathbf{x})}{h}\Big)\mathbf{r}_p\Big(\frac{D_i(\mathbf{x})}{h}\Big)^\top K_h(D_i(\mathbf{x}))\mathbb{1}(D_i(\mathbf{x})\in\mathscr{I}_t)\Big], \qquad \text{and}\\
\mathbf{O}_{t,\mathbf{x}} &= \mathbb{E}_n\Big[\mathbf{r}_p\Big(\frac{D_i(\mathbf{x})}{h}\Big)K_h(D_i(\mathbf{x}))(Y_i - \theta_{t,\mathbf{x}}^*(D_i(\mathbf{x})))\mathbb{1}(D_i(\mathbf{x})\in\mathscr{I}_t)\Big].
\end{aligned}$$

In the decomposition (1), the first term is the non-random mean square approximation bias, the second term is the stochastic linear approximation of the centered estimator (an average of unconditional mean-zero random variables), and the third term is a "linearization" error arising from the convergence of the Gram matrix associated with the (weighted) least squares estimator. Importantly, our results do not assume a conditional mean-zero restriction; this allows us to handle potential misspecification broadly.

Theorem 1 and the decomposition (1) imply that

$$\widehat{\vartheta}(\mathbf{x}) - \tau(\mathbf{x}) = \mathfrak{B}(\mathbf{x}) + \mathfrak{L}(\mathbf{x}) + \mathfrak{Q}(\mathbf{x}), \qquad \mathbf{x}\in\mathscr{B},$$

where $\mathfrak{B}(\mathbf{x}) = \theta_{1,\mathbf{x}}^*(0) - \theta_{0,\mathbf{x}}^*(0) - \tau(\mathbf{x})$ is the nonrandom approximation bias of the estimator, $\mathfrak{L}(\mathbf{x}) = \mathbf{e}_1^\top\boldsymbol{\Psi}_{1,\mathbf{x}}^{-1}\mathbf{O}_{1,\mathbf{x}} - \mathbf{e}_1^\top\boldsymbol{\Psi}_{0,\mathbf{x}}^{-1}\mathbf{O}_{0,\mathbf{x}}$ is a unconditional mean-zero linear (sample average) statistic,



(a) Distance to $\mathbf{b} \in \mathscr{B}$.

(b) Distance-based Conditional Expectation.

Figure 1: Lack of smoothness of distance-based conditional expectation near a kink.
Note: Analytic example of $\theta_{1,\mathbf{b}}(r) = \mathbb{E}[Y(1)|D_i(\mathbf{b}) = r]$, $r \geq 0$, for distance transformation $D_i(\mathbf{b}) = \mathscr{d}(\mathbf{X}_i, \mathbf{b}) = \|\mathbf{X}_i - \mathbf{b}\|$ to point $\mathbf{b} \in \mathscr{B}$ near a kink point on the boundary, based on location $\mathbf{X}_i = (X_{1i}, X_{2i})^\top$. The induced univariate conditional expectation $r \mapsto \theta_{1,\mathbf{b}}(r)$ is continuous but not differentiable at $r = r_3$.

and $\mathfrak{Q}(\mathbf{x}) = \mathbf{e}_1^\top(\widehat{\mathbf{\Psi}}_{1,\mathbf{x}}^{-1} - \mathbf{\Psi}_{1,\mathbf{x}}^{-1})\mathbf{O}_{1,\mathbf{x}} - \mathbf{e}_1^\top(\widehat{\mathbf{\Psi}}_{0,\mathbf{x}}^{-1} - \mathbf{\Psi}_{0,\mathbf{x}}^{-1})\mathbf{O}_{0,\mathbf{x}}$ is the higher-order linearization error.

In standard large-sample local polynomial regression settings, $\mathfrak{B}(\mathbf{x})$ is of order h^{p+1}, $\mathfrak{L}(\mathbf{x})$ is conditionally mean-zero and approximately Gaussian, and $\mathfrak{Q}(\mathbf{x})$ is neglected. However, in our setting these standard results are not valid in general because estimation is conducted along the one-dimensional assignment boundary $\mathscr{B}$ (a particular submanifold) and is based on distance to each point on $\mathscr{B}$ (isotropic nonparametric smoothing). We show that additional conditions may be needed to attain the usual local polynomial approximation results, and that some of these results may be invalid in some settings.

4 Approximation Bias

The smoothness of the induced distance-based conditional expectation function $r \mapsto \theta_{t,\mathbf{x}}(r) = \mathbb{E}[Y_i(t)|D_i(\mathbf{x}) = r]$ depends on the smoothness of the conditional expectation $\mu_t(\mathbf{x}) = \mathbb{E}[Y_i(t)|\mathbf{X}_i = \mathbf{x}]$, the distance function $\mathscr{d}(\cdot)$, and the "regularity" of the boundary $\mathscr{B}$. This means that the bias of the distance-based local polynomial estimator can be affected by the shape of the boundary $\mathscr{B}$, regardless of how large p is (Assumption 1).

Figure 1 demonstrates the problem graphically: for a point $\mathbf{x} \in \mathscr{B}$ that is close enough to a kink point on the boundary, the conditional expectation $\theta_{1,\mathbf{x}}(r)$ is not differentiable for all $r \geq 0$. The

problem arises because, given the distance function $\mathscr{d}(\cdot,\cdot)$, for any point $\mathbf{x} \in \mathscr{B}$ near a kink, a "small" r gives a complete arc $\{\mathbf{x} \in \mathscr{A}_1 : \mathscr{d}(\mathbf{X}_i, \mathbf{x}) = r\}$, while the arc is truncated by the boundary for a "large" r. As a result, for the example in Figure 1, $\theta_{1,\mathbf{x}}(r)$ is smooth for all $r \leq r_3$ and $r > r_3$, but the function is not differentiable at $r = r_3$. Furthermore, at $r = r_3$, the left derivative is constant, but the right derivative is equal to infinity. The supplemental appendix gives details on this analytic example.

Although smoothness of the boundary $\mathscr{B}$ can affect the smoothness of $\theta_{t,\mathbf{x}}(r)$, the fact that the distance-based estimator is "local" means that the *approximation error* will be no greater than that of a local constant estimator, regardless of the polynomial order used and smoothness assumed (controlled by p in Assumption 1). The following theorem formalizes this result, and also shows that this bias order cannot be improved by increasing $p \geq 1$.

Theorem 2 (Approximation Bias: Uniform Guarantee)**.** *For some $L > 0$, let $\mathscr{P}$ be the class of data generating processes satisfying Assumptions 1(i)-(iii) with $\mathscr{X} \subseteq [-L, L]^2$, Assumption 2, and the following three additional conditions:*

(i) $L^{-1} \leq \inf_{\mathbf{x}\in\mathscr{X}} f_X(\mathbf{x}) \leq \sup_{\mathbf{x}\in\mathscr{X}} f_X(\mathbf{x}) \leq L$,

(ii) $\max_{0\leq|\boldsymbol{\nu}|\leq p} \sup_{\mathbf{x}\in\mathscr{X}} |\partial^{\boldsymbol{\nu}}\mu_t(\mathbf{x})| + \max_{0\leq|\boldsymbol{\nu}|\leq p} \sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}} \frac{|\partial^{\boldsymbol{\nu}}\mu_t(\mathbf{x}) - \partial^{\boldsymbol{\nu}}\mu_t(\mathbf{y})|}{\|\mathbf{x}-\mathbf{y}\|} \leq L$, *and*

(iii) $\liminf_{h\downarrow 0} \inf_{\mathbf{x}\in\mathscr{B}} \int_{\mathscr{A}_t} k_h(\mathscr{d}(\mathbf{u}, \mathbf{x}))d\mathbf{u} \geq L^{-1}$.

For any $p \geq 1$, if $nh^2 \to \infty$ and $h \to 0$, then

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

This theorem characterizes precisely the uniform over $\mathscr{B}$ (and data generating processes) bias of the distance-based local polynomial estimator $\widehat{\vartheta}(\mathbf{x})$. The upper bound in Theorem 2 is established uniformly over the class of data generating processes because we can show that $|\theta_t(0) - \theta_t(r)| \lesssim r$ for $t \in \{0, 1\}$ in general. The lower bound is shown using the following example. Suppose $\mathbf{X}_i \sim \mathsf{Uniform}([-2, 2]^2)$, $\mu_0(x_1, x_2) = 0$, $\mu_1(x_1, x_2) = x_2$ for all $(x_1, x_2) \in [-2, 2]^2$, and $Y_i(0)|\mathbf{X}_i \sim \mathsf{Normal}(\mu_0(\mathbf{X}_i), 1)$ and $Y_i(1)|\mathbf{X}_i \sim \mathsf{Normal}(\mu_1(\mathbf{X}_i), 1)$. Let $\mathscr{d}(\cdot,\cdot)$ be the Euclidean distance, and suppose that the control and treatment regions are $\mathscr{A}_1 = \{(x, y) \in \mathbb{R}^2 : x \leq 0, y \geq 0\}$ and $\mathscr{A}_0 = \mathbb{R}^2/\mathscr{A}_1$, respectively, and hence $\mathscr{B} = \{(x, y) \in \mathbb{R} : -2 \leq x \leq 0, y = 0 \text{ or } x = 0, 0 \leq y \leq 2\}$.

This boundary is L-shaped as in the empirical application used in Section 7, having a 90° kink at $\mathbf{x} = (0, 0)$. It follows that the conditions of Theorem 2 hold, and hence this is an allowed data generating process. The supplemental appendix establishes the lower bound by careful analysis of the resulting approximation bias.

As a point of contrast, note that it is intuitive to expect that $\mathfrak{B}(\mathbf{x}) \lesssim h^{p+1}$ pointwise in $\mathbf{x} \in \mathscr{B}$, for small enough h, provided that the kinks on the boundary $\mathscr{B}$ are sufficiently far apart of each other relative to the bandwidth, and other regularity conditions hold. However, Theorem 2 demonstrates that, no matter how large the sample size is (i.e., how small the bandwidth is), there will always be a region near a kink of the boundary $\mathscr{B}$ where the misspecification bias of the distance-based local polynomial estimator $\widehat{\vartheta}(\mathbf{x})$ is at best of order h (i.e., not of order h^{p+1} as expected), regardless of the value of p. The problem arises because the boundary $\mathscr{B}$ changes non-smoothly, leading to a non-differentiable induced conditional expectation $\theta_{t,\mathbf{x}}(r)$, as illustrated in Figure 1b.

On the other hand, if the boundary $\mathscr{B}$ is smooth enough, then a better misspecification (smoothing) bias can be established.

Theorem 3 (Approximation Bias: Smooth Boundary)**.** *Suppose Assumption 1(i)-(iii) and Assumptions 2 hold, with $\mathscr{d}(\cdot,\cdot)$ the Euclidean distance. Let $h \to 0$.*

(i) *For $\mathbf{x} \in \mathscr{B}$, and for some $\delta, \varepsilon > 0$, suppose that $\mathscr{B} \cap \{\mathbf{y} : \|\mathbf{y} - \mathbf{x}\| \leq \varepsilon\} = \gamma([-\delta, \delta])$, where $\gamma : \mathbb{R} \to \mathbb{R}^2$ is a one-to-one function in $C^{\kappa+2}([-\delta, \delta], \mathbb{R}^2)$. Then, $\theta_{0,\mathbf{x}}(\cdot)$ and $\theta_{1,\mathbf{x}}(\cdot)$ are $(\kappa \wedge (p+1))$-times continuously differentiable on $[0, \varepsilon]$. Therefore, there exists a positive constant C such that $|\mathfrak{B}(\mathbf{x})| \leq Ch^{\kappa \wedge (p+1)}$ for all $h \in [0, \varepsilon]$ and $\mathbf{x} \in \mathscr{B}$.*

(ii) *Suppose $\mathscr{B} = \gamma([0, L])$ where γ is a one-to-one function in $C^{\iota+2}([0, L], \mathbb{R}^2)$ for some $L > 0$. Suppose there exists $\delta, \varepsilon > 0$ such that for all $\mathbf{x} \in \mathscr{B}^o = \gamma([\delta, L - \delta])$, $r \in [0, \varepsilon]$, and $t = 0, 1$, and the set $\{\mathbf{u} \in \mathbb{R}^2 : \mathscr{d}(\mathbf{u}, \mathbf{x}) = r\}$ intersects with $\mathtt{bd}(\mathscr{A}_t)$ at only two points in $\mathscr{B}$. Then, for all $\mathbf{x} \in \gamma([\delta, L - \delta])$, $\theta_{0,\mathbf{x}}(\cdot)$ and $\theta_{1,\mathbf{x}}(\cdot)$ are $(\iota \wedge (p+1))$-times continuously differentiable on $[0, \varepsilon]$, and $\lim_{r \downarrow 0} \frac{d^v}{dr^v} \theta_{t,\mathbf{x}}(r)$ exists and is finite for all $0 \leq v \leq p+1$ and $t \in \{0, 1\}$. Therefore, there exists a positive constant C such that $\sup_{\mathbf{x} \in \mathscr{B}^o} |\mathfrak{B}(\mathbf{x})| \leq Ch^{\iota \wedge (p+1)}$ for all $h \in [0, r]$.*

This theorem gives sufficient conditions in terms of smoothness of the boundary $\mathscr{B}$ to achieve a usual nonparametric smoothing misspecification bias for $\widehat{\vartheta}(\mathbf{x})$, which improves the minimal guarantee given by Theorem 2. For a uniform bound on misspecification bias, the sufficient condition

in Theorem 3 requires that the boundary $\mathscr{B}$ be *uniformly smooth* in that it can be parameterized by *one* smooth function. The sufficient condition ensuring a smooth boundary everywhere is crucial because, as shown in Theorem 2, even a piecewise smooth $\mathscr{B}$ with only one kink will have detrimental effects on the convergence rate of the misspecification bias of $\widehat{\vartheta}(\mathbf{x})$; cf., Figure 1.

5 Estimation and Inference

The results in this section are established for a general misspecification bias $\mathfrak{B}(\mathbf{x})$ of the distance-based estimator $\widehat{\vartheta}(\mathbf{x})$, which is determined by the properties of the kernel and distance functions as well as the geometry (smoothness) of the assignment boundary $\mathscr{B}$, as demonstrated by Theorems 2 and 3. We also discuss the implications of these results for implementation, and compare them to standard methods from univariate RD designs.

5.1 Convergence Rates

Using technical results established in the supplemental appendix, we obtain the following convergence rates for the univariate distance-based local polynomial treatment effect estimator.

Theorem 4 (Convergence Rates)**.** *Suppose Assumptions 1 and 2 hold. If* $nh^2/\log(1/h) \to \infty$ *and* $h \to 0$*, then*

(i) $|\widehat{\vartheta}(\mathbf{x}) - \tau(\mathbf{x})| \lesssim_{\mathbb{P}} \frac{1}{\sqrt{nh^2}} + \frac{1}{n^{\frac{1+v}{2+v}}h^2} + |\mathfrak{B}(\mathbf{x})|$ *for* $\mathbf{x} \in \mathscr{B}$*, and*

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

This theorem establishes the pointwise (for each $\mathbf{x} \in \mathscr{B}$) and uniform (over $\mathscr{B}$) convergence rates for the distance-based treatment effect estimator. At the minimum, by Theorem 2, $\widehat{\vartheta}(\mathbf{x}) \to_{\mathbb{P}} \tau(\mathbf{x})$, both pointwise and uniformly, because $h \to 0$. Furthermore, by Theorem 3(i), pointwise in $\mathbf{x} \in \mathscr{B}$, we obtain $|\mathfrak{B}(\mathbf{x})| \lesssim h^{p+1}$ for small enough h, and under regularity on $\mathscr{B}$. Similarly, by Theorem 3(ii), the uniform bias rate of $\widehat{\vartheta}(\mathbf{x})$ improves as the assignment boundary becomes smoother.

The "variance" components of the distance-based estimator $\widehat{\vartheta}(\mathbf{x})$ are of order $(nh^2)^{-1}$ despite its being a univariate local polynomial estimator, which would have naïvely suggested a convergence of order $(nh)^{-1}$ instead. But this is expected because the target estimand is the difference of two bivariate regression functions, and the nonparametric distance-based estimation approach cannot circumvent the curse of dimensionality. See also Section 6 for more discussion.

Combining Theorems 2 and 4(ii), and for appropriate choice of h, it follows that the distance-based local polynomial estimator $\widehat{\vartheta}(\mathbf{x})$ can achieve the uniform convergence rate $(n/\log n)^{-1/4}$. In Section 6, we show that for the class of rectifiable boundaries, the best isotropic nonparametric estimator (not just the distance-based local polynomial estimator) can achieve a uniform convergence rate no better than $n^{-1/4}$, which is the nonparametric mean square minimax rate for estimating bivariate Lipschitz functions [see, e.g., Tsybakov, 2008, and references therein]. This minimax result means that no matter which univariate distance-based estimator is used, it is always possible to find a rectifiable boundary and a data generating process such that the estimation error is no better than $n^{-1/4}$ for the best possible isotropic nonparametric estimator. It follows that the distance-based estimator $\widehat{\vartheta}(\mathbf{x})$ is minimax optimal, up to the $\log^{1/4}(n)$ term, for an appropriate choice of bandwidth sequence.

Finally, it is possible to establish valid pointwise, or integrated over $\mathscr{B}$, MSE convergence rates matching those in Theorem 4(i). We provide details in the supplemental appendix.

5.2 Uncertainty Quantification

To develop companion pointwise and uniform inference procedures along the treatment assignment boundary $\mathscr{B}$, we consider the feasible t-statistic for given a bandwidth choice and for each boundary point $\mathbf{x} \in \mathscr{B}$:

$$\widehat{\mathrm{T}}(\mathbf{x}) = \frac{\widehat{\vartheta}(\mathbf{x}) - \tau(\mathbf{x})}{\sqrt{\widehat{\Xi}_{\mathbf{x},\mathbf{x}}}},$$

where, using standard least squares algebra, for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathscr{B}$ and $t \in \{0,1\}$, we define

$$\widehat{\Xi}_{\mathbf{x}_1,\mathbf{x}_2} = \widehat{\Xi}_{0,\mathbf{x}_1,\mathbf{x}_2} + \widehat{\Xi}_{1,\mathbf{x}_1,\mathbf{x}_2}, \qquad \widehat{\Xi}_{t,\mathbf{x}_1,\mathbf{x}_2} = \frac{1}{nh^2}\mathbf{e}_1^\top \widehat{\boldsymbol{\Psi}}_{t,\mathbf{x}_1}^{-1} \widehat{\boldsymbol{\Upsilon}}_{t,\mathbf{x}_1,\mathbf{x}_2} \widehat{\boldsymbol{\Psi}}_{t,\mathbf{x}_2}^{-1} \mathbf{e}_1,$$

and

$$\begin{aligned}\widehat{\boldsymbol{\Upsilon}}_{t,\mathbf{x}_1,\mathbf{x}_2} = h^2 \mathbb{E}_n\Big[&\mathbf{r}_p\Big(\frac{D_i(\mathbf{x}_1)}{h}\Big) k_h(D_i(\mathbf{x}_1)) \big(Y_i - \mathbf{r}_p(D_i(\mathbf{x}_1))^\top \widehat{\boldsymbol{\gamma}}_t(\mathbf{x}_1)\big) \mathbb{1}(D_i(\mathbf{x}_1) \in \mathscr{I}_t) \\ &\times \mathbf{r}_p\Big(\frac{D_i(\mathbf{x}_2)}{h}\Big)^\top k_h(D_i(\mathbf{x}_2)) \big(Y_i - \mathbf{r}_p(D_i(\mathbf{x}_2))^\top \widehat{\boldsymbol{\gamma}}_t(\mathbf{x}_2)\big) \mathbb{1}(D_i(\mathbf{x}_2) \in \mathscr{I}_t)\Big].\end{aligned}$$

Thus, feasible confidence intervals and confidence bands over $\mathscr{B}$ take the form:

$$\widehat{\mathrm{I}}_\alpha(\mathbf{x}) = \left[\widehat{\vartheta}(\mathbf{x}) - \mathit{q}_\alpha \sqrt{\widehat{\Xi}_{\mathbf{x},\mathbf{x}}} \ , \ \widehat{\vartheta}(\mathbf{x}) + \mathit{q}_\alpha \sqrt{\widehat{\Xi}_{\mathbf{x},\mathbf{x}}} \right], \qquad \mathbf{x} \in \mathscr{B},$$

for any $\alpha \in (0,1)$, and where q_α denotes the appropriate quantile depending on the desired inference procedure.

For pointwise inference, we show in the supplemental appendix that $\sup_{t\in\mathbb{R}} |\mathbb{P}[\widehat{\mathrm{T}}(\mathbf{x}) \leq t] - \Phi(t)| \to 0$ for each $\mathbf{x} \in \mathscr{B}$, under standard regularity conditions, and provided that $nh^2\mathfrak{B}^2(\mathbf{x}) \to 0$. Thus, in this case, $\mathit{q}_\alpha = \Phi^{-1}(1-\alpha/2)$ is an asymptotically valid choice. For uniform inference (over $\mathscr{B}$), on the other hand, the stochastic process $(\widehat{\mathrm{T}}(\mathbf{x}) : \mathbf{x} \in \mathscr{B})$ is not asymptotically tight, and thus it does not converge weakly in the space of uniformly bounded real functions supported on $\mathscr{B}$ and equipped with the uniform norm [van der Vaart and Wellner, 1996, Giné and Nickl, 2016]. This implies that distribution theory cannot be based on the usual weak convergence arguments for empirical processes. Furthermore, the geometry of the manifold $\mathscr{B}$, and the fact that the estimator is based on the signed distance score, requires additional special technical care for establishing asymptotically valid inference procedures. We address these issues by we first establishing a novel strong approximation for the entire stochastic process $(\widehat{\mathrm{T}}(\mathbf{x}) : \mathbf{x} \in \mathscr{B})$, assuming the usual "small" bias property $nh^2 \sup_{\mathbf{x}\in\mathscr{B}} \mathfrak{B}^2(\mathbf{x}) \to 0$. We then choose the appropriate q_α for uniform inference because

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

Our technical work combines a new strong approximation theorem (Theorem SA-8 in the supplemental appendix) and important technical results from Chernozhukov et al. [2014a,b] and Chernozhuokov et al. [2022]. See the supplemental appendix for omitted technical details, and other theoretical results that may be of independent interest. Let $\mathscr{U}_n$ be the σ-algebra generated by $((Y_i, (D_i(\mathbf{x}) : \mathbf{x} \in \mathscr{B})) : 1 \leq i \leq n)$.

Theorem 5 (Statistical Inference)**.** *Suppose Assumptions 1 and 2 hold.*

(i) *For all* $\mathbf{x} \in \mathscr{B}$*, if* $n^{\frac{v}{2+v}} h^2 \to \infty$ *and* $nh^2 \mathfrak{B}^2(\mathbf{x}) \to 0$*, then*

$$\mathbb{P}\big[\tau(\mathbf{x}) \in \widehat{\mathrm{I}}_\alpha(\mathbf{x})\big] \to 1-\alpha,$$

for $\mathscr{q}_\alpha = \Phi^{-1}(1-\alpha/2)$.

(ii) *If* $n^{\frac{v}{2+v}} h^2/\log n \to \infty$*,* $\liminf_{n\to\infty} \frac{\log h}{\log n} > -\infty$*,* $nh^2 \sup_{\mathbf{x}\in\mathscr{B}} \mathfrak{B}^2(\mathbf{x}) \to 0$*, and* $\operatorname{perim}(\{\mathbf{y} \in \mathscr{A}_t : \mathscr{d}(\mathbf{y},\mathbf{x})/h \in \operatorname{Supp}(k)\}) \lesssim h$ *for all* $\mathbf{x} \in \mathscr{B}$ *and* $t \in \{0,1\}$*, then*

$$\mathbb{P}\big[\tau(\mathbf{x}) \in \widehat{\mathrm{I}}_\alpha(\mathbf{x}),\ \textit{for all } \mathbf{x} \in \mathscr{B}\big] \to 1-\alpha,$$

for $\mathscr{q}_\alpha = \inf\{c>0 : \mathbb{P}[\sup_{\mathbf{x}\in\mathscr{B}} |\widehat{Z}_n(\mathbf{x})| \geq c | \mathcal{U}_n] \leq \alpha\}$*, where* $(\widehat{Z}_n : \mathbf{x} \in \mathscr{B})$ *is a Gaussian process conditional on* $\mathcal{U}_n$*, with* $\mathbb{E}[\widehat{Z}_n(\mathbf{x}_1)|\mathcal{U}_n] = 0$ *and* $\mathbb{E}[\widehat{Z}_n(\mathbf{x}_1)\widehat{Z}_n(\mathbf{x}_2)|\mathcal{U}_n] = \widehat{\Xi}_{\mathbf{x}_1,\mathbf{x}_2} / \sqrt{\widehat{\Xi}_{\mathbf{x}_1,\mathbf{x}_1}\widehat{\Xi}_{\mathbf{x}_2,\mathbf{x}_2}}$*, for all* $\mathbf{x}_1, \mathbf{x}_2 \in \mathscr{B}$.

This theorem establishes asymptotically valid inference procedures using the distance-based local polynomial treatment effect estimator $\widehat{\vartheta}(\mathbf{x})$. For uniform inference, an additional restriction on the assignment boundary $\mathscr{B}$ is imposed: the De-Giorgi perimeter condition can be verified when the boundary of $\{\mathbf{y} \in \mathscr{A}_t : \mathscr{d}(\mathbf{y},\mathbf{x})/h \in \operatorname{Supp}(k)\}$ is a curve of length no greater than h up to a constant; since the set is contained in a h-ball centered at $\mathbf{x}$, the curve length condition holds as long as the curve $\mathscr{B} \cap \{\mathbf{y} \in \mathbb{R}^2 : \mathscr{d}(\mathbf{y},\mathbf{x}) \leq h\}$ is not "too wiggly". In other words, the one-dimensional boundary $\mathscr{B}$ can not be "too long".

5.3 Discussion and Implementation

Although the distance-based estimator $\widehat{\vartheta}(\mathbf{x})$ looks like a univariate local polynomial estimation procedure based on the scalar score variable $D_i(\mathbf{x})$, Theorem 4 shows that its pointwise and uniform variance convergence rates are equal to that of a bivariate nonparametric estimator (which are unimprovable). Theorem 5 shows that inference results derived using univariate local polynomial regression methods can be deployed directly in distance-based settings, provided the side conditions are satisfied. This result follows because $\widehat{\mathrm{T}}(\mathbf{x})$ is constructed as a self-normalizing statistic, and is therefore *adaptive* to the fact that the univariate covariate $D_i(\mathbf{x})$ is actually based on the bivariate covariate $\mathbf{X}_i$. This finding documents another advantage of employing pre-asymptotic variance

estimators and self-normalizing statistics for distributional approximation and inference [Calonico et al., 2018]. Therefore, standard estimation and inference methods from the univariate RD design literature [Calonico et al., 2014] can be deployed to BD designs employing univariate distance to the boundary, provided an appropriate bandwidth is chosen to ensure that the bias due to the shape of the assignment boundary $\mathscr{B}$ (documented in Section 4) is "small".

For implementation, consider first the case that $\mathfrak{B}(\mathbf{x}) \lesssim h^{p+1}$, that is, the assignment boundary $\mathscr{B}$ is smooth (in the sense of Theorem 3), or otherwise the bias is small. Establishing a precise MSE expansion for $\widehat{\vartheta}(\mathbf{x})$ is cumbersome due to the added complexity introduced by the distance transformation, but the convergence rates can be deduced from Theorem 4. The *incorrect* univariate MSE-optimal bandwidth is $h_{\mathtt{1d},\mathbf{x}} \asymp n^{-1/(3+2p)}$, while the *correct* MSE-optimal bandwidth is $h_{\mathtt{MSE},\mathbf{x}} \asymp n^{-1/(4+2p)}$, implying that taking the distance variable as the univariate score, and thus ignoring its intrinsic bivariate dimension, leads to a smaller-than-optimal bandwidth choice because $n^{-1/(3+2p)} < n^{-1/(4+2p)}$, thereby undersmoothing the point estimator (relative to the *correct* MSE-optimal bandwidth choice). As a consequence, the point estimator will not be MSE-optimal but rather exhibit more variance and less bias, and the associated inference procedures will be more conservative. A simple solution is to rescale the *incorrect* univariate bandwidth, $\frac{n^{1/(3+2p)}}{n^{1/(4+2p)}} h_{\mathtt{1d},\mathbf{x}}$, but this may not be necessary if the empirical implementation of $h_{\mathtt{1d},\mathbf{x}}$ employs a pre-asymptotic variance estimator, as in the software package `rdrobust` (`https://rdpackages.github.io/rdrobust/`). See Calonico et al. [2020], and references therein, for details on bandwidth selection for standard univariate RD designs.

When the assignment boundary $\mathscr{B}$ exhibits kinks or other irregularities, no matter how large the sample size or the polynomial order, there will always be a region near each kink where the bias will be "large", that is, of irreducible order h, as shown by Theorem 2. In this case, near each kink the *correct* MSE-optimal bandwidth is $h \asymp n^{-1/4}$. Away from each kink, the MSE-optimal bandwidths are as discussed in the preceding paragraph. This phenomenon is generated by the lack of smoothness of $\mathscr{B}$, and leads to different MSE-optimal bandwidths for each $\mathbf{x} \in \mathscr{B}$, thus making automatic implementation more difficult. A simple solution to this problem is to set $h = \widehat{\mathtt{C}} \cdot n^{-1/4}$ for all $\mathbf{x} \in \mathscr{B}$, and with $\widehat{\mathtt{C}}$ a rule-of-thumb estimate, which is generically suboptimal but always no larger than the pointwise MSE-optimal bandwidth, thus delivering a more variable (than optimal) point estimator and more conservative (than possible) associated inference procedure. Another

simple solution is to employ different bandwidths for different values of $\mathbf{x} \in \mathscr{B}$, taking into account the relative proximity to a kink point on $\mathscr{B}$. Specifically, estimation can employ the best approach depending on whether the evaluation point is "near" or "away" from a kink of $\mathscr{B}$:

$$\widehat{h}_{\texttt{kink},\mathbf{x}}(\mathscr{B}) = \min\left\{\widehat{h}_{\texttt{MSE},\mathbf{x}}, \max\left\{\widehat{\texttt{C}} \cdot n^{-1/4}, \min_{\mathbf{b}\in\mathscr{B}_{\texttt{kink}}} \mathscr{d}(\mathbf{x},\mathbf{b})\right\}\right\}.$$

for all $\mathbf{x} \in \mathscr{B}$, and where $\mathscr{B}_{\texttt{kink}}$ denotes the collection of kink points on $\mathscr{B}$.

Given a choice of bandwidth h, valid statistical inference can be developed by controlling the remaining misspecification bias. When the boundary $\mathscr{B}$ is smooth, robust bias correction for standard univariate RD designs continues to be valid in the context of distance-based estimation [Calonico et al., 2014, 2018, 2022]. When $\mathscr{B}$ exhibits kinks, undersmoothing relative to the MSE-optimal bandwidth for $p = 0$ is needed due to the fact that increasing p does not necessarily imply a reduction in misspecification bias, and thus bias correction techniques are ineffective (uniformly over $\mathbf{x} \in \mathscr{B}$). A simple solution is to employ different bandwidths and polynomial orders for different values of $\mathbf{x} \in \mathscr{B}$, taking into account the relative proximity to a kink point on $\mathscr{B}$.

The companion software package `rd2d` employs a simple rule-of-thumb bandwidth selector targeting $h \asymp n^{-1/4}$ to implement the univariate distance-based estimator $\widehat{\vartheta}(\mathbf{x})$ as default for all $\mathbf{x} \in \mathscr{B}$ when it is unknown whether $\mathscr{B}$ has kinks or other irregularities o, which is valid for estimation whether $\mathscr{B}$ is smooth or not. For inference, following Calonico et al. [2018], the package employs the undersmoothed choice of order $n^{-1/3}$, which targets coverage error optimality of confidence interval estimators. If $\mathscr{B}$ is known to be smooth, then the package implements a rule-of-thumb bandwidth selector targeting $h \asymp n^{-1/(4+2p)}$ for point estimation, and them employs robust bias corrected inference based on that bandwidth choice. Furthermore, if the location of the kinks is known, then the more adaptive bandwidth selection procedure $\widehat{h}_{\texttt{kink},\mathbf{x}}(\mathscr{B})$ can be used. Section 7 illustrates the performance of these methods, and also considers the software package `rdrobust` for bandwidth selection with $p = 0$ as another simple rule-of-thumb bandwidth implementation. Further implementation details and simulation evidence are given in Cattaneo et al. [2025d].

Finally, we explain how uniform inference is implemented based on Theorem 5(ii). In practice, we discretize along the boundary with point of evaluations $\mathbf{b}_1, \cdots, \mathbf{b}_M \in \mathscr{B}$, and hence the feasible (conditional) Gaussian process $(\widehat{Z}_n(\mathbf{x}) : \mathbf{x} \in \mathscr{B})$ is reduced to the M-dimensional (conditional)

Gaussian random vector $\widehat{\mathbf{Z}}_n = (\widehat{Z}_n(\mathbf{b}_1), \ldots, \widehat{Z}_n(\mathbf{b}_M))$ with covariance matrix having a typical element $\mathbb{E}[\widehat{Z}_n(\mathbf{b}_1)\widehat{Z}_n(\mathbf{b}_2)|\mathcal{U}]$. The estimated (discretized) covariate function may not be positive define in finite samples, in which case a simple regularization can be implemented as needed [see, e.g., Cattaneo et al., 2024a, for discussion and related technical results]. Finding $\mathscr{q}_\alpha$ reduces to finding the α-quantile of the distribution of $\max_{1\leq l\leq M} |\widehat{Z}_n(\mathbf{b}_l)|$, which can be easily simulated.

6 Distance-based Minimax Convergence Rate

Theorems 2 and 3 give precise conditions characterizing the pointwise and uniform (over $\mathscr{B}$) bias and convergence rate of the distance-based local polynomial estimator $\widehat{\tau}(\mathbf{x})$. It follows that the estimator can achieve at best the uniform convergence rate $(n/\log(n))^{-1/4}$ for an appropriate choice of the bandwidth sequence h. A natural follow-up theoretical (and potentially practical) question is whether a different distance-based estimator could do better. This section provide one answer to this question: if the boundary $\mathscr{B}$ is rectifiable, then no isotropic nonparametric estimator of a bivariate regression function can do better, up to polylog(n) terms, no matter how much additional smoothness in the underlying conditional expectation function is assumed.

This section is self-contained because it pertains to minimax theory for nonparametric curve estimation [see, e.g., Tsybakov, 2008, and references therein]. The following theorem presents a minimax uniform convergence rate result for estimation of bivariate compact supported nonparametric functions employing a class of isotropic nonparametric estimators.

Theorem 6 (Distance-based Minimax Convergence Rate)**.** *For a constants $q \geq 1$ and $L > 0$, let $\mathscr{P}_{\texttt{NP}} = \mathscr{P}_{\texttt{NP}}(L, q)$ be the class of (joint) probability laws $\mathbb{P}$ of $(Y_1, \mathbf{X}_1), \cdots, (Y_n, \mathbf{X}_n)$ satisfying the following:*

(i) *$((Y_i, \mathbf{X}_i) : 1 \leq i \leq n)$ are i.i.d. taking values in $\mathbb{R} \times \mathbb{R}^2$.*

(ii) *$\mathbf{X}_i$ admits a Lebesgue density f that is continuous on its compact support $\mathcal{X} \subseteq [-L, L]^2$, with $L^{-1} \leq \inf_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x}) \leq \sup_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x}) \leq L$, and $\mathscr{B} = \texttt{bd}(\mathcal{X})$ is a rectifiable curve.*

(iii) *$\mu(\mathbf{x}) = \mathbb{E}[Y_i|\mathbf{X}_i = \mathbf{x}]$ is q-times continuously differentiable on $\mathcal{X}$ with*

$$\max_{0\leq|\boldsymbol{\nu}|\leq\lfloor q\rfloor} \sup_{\mathbf{x}\in\mathcal{X}} |\partial^{\boldsymbol{\nu}}\mu(\mathbf{x})| + \max_{|\boldsymbol{\nu}|=q} \sup_{\mathbf{x}_1,\mathbf{x}_2\in\mathcal{X}} \frac{|\partial^{\boldsymbol{\nu}}\mu(\mathbf{x}_1) - \partial^{\boldsymbol{\nu}}\mu(\mathbf{x}_2)|}{\|\mathbf{x}_1 - \mathbf{x}_2\|^{q-\lfloor q\rfloor}} \leq L.$$

(iv) $\sigma^2(\mathbf{x}) = \mathbb{V}[Y_i|\mathbf{X}_i = \mathbf{x}]$ *is continuous on* $\mathcal{X}$ *with* $L^{-1} \leq \inf_{\mathbf{x}\in\mathcal{X}} \sigma^2(\mathbf{x}) \leq \sup_{\mathbf{x}\in\mathcal{X}} \sigma^2(\mathbf{x}) \leq L$.

In addition, let $\mathscr{T}$ *be the class of all distance-based estimators* $T_n(\mathbf{U}_n(\mathbf{x}))$ *with* $\mathbf{U}_n(\mathbf{x}) = ((Y_i, \|\mathbf{X}_i - \mathbf{x}\|)^\top : 1 \leq i \leq n)$ *for each* $\mathbf{x} \in \mathcal{X}$. *Then,*

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

where $\mathbb{E}_{\mathbb{P}}[\cdot]$ *denotes an expectation taken under the data generating process* $\mathbb{P}$.

In sharp contrast to Theorem 6, under the assumptions imposed, the uniform minimax convergence rate among all possible nonparametric estimators is

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

where $\mathcal{S}$ is the (unrestricted class) of estimators based on $(\mathbf{W}_n = (Y_i, \mathbf{X}_i^\top)^\top : 1 \leq i \leq n)$. Therefore, Theorem 6 shows that if we restrict the class of estimators to only using the distance $\|\mathbf{X}_i - \mathbf{x}\|$ to the evaluation point $\mathbf{x}$ (as opposed to using $\mathbf{X}_i$ directly) for estimation of $\mu(\mathbf{x})$, then for any such estimator there is a data generating process such that the largest estimation error uniformly along the boundary is at least of order $n^{-1/4}$ when the boundary can have countably many kinks. Notably, increasing the smoothness of the underlying bivariate regression function $\mu(\mathbf{x})$ does not improve the uniform estimation accuracy of estimators in $\mathscr{T}$ due to the possible lack of smoothness introduced by a non-smooth boundary $\mathscr{B}$ of the support $\mathcal{X}$.

Finally, we can connect the results in this section to the results for BD designs. For clarity, we consider one nonparametric regression estimation, as opposed to the difference of two such estimators. By Theorem 4 and Theorem 2, the univariate distance-based p-th order local polynomial nonparametric regression estimator

$$\widehat{\mu}(\mathbf{x}) = \mathbf{e}_1^\top \widehat{\boldsymbol{\gamma}}(\mathbf{x}), \qquad \widehat{\boldsymbol{\gamma}}(\mathbf{x}) = \operatorname*{arg\,min}_{\boldsymbol{\gamma}\in\mathbb{R}^{p+1}} \mathbb{E}_n\Big[\big(Y_i - \mathbf{r}_p(D_i(\mathbf{x}))^\top\boldsymbol{\gamma}\big)^2 k_h(D_i(\mathbf{x}))\Big], \qquad \mathbf{x} \in \mathscr{B},$$

satisfies

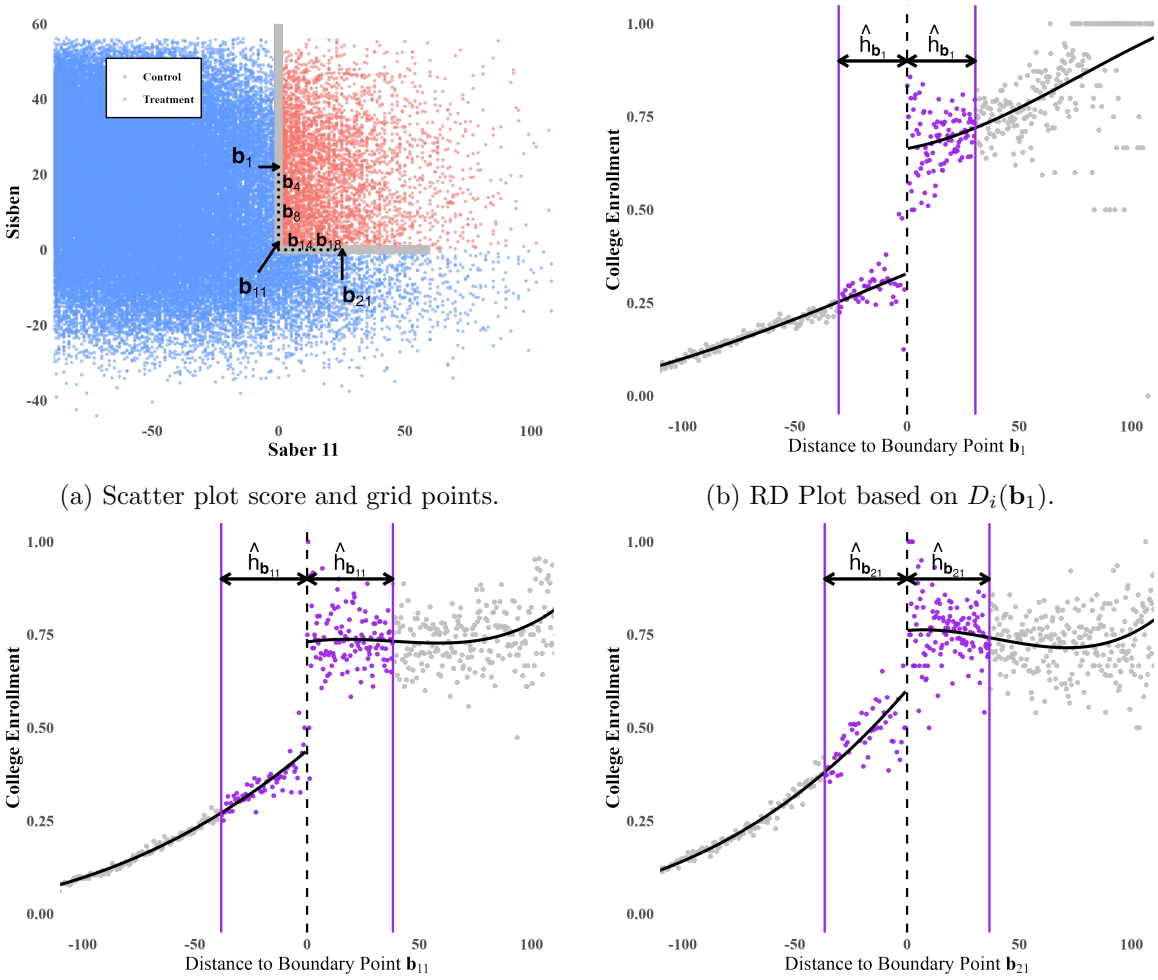
$$\limsup_{M\to\infty} \limsup_{n\to\infty} \sup_{\mathbb{P}\in\mathscr{P}} \mathbb{P}\Big[\Big(\frac{n}{\log n}\Big)^{1/4} \sup_{\mathbf{x}\in\mathscr{B}} \big|\widehat{\mu}(\mathbf{x}) - \mu(\mathbf{x})\big| \geq M\Big] = 0.$$

Since $\widehat{\mu}(\mathbf{x}) \in \mathscr{T}$ in Theorem 6, it follows that the distance-based local polynomial regression estimator is minimax optimal in the sense of Theorem 6, up to the $\log^{1/4}(n)$ factor, which we conjecture is unimprovable (i.e., the lower bound in Theorem 6 is only loose by the factor $\log^{1/4} n$).

7 Small Simulation Study

We illustrate the numerical performance of the distance-based methods with simulations calibrated using the real-world dataset from Londoño-Vélez et al. [2020], who analyzed the Colombian governmental post-secondary education subsidy *Ser Pilo Paga* (SPP). This anti-poverty policy provided tuition support for undergraduate college students attending high-quality, government-certified higher education institutions. SPP eligibility was based on a deterministic, bivariate cutoff combining merit and economic need: students had to achieve a SABER 11 high school exit exam score in the top 9 percent and have a SISBEN wealth index below a region-specific threshold. The resulting treatment assignment rule forms an L-shaped boundary $\mathscr{B}$, with one kink at the lowest levels of eligibility in both dimensions of the score (and two other potential kinks at the boundary of the support), but otherwise linear (hence "nice" and "smooth" in those sub-regions).

The dataset has $n = 363,096$ complete observations for the first cohort of the program (2014). Each observation $i = 1, \ldots, n$ corresponds to one student, and the bivariate score $\mathbf{X}_i = (X_{1i}, X_{2i})^\top = (\texttt{SABER11}_i, \texttt{SISBEN}_i)^\top$ is comprised of the student's SABER11 test score (ranging from -310 to 172) and SISBEN wealth index (ranging from -103.41 to 127.21). Without loss of generality, each component of the score is recentered at its cutoff for program eligibility and standardized to more naturally accommodate a common distance function and scalar bandwidth parameter. The resulting treatment assignment boundary is $\mathscr{B} = \{(\texttt{SABER11}, \texttt{SISBEN}) : (\texttt{SABER11}, \texttt{SISBEN}) \in \{\texttt{SABER11} \geq 0 \text{ and } \texttt{SISBEN} = 0\} \cup \{\texttt{SABER11} = 0 \text{ and } \texttt{SISBEN} \geq 0\}\}$. Since the support of the score $\mathscr{X}$ is compact, technically speaking, there are three kinks on the boundary $\mathscr{B}$: the point $(0, 0)$, which is the point of main interest for our purposes, and the two points where $\mathscr{B}$ intersects the boundary of $\mathscr{X}$. Our empirical analysis focuses on 20 evenly-spaced cutoff points, $\{\mathbf{b}_1, \ldots, \mathbf{b}_{20}\} \in \mathscr{B}$, which are positioned near the inner kink and away from the boundary of $\mathscr{X}$ to demonstrate the effect of having one kink on the bias of the distance-based estimator. We employ the Euclidean distance so that $D_i(\mathbf{x}) = \|\mathbf{X}_i - \mathbf{x}\|$ for each $\mathbf{x} \in \mathscr{B}$.



(a) Scatter plot score and grid points.

(b) RD Plot based on $D_i(\mathbf{b}_1)$.

(c) RD Plot based on $D_i(\mathbf{b}_{11})$.

(d) RD Plot based on $D_i(\mathbf{b}_{21})$.

Figure 2: SPP Data: Scatter plot, and sample of distance-based RD Plots.

All the results in this section are implemented using our companion software package `rd2d`, and the package `rdrobust` for RD designs with univariate score. Omitted details are given in the replication files available at https://rdpackages.github.io/. Figure 2 introduces the SPP data. First, Figure 2a presents a scatterplot of the bivariate score, the boundary, and the 20 grid points. Second, Figures 2b, 2c and 2d present three RD plots using the induced signed (univariate) distance scores $D_i(\mathbf{b}_1)$, $D_i(\mathbf{b}_{11})$ and $D_i(\mathbf{b}_{20})$, respectively.

The score $\mathbf{X}_i = (X_{1i}, X_{2i})^\top$ is taken to follow the distribution $(100 \cdot \mathsf{Beta}(3,4) - 25, 100 \cdot \mathsf{Beta}(3,4) - 25)^\top$ with independent components, which roughly mimics the SPP dataset. In addition, we employ the same L-shaped boundary as in the SPP application, with a kink at $\mathbf{x} = (0,0)^\top$. The potential

outcomes are generated by the regression models

$$Y_i(t) = \beta_{t,0} + X_{1i}\beta_{t,1} + X_{2i}\beta_{t,2} + \varepsilon_{t,i}, \qquad t \in \{0,1\}$$

where $\mathbf{X}_i$, $\varepsilon_{0,i}$ and $\varepsilon_{1,i}$ are mutually independent, and $\varepsilon_{t,i} \sim \mathsf{Normal}(0, \sigma_t^2)$, for $i = 1, 2, \cdots, n$. The coefficients were estimated using the SPP dataset. See Table 1 for the exact numerical values.

	$t=0$	$t=1$
$\beta_{t,0}$	3.35×10^{-1}	6.98×10^{-1}
$\beta_{t,1}$	2.52×10^{-3}	2.74×10^{-3}
$\beta_{t,2}$	-1.72×10^{-3}	-6.05×10^{-4}
σ_t	3.32×10^{-1}	4.35×10^{-1}

Table 1: Parameter values calibrated using SPP dataset.

The simulation study considers $n = 20,000$ and $2,000$ replications. We investigate measures of point estimation and inference, both pointwise and uniform over the boundary $\mathscr{B}$. The numerical results for the distance-based local polynomial regression procedures are reported in four tables: Table 2 assumes the boundary is smooth and thus employs the bandwidth choice $h = \widehat{h}_{\mathtt{MSE},\mathbf{b}}$; Table 3 employs the kink-adaptive bandwidth choice $h = \widehat{h}_{\mathtt{MSE},\mathbf{b}}(\mathscr{B})$, which leverages knowledge of the kink location; Table 4 employs the bandwidth choice $h = \widehat{C} \cdot n^{-1/4}$, which is valid regardless of whether one or more kinks are present in $\mathscr{B}$; and Table 5 employs the bandwidth choice $h = h_{\mathtt{1d},\mathbf{b}}$ from the software package `rdrobust`, which is valid for univariate RD designs, and produces undersmoothing in this BD design setting.

The main results are summarized in Figure 3, which plots the average of each point estimator across the $2,000$ replications, along with the population (linear) conditional expectation. The figure demonstrates that distance-based methods that ignore the presence of the kink in $\mathscr{B}$ have a larger bias than distance-based procedures that account for kinks or other irregularities in $\mathscr{B}$. This numerical finding corroborates our main theoretical result in Theorem 2. Interestingly, in this small calibrated simulation study, the relatively large bias does not have much of an impact on inference performance: the signal-to-noise ratio is relatively small, and hence coverage of confidence intervals and confidence bands is roughly close to their targeted nominal levels in all cases. This finding indicates that the large bias uncovered in this paper for distance-based methods can affect

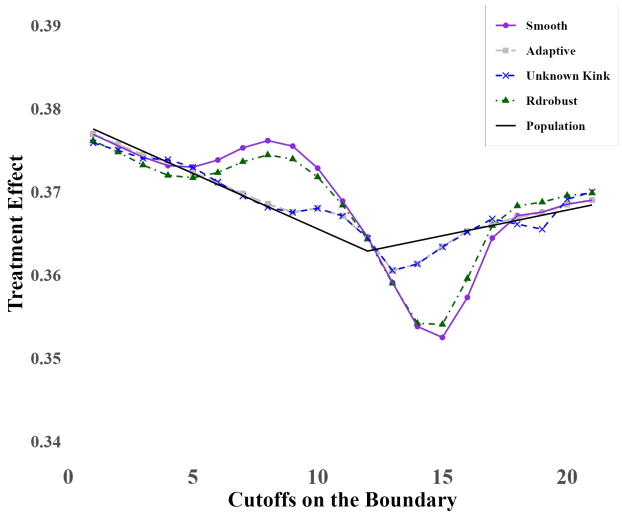


Figure 3: Average of Treatment effect Estimators: (i) *Smooth* denotes using bandwidths $h = \widehat{h}_{\mathtt{MSE},\mathbf{x}}$ under the assumption that the boundary is smooth; (ii) *Adaptive* denotes using bandwidths $\widehat{h}_{\mathtt{kink},\mathbf{x}}(\mathscr{B})$ that adapts to the kink given information about the kink location; (iii) *Unknown Kink* denotes using the bandwidth $h = \widehat{\mathtt{C}} \cdot n^{-1/4}$ under kink rates; (iv) *Rdrobust* denotes the (incorrect) univariate MSE optimal bandwidth $h_{\mathtt{1d},\mathbf{x}}$ from `rdrobust`; and (v) *Population* denotes the population BATEC causal parameter, $\tau(\mathbf{x})$, calibrated using the SPP dataset (see Table 1).

point estimation, without necessarily affecting uncertainty quantification. Of course, this is just one small calibrated simulation study, and in other settings with a larger signal-to-noise ratio, inference could also be affected whenever the distance-based procedures do not account for the presence of the kinks or other irregularities in the assignment boundary.

8 Final Remarks

We studied the statistical properties of distance-based (isotropic) local polynomial estimation in the context of BD designs. We presented necessary and sufficient conditions for identification, estimation and inference in large samples, both pointwise and uniformly along the assignment boundary. Our theoretical results highlighted the crucial role played by the regularity ("smoothness" and "length") of the boundary (a one-dimensional manifold) over which estimation and inference is conducted. Using our theoretical results, we offered precise guidance for empirical practice, and demonstrated the performance of our methods with simulated and real-world data. The companion general-purpose software package `rd2d` implements the main results [Cattaneo et al., 2025d]. Our results complement other recent theoretical and methodological work for analysis and interpreta-

tion of BD discontinuity designs, including Cattaneo et al. [2025b] for location-based methods and Cattaneo et al. [2025c] for boundary average treatment effect methods. See Cattaneo et al. [2026] for a recent review of the literature.

Our work can be generalized in several ways. For example, the results in the supplemental appendix could be extended to consider kink RD designs [Card et al., 2015], fuzzy RD designs [Arai et al., 2022], multi-cutoff RD designs [Cattaneo et al., 2016], and pre-treatment covariates for efficiency improvements [Calonico et al., 2019] or heterogeneity analysis [Calonico et al., 2025]. Furthermore, motivated by Theorem 3, which shows that distance-based methods can exhibit improved misspecification bias for "smooth" boundaries, we conjecture that new distance-based (isotropic) nonparametric smoothing estimators that either (i) take into account the specific geometry of the boundary (e.g., knowledge of the location of kinks in $\mathscr{B}$), or (ii) rely on regularization of the boundary (e.g., a smooth approximation of $\mathscr{B}$), could enjoy smaller misspecification bias. These generalizations are fruitful avenues for future research.

$\mathbf{b} \in \mathscr{B}$	h	Bias	SD	RMSE	EC	IL
$\mathbf{b}_1$	18.125	-0.001	0.019	0.020	0.948	0.137
$\mathbf{b}_2$	17.553	-0.001	0.020	0.020	0.946	0.140
$\mathbf{b}_3$	17.071	-0.001	0.020	0.020	0.946	0.142
$\mathbf{b}_4$	16.845	-0.000	0.020	0.020	0.952	0.143
$\mathbf{b}_5$	16.917	0.001	0.021	0.021	0.963	0.143
$\mathbf{b}_6$	17.231	0.003	0.021	0.021	0.955	0.140
$\mathbf{b}_7$	17.640	0.006	0.021	0.021	0.949	0.138
$\mathbf{b}_8$	18.056	0.008	0.020	0.022	0.944	0.136
$\mathbf{b}_9$	18.172	0.009	0.020	0.021	0.948	0.138
$\mathbf{b}_{10}$	18.267	0.007	0.020	0.021	0.949	0.140
$\mathbf{b}_{11}$	18.411	0.005	0.020	0.021	0.948	0.144
$\mathbf{b}_{12}$	19.992	0.002	0.020	0.020	0.947	0.147
$\mathbf{b}_{13}$	18.350	-0.004	0.020	0.021	0.953	0.142
$\mathbf{b}_{14}$	18.404	-0.010	0.019	0.022	0.956	0.137
$\mathbf{b}_{15}$	18.456	-0.012	0.019	0.023	0.952	0.133
$\mathbf{b}_{16}$	17.803	-0.008	0.021	0.022	0.957	0.136
$\mathbf{b}_{17}$	16.731	-0.002	0.021	0.021	0.961	0.144
$\mathbf{b}_{18}$	15.953	0.001	0.021	0.021	0.954	0.151
$\mathbf{b}_{19}$	16.058	0.000	0.021	0.021	0.947	0.151
$\mathbf{b}_{20}$	17.063	0.001	0.020	0.020	0.941	0.146
$\mathbf{b}_{21}$	18.746	0.001	0.019	0.019	0.941	0.137
Uniform					0.936	0.213

Table 2: Simulation results with $h = \widehat{h}_{\mathtt{MSE},\mathbf{b}}$ (smoothed $\mathscr{B}$)

$\mathbf{b} \in \mathscr{B}$	h	Bias	SD	RMSE	EC	IL
$\mathbf{b}_1$	18.087	-0.001	0.020	0.020	0.948	0.138
$\mathbf{b}_2$	17.354	-0.001	0.020	0.020	0.945	0.141
$\mathbf{b}_3$	16.511	-0.000	0.020	0.020	0.947	0.146
$\mathbf{b}_4$	15.516	-0.000	0.021	0.021	0.943	0.153
$\mathbf{b}_5$	13.954	0.000	0.023	0.023	0.959	0.169
$\mathbf{b}_6$	12.000	0.000	0.026	0.026	0.941	0.196
$\mathbf{b}_7$	10.000	0.000	0.032	0.032	0.941	0.235
$\mathbf{b}_8$	8.294	0.000	0.040	0.040	0.936	0.286
$\mathbf{b}_9$	7.963	0.001	0.043	0.043	0.940	0.305
$\mathbf{b}_{10}$	8.003	0.002	0.041	0.041	0.951	0.309
$\mathbf{b}_{11}$	8.066	0.003	0.043	0.043	0.937	0.316
$\mathbf{b}_{12}$	8.759	0.002	0.046	0.046	0.932	0.334
$\mathbf{b}_{13}$	8.039	-0.003	0.043	0.043	0.949	0.315
$\mathbf{b}_{14}$	8.063	-0.003	0.041	0.041	0.934	0.303
$\mathbf{b}_{15}$	8.548	-0.001	0.037	0.037	0.945	0.277
$\mathbf{b}_{16}$	11.200	-0.000	0.028	0.028	0.950	0.210
$\mathbf{b}_{17}$	13.951	0.000	0.023	0.023	0.949	0.169
$\mathbf{b}_{18}$	15.455	0.000	0.021	0.021	0.956	0.154
$\mathbf{b}_{19}$	15.988	0.000	0.021	0.021	0.948	0.152
$\mathbf{b}_{20}$	17.054	0.001	0.020	0.020	0.941	0.146
$\mathbf{b}_{21}$	18.746	0.001	0.019	0.019	0.941	0.137
Uniform					0.918	0.334

Table 3: Simulation results with $h = \widehat{h}_{\mathtt{MSE},\mathbf{b}}(\mathscr{B})$ (kink adaptive)

$\mathbf{b} \in \mathscr{B}$	h	Bias	SD	RMSE	EC	IL
$\mathbf{b}_1$	7.941	-0.002	0.042	0.042	0.945	0.158
$\mathbf{b}_2$	7.690	-0.001	0.042	0.042	0.944	0.161
$\mathbf{b}_3$	7.479	-0.001	0.043	0.043	0.943	0.163
$\mathbf{b}_4$	7.380	0.000	0.043	0.043	0.946	0.165
$\mathbf{b}_5$	7.411	0.001	0.042	0.042	0.960	0.164
$\mathbf{b}_6$	7.549	0.000	0.042	0.042	0.946	0.162
$\mathbf{b}_7$	7.728	-0.000	0.042	0.042	0.947	0.159
$\mathbf{b}_8$	7.911	-0.000	0.042	0.042	0.946	0.158
$\mathbf{b}_9$	7.961	0.001	0.043	0.043	0.943	0.159
$\mathbf{b}_{10}$	8.003	0.002	0.041	0.041	0.951	0.162
$\mathbf{b}_{11}$	8.066	0.003	0.043	0.043	0.943	0.166
$\mathbf{b}_{12}$	8.759	0.002	0.046	0.046	0.941	0.173
$\mathbf{b}_{13}$	8.039	-0.003	0.043	0.043	0.947	0.164
$\mathbf{b}_{14}$	8.063	-0.003	0.041	0.041	0.954	0.158
$\mathbf{b}_{15}$	8.086	-0.001	0.039	0.039	0.953	0.154
$\mathbf{b}_{16}$	7.800	-0.000	0.040	0.040	0.954	0.157
$\mathbf{b}_{17}$	7.330	0.001	0.043	0.043	0.941	0.166
$\mathbf{b}_{18}$	6.989	-0.000	0.044	0.044	0.956	0.174
$\mathbf{b}_{19}$	7.035	-0.002	0.045	0.045	0.946	0.175
$\mathbf{b}_{20}$	7.475	0.001	0.043	0.043	0.942	0.168
$\mathbf{b}_{21}$	8.213	0.002	0.041	0.041	0.940	0.156
Uniform					0.930	0.248

Table 4: Simulation results with $h = \widehat{C} \cdot n^{-1/4}$ (unknown kink location)

$\mathbf{b} \in \mathscr{B}$	h	Bias	SD	RMSE	EC	IL
$\mathbf{b}_1$	17.171	-0.001	0.021	0.021	0.950	0.145
$\mathbf{b}_2$	16.692	-0.001	0.021	0.021	0.945	0.147
$\mathbf{b}_3$	16.382	-0.002	0.021	0.021	0.951	0.149
$\mathbf{b}_4$	16.342	-0.002	0.022	0.022	0.955	0.149
$\mathbf{b}_5$	16.521	-0.001	0.022	0.022	0.964	0.148
$\mathbf{b}_6$	16.923	0.001	0.023	0.023	0.960	0.146
$\mathbf{b}_7$	17.517	0.004	0.023	0.024	0.951	0.143
$\mathbf{b}_8$	18.050	0.006	0.023	0.024	0.948	0.140
$\mathbf{b}_9$	18.633	0.007	0.022	0.023	0.945	0.138
$\mathbf{b}_{10}$	19.099	0.006	0.021	0.022	0.950	0.138
$\mathbf{b}_{11}$	19.707	0.004	0.021	0.021	0.953	0.138
$\mathbf{b}_{12}$	21.348	0.001	0.021	0.021	0.953	0.141
$\mathbf{b}_{13}$	19.520	-0.004	0.021	0.021	0.952	0.137
$\mathbf{b}_{14}$	19.133	-0.010	0.021	0.023	0.956	0.135
$\mathbf{b}_{15}$	18.682	-0.011	0.022	0.024	0.954	0.136
$\mathbf{b}_{16}$	17.531	-0.006	0.024	0.025	0.958	0.143
$\mathbf{b}_{17}$	16.033	-0.000	0.024	0.024	0.958	0.154
$\mathbf{b}_{18}$	15.139	0.002	0.022	0.022	0.958	0.160
$\mathbf{b}_{19}$	15.304	0.002	0.022	0.022	0.950	0.159
$\mathbf{b}_{20}$	16.065	0.002	0.022	0.022	0.946	0.154
$\mathbf{b}_{21}$	16.896	0.001	0.022	0.022	0.947	0.150
Uniform					0.947	0.219

Table 5: Simulation results: $h = h_{\texttt{1d},\mathbf{b}}$ (`rdrobust` + `rd2d.dist`)

References

Yoichi Arai, Yu-Chin Hsu, Toru Kitagawa, Ismael Mourifié, and Yuanyuan Wan. Testing identifying assumptions in fuzzy regression discontinuity designs. *Quantitative Economics*, 13(1):1–28, 2022.

Sudipto Banerjee. On geodetic distance computations in spatial modeling. *Biometrics*, 61(2): 617–625, 2005.

Sebastian Calonico, Matias D. Cattaneo, and Rocio Titiunik. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326, 2014.

Sebastian Calonico, Matias D. Cattaneo, and Max H. Farrell. On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association*, 113(522):767–779, 2018.

Sebastian Calonico, Matias D. Cattaneo, Max H. Farrell, and Rocio Titiunik. Regression discontinuity designs using covariates. *Review of Economics and Statistics*, 101(3):442–451, 2019.

Sebastian Calonico, Matias D. Cattaneo, and Max H. Farrell. Optimal bandwidth choice for robust bias corrected inference in regression discontinuity designs. *Econometrics Journal*, 23(2):192–210, 2020.

Sebastian Calonico, Matias D. Cattaneo, and Max H. Farrell. Coverage error optimal confidence intervals for local polynomial regression. *Bernoulli*, 28(4):2998–3022, 2022.

Sebastian Calonico, Matias D. Cattaneo, Max H. Farrell, Filippo Palomba, and Rocio Titiunik. Treatment effect heterogeneity in regression discontinuity designs. *Working paper*, 2025.

David Card, David S. Lee, Zhuan Pei, and Andrea Weber. Inference on causal effects in a generalized regression kink design. *Econometrica*, 83(6):2453–2483, 2015.

Matias D. Cattaneo and Rocio Titiunik. Regression discontinuity designs. *Annual Review of Economics*, 14:821–851, 2022.

Matias D. Cattaneo and Ruiqi (Rae) Yu. Strong approximations for empirical processes indexed by lipschitz functions. *Annals of Statistics*, 53(3):1203–1229, 2025.

Matias D. Cattaneo, Luke Keele, Rocio Titiunik, and Gonzalo Vazquez-Bare. Interpreting regression discontinuity designs with multiple cutoffs. *Journal of Politics*, 78(4):1229–1248, 2016.

Matias D. Cattaneo, Nicolás Idrobo, and Rocio Titiunik. *A Practical Introduction to Regression Discontinuity Designs: Foundations*. Cambridge University Press, 2020.

Matias D. Cattaneo, Yingjie Feng, and William G. Underwood. Uniform inference for kernel density estimators with dyadic data. *Journal of the American Statistical Association*, 119(524): 2695–2708, 2024a.

Matias D. Cattaneo, Nicolás Idrobo, and Rocio Titiunik. *A Practical Introduction to Regression Discontinuity Designs: Extensions*. Cambridge University Press, 2024b.

Matias D. Cattaneo, Ricardo Masini, and William G. Underwood. Yurinskii's coupling for martingales. *Annals of Statistics*, forthcoming, 2025a.

Matias D. Cattaneo, Rocio Titiunik, and Ruiqi (Rae) Yu. Estimation and inference in boundary discontinuity designs: Location-based methods. *arXiv:2505.05670*, 2025b.

Matias D. Cattaneo, Rocio Titiunik, and Ruiqi (Rae) Yu. Estimation and inference in boundary discontinuity designs: Pooling-based methods. *Working paper*, 2025c.

Matias D. Cattaneo, Rocio Titiunik, and Ruiqi (Rae) Yu. `rd2d`: Causal inference in boundary discontinuity designs. *arXiv:2505.07989*, 2025d.

Matias D. Cattaneo, Rocio Titiunik, and Ruiqi (Rae) Yu. Boundary discontinuity designs: Theory and practice. In *Invited book chapter for the 2025 Econometric Society World Congress*, volume 1, chapter 2. Cambridge University Press, 2026.

Xiaohong Chen and Wayne Yuan Gao. Semiparametric learning of integral functionals on submanifolds. *arXiv preprint arXiv:2507.12673*, 2025.

Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Anti-concentration and honest, adaptive confidence bands. *Annals of Statistics*, 42(5):1787–1818, 2014a.

Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Gaussian approximation of suprema of empirical processes. *Annals of Statistics*, 42(4):1564–1597, 2014b.

Victor Chernozhuokov, Denis Chetverikov, Kengo Kato, and Yuta Koike. Improved central limit theorem and bootstrap approximations in high dimensions. *Annals of Statistics*, 50(5):2562–2586, 2022.

Juan D Diaz and Jose R Zubizarreta. Complex discontinuity designs using covariates for policy impact evaluation. *Annals of Applied Statistics*, 17(1):67–88, 2023.

Herbert Federer. *Geometric measure theory*. Springer, 2014.

Sebastian Galiani, Patrick J. McEwan, and Brian Quistorff. External and internal validity of a geographic quasi-experiment embedded in a cluster-randomized experiment. In Matias D. Cattaneo and Juan Carlos Escanciano, editors, *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, volume 38)*, pages 195–236. Emerald Group Publishing, 2017.

Evarist Giné and Richard Nickl. *Mathematical Foundations of Infinite-dimensional Statistical Models*. Cambridge University Press, New York, 2016.

Jinyong Hahn, Petra Todd, and Wilbert van der Klaauw. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209, 2001.

Wolfgang Härdle, Marlene Müller, Stefan Sperlich, and Axel Werwatz. *Nonparametric and Semiparametric Models*. Springer, Heidelberg, 2004.

Miguel A. Hernán and James M. Robins. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC, 2020.

Luke Keele and Rocio Titiunik. Natural experiments based on geography. *Political Science Research and Methods*, 4(1):65–95, 2016.

Luke J. Keele and Rocio Titiunik. Geographic boundaries as regression discontinuities. *Political Analysis*, 23(1):127–155, 2015.

Luke J. Keele, Rocio Titiunik, and Jose Zubizarreta. Enhancing a geographic regression discontinuity design through matching to estimate the effect of ballot initiatives on voter turnout. *Journal of the Royal Statistical Society: Series A*, 178(1):223–239, 2015.

Luke J. Keele, Scott Lorch, Molly Passarella, Dylan Small, and Rocio Titiunik. An overview of geographically discontinuous treatment assignments with an application to children's health insurance. In Matias D. Cattaneo and Juan Carlos Escanciano, editors, *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, volume 38)*, pages 147–194. Emerald Group Publishing, 2017.

Juliana Londoño-Vélez, Catherine Rodríguez, and Fabio Sánchez. Upstream and downstream impacts of college merit-based financial aid for low-income students: Ser pilo paga in colombia. *American Economic Journal: Economic Policy*, 12(2):193–227, 2020.

John P Papay, John B Willett, and Richard J Murnane. Extending the regression-discontinuity approach to multiple assignment variables. *Journal of Econometrics*, 161(2):203–207, 2011.

Sean F Reardon and Joseph P Robinson. Regression discontinuity designs with multiple rating-score variables. *Journal of Research on Educational Effectiveness*, 5(1):83–104, 2012.

Maxime Rischard, Zach Branson, Luke Miratrix, and Luke Bornn. Do school districts affect nyc house prices? identifying border differences using a bayesian nonparametric approach to geographic regression discontinuity designs. *Journal of the American Statistical Association*, 116 (534):619–631, 2021.

Leon Simon et al. *Lectures on geometric measure theory*. Centre for Mathematical Analysis, Australian National University Canberra, 1984.

A.B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, 2008.

Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes*. Springer, 1996.

Vivian C Wong, Peter M Steiner, and Thomas D Cook. Analyzing regression-discontinuity designs with multiple assignment variables: A comparative study of four estimation methods. *Journal of Educational and Behavioral Statistics*, 38(2):107–141, 2013.