Estimation and Inference in Boundary Discontinuity Designs*

Matias D. Cattaneo† Rocio Titiunik‡ Ruiqi Yu§

March 2, 2025

Abstract

Boundary discontinuity designs are used to learn about treatment effects along a continuous boundary that splits units into control and treatment groups according to their bivariate score variable. These research designs are also called Multi-Score Regression Discontinuity designs, a leading special case being Geographic Regression Discontinuity designs. We study the statistical properties of commonly used local polynomial treatment effects estimators along the continuous treatment assignment boundary. We consider two distinct approaches: one based explicitly on the bivariate score variable for each unit, and the other based on their univariate distance to the boundary. For each approach, we present pointwise and uniform estimation and inference methods for the treatment effect function over the assignment boundary. Importantly, we show that methods based on univariate distance to the boundary exhibit an irreducible large misspecification bias when the assignment boundary has kinks or other irregularities, making the distance-based approach unsuitable for empirical work in those settings. In contrast, methods based on the bivariate score/location variable do not suffer from this drawback. We illustrate our methods with an empirical application and in simulations.

Keywords: regression discontinuity, uniform estimation and inference, causal inference.

*We thank Boris Hanin, Jason Klusowski, Juliana Londoño-Vélez, Xinwei Ma, Boris Shigida, and Mykhaylo Shkolnikov for comments. Cattaneo and Titiunik gratefully acknowledge financial support from the National Science Foundation (SES-2019432 and SES-2241575).

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

‡Department of Politics, Princeton University.

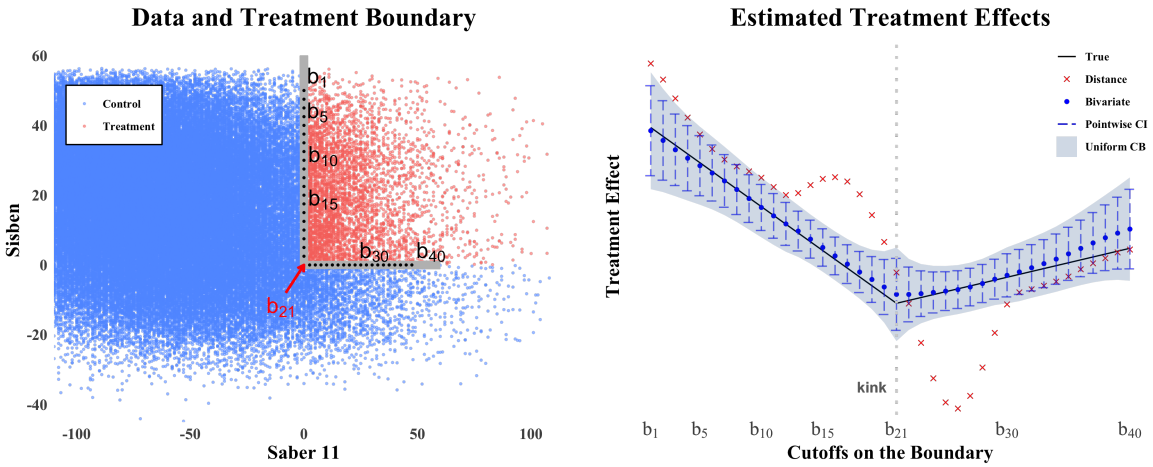
§Department of Operations Research and Financial Engineering, Princeton University.

1 Introduction

We study estimation and inference in boundary discontinuity designs (Black, 1999; Dell, 2010; Jardim et al., 2024), where the goal is to learn about causal treatment effects along a continuous boundary that splits units into control and treatment groups according to the value of their bivariate score/location variable. This setup is also known as a Multi-Score Regression Discontinuity (RD) design (Reardon and Robinson, 2012), a leading special case being the Geographic RD design and variations thereof (Keele and Titiunik, 2015, 2016; Keele et al., 2017; Galiani et al., 2017; Diaz and Zubizarreta, 2023). See Cattaneo and Titiunik (2022, Section 2.3) for an overview of the literature on Multi-dimensional RD designs, and Cattaneo et al. (2024, Section 5) for a practical introduction.

To describe the setup formally, suppose that $(Y_i(0), Y_i(1), \mathbf{X}_i)$, $i = 1, 2, \ldots, n$, is a random sample, where $Y_i(0)$ and $Y_i(1)$ denote the scalar potential outcomes for unit i under control and treatment assignment, respectively, and the score $\mathbf{X}_i = (X_{1i}, X_{2i})'$ is a continuous bivariate vector with support $\mathscr{X} \subseteq \mathbb{R}^2$. Units are assigned to either the control group or treatment group according to their location $\mathbf{X}_i$ relative to a known one-dimensional boundary curve $\mathscr{B}$ splitting the support $\mathscr{X}$ in 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 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 $\mathscr{A}_t$. 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 $\mathscr{B} \cap \mathscr{A}_0 = \emptyset$.

Boundary discontinuity designs are commonly used in quantitative social, behavioral, and biomedical sciences. For example, consider the substantive application analyzed in Londoño-Vélez, Rodríguez and Sánchez (2020). The authors studied the effects of a governmental subsidy for post-secondary education in Colombia, called *Ser Pilo Paga* (SPP), a social anti-poverty policy providing tuition support for four-year or five-year undergraduate college students in any government-certified higher education institution (HEI) with high-quality status. Eligibility to the program SSP was based on both merit and economic need: in order to qualify for the program, students had to obtain a high grade in Colombia's national standardized high school exit exam, SABER 11, and they had also to come from economically disadvantaged families, measured by a survey-based wealth index known as SISBEN. Eligibility followed a deterministic rule with a fixed bivariate cutoff: students had to obtain a SABER 11 score in the top 9 percent of scores or better, and they had to come



(a) Scatterplot and Boundary (SPP data). (b) Estimation and Inference (simulations).

Figure 1: Scatterplot, Treatment Boundary, and Estimation and Inference.

Note: Panel (a) presents a scatterplot of the bivariate score $\mathbf{X}_i$ using the SPP data, and also plots the treatment boundary $\mathscr{B}$ with 40 marked grid points. Panel (b) presents estimation and inference results over the 40 boundary grid points depicted in Panel (a) based on one simulated dataset calibrated using the real SPP data, and the solid black line depicts the population treatment effect curve $\tau(\mathbf{x})$ assumed linear and calibrated using the real SPP data.

from a household with SISBEN index below a region-specific threshold. Formally, each student was assigned a bivariate score $\mathbf{X}_i = (X_{1i}, X_{2i})^\top = (\texttt{SABER11}_i, \texttt{SISBEN}_i)^\top$, where $X_{1i} = \texttt{SABER11}_i$ recorded the SABER11 score and $X_{2i} = \texttt{SISBEN}_i$ recorded the SISBEN wealth score. After recentering each variable at its corresponding threshold, the treatment assignment boundary becomes $\mathscr{B} = \{(\texttt{SABER11}, \texttt{SISBEN}) : (\texttt{SABER11}, \texttt{SISBEN}) \in \{\texttt{SABER11} \geq 0 \text{ and } \texttt{SISBEN} = 0\} \cup \{\texttt{SABER11} = 0 \text{ and } \texttt{SISBEN} \geq 0\}\}$. Figure 1a presents a scatterplot of the bivariate score of the data of students in the 2014 cohort ($n = 363,096$ observations), and also plots the bivariate assignment boundary $\mathscr{B}$ together with 40 evenly-spaced cutoff points along the boundary.

Section 2 presents the core assumption underlying the causal inference framework used in this paper, which generalizes the standard unidimensional RD design to boundary discontinuity designs. Given the standard (continuity and finite-moments) conditions in Assumption 1, the goal is to conduct estimation and inference for the *average treatment effect curve along the boundary*:

$$\tau(\mathbf{x}) = \mathbb{E}[Y_i(1) - Y_i(0) | \mathbf{X}_i = \mathbf{x}], \qquad \mathbf{x} \in \mathscr{B},$$

both pointwise for each $\mathbf{x} \in \mathscr{B}$, and uniformly over $\mathscr{B}$. For example, in the SPP application, the outcome variable $Y_i = 1$ if student i attended college or $Y_i = 0$ otherwise, and thus the

causal parameter $\tau(\mathbf{x})$ captures the treatment effect of SPP on the probability of college education for students at the margin of program eligibility, as determined by their bivariate score $\mathbf{X}_i = (\texttt{SABER11}_i, \texttt{SISBEN}_i) = \mathbf{x} \in \mathscr{B}$. As a result, the parameter $\tau(\mathbf{x})$ captures policy-relevant heterogenous treatment effects along the boundary $\mathscr{B}$: for example, in Figure 1a, $\tau(\mathbf{b}_1)$ corresponds to the (local) average treatment effect for students with high Sisben score (wealth) and low Saber 11 Score (academic), while $\tau(\mathbf{b}_{40})$ corresponds to the (local) average treatment effect for students with low Sisben score and higher Saber 11 score. Identification of these boundary treatment effects parallels standard continuity-based univariate RD design arguments (Hahn et al., 2001): treatment assignment changes abruptly along the boundary $\mathscr{B}$, which implies that conditional expectations on each side of the assignment boundary can be used to identify $\tau(\mathbf{x})$ whenever there is no systematic "sorting" of units across the boundary, that is, whenever $\mathbb{E}[Y_i(0)|\mathbf{X}_i = \mathbf{x}]$ and $\mathbb{E}[Y_i(1)|\mathbf{X}_i = \mathbf{x}]$ are continuous for all $\mathbf{x} \in \mathscr{B}$ (Assumption 1 below). See Reardon and Robinson (2012), Keele and Titiunik (2015), and references therein, for more discussion.

Motivated by the local to the assignment boundary identifiability of $\tau(\mathbf{x})$, researchers employ flexible regression methods using only observations with score near the boundary $\mathscr{B}$. Local polynomial methods are the preferred choice for estimation and inference because they are simple (weighted) linear regression methods that intuitively incorporate localization to the assignment boundary. Two distinct implementations can be considered in this setting:

(i) regression analysis based on the *univariate distance* to the boundary $\mathscr{B}$, or

(ii) regression analysis based on the *bivariate location* relative to the boundary $\mathscr{B}$.

The first approach is more commonly used in practice because it is perceived as simpler, while the second approach is sometimes also encountered in applications. Despite their widespread use, however, there is no foundational understanding of their statistical properties and relative merits. This paper fills this gap in the literature by providing a comprehensive set of large sample results for each of the two approaches, which we then use to offer specific practical recommendations for the analysis and interpretation of boundary discontinuity designs. In particular, we provide pointwise and uniform (over $\mathscr{B}$) estimation and inference methods for both local polynomial regression approaches, and demonstrate theoretical and practical advantages of the approach based on bivariate location over the approach based on distance (albeit both are shown to be valid under appropriate

assumptions).

Using simulated data calibrated with the real SPP data (see the supplemental appendix for details), Figure 1b illustrates graphically the two local polynomial approaches for estimation of $\tau(\mathbf{x})$. The solid black line corresponds to the population treatment effect curve $\tau(\mathbf{x})$, taken to be linear in order to remove smoothing bias. The red crosses and blue dots correspond to local polynomial estimators based on, respectively, univariate distance to the boundary and bivariate location relative to the boundary, for the 40 grid points depicted on the assignment boundary $\mathscr{B}$ in Figure 1a. Noticeably, the estimator based on distance exhibits higher bias near the boundary kink point $\mathbf{x} = \mathbf{b}_{21}$, relative to the bivariate local polynomial estimator based on the bivariate score. The shaded region corresponds to a confidence band estimator based on the latter estimator. The underlying theoretical justification for these methods, as well as related implementation details such are bandwidth selection and robust bias correction inference, are novel contributions presented in the upcoming sections.

Specifically, Section 3 presents pointwise and uniform estimation and inference methods for local polynomial regression methods based on the univariate distance to the boundary $\mathscr{B}$. We begin by providing interpretable sufficient conditions (Assumption 2 below) for identification of $\tau(\mathbf{x})$, which restrict the distance function in conjunction with the univariate kernel function used and the shape of the boundary $\mathscr{B}$. We then present an important negative result: near kinks of the boundary $\mathscr{B}$, a pth order distance-based local polynomial estimator exhibits an irreducible bias of order h, the bandwidth used for implementation, no matter the polynomial order p used. This drawback is due to the fact that the underlying population regression function is at most Lipschitz continuous near kinks of the boundary $\mathscr{B}$. Figure 1b illustrated the phenomenon numerically, but our paper appears to be the first to provide a theoretical explanation for the large bias of the distance-based treatment effect estimator. In contrast, when the boundary $\mathscr{B}$ is smooth, we show that the pth order distance-based local polynomial estimator exhibits the usual bias of order h^{p+1}. Thus, our results show that the standard distance-based treatment effect estimator is consistent for $\tau(\mathbf{x})$, both pointwise and uniformly over $\mathscr{B}$, but it can exhibit a large bias affecting bandwidth selection and statistical inference in applications whenever the $\mathscr{B}$ has kinks or other irregularities. For uncertainty quantification, Section 3 presents pointwise and uniform (over $\mathscr{B}$) large sample distribution theory, which is used to propose both confidence intervals for $\tau(\mathbf{x})$ and confidence

bands for the entire treatment effect curve $(\tau(\mathbf{x}) : \mathbf{x} \in \mathscr{B})$. The last part of Section 3 discusses implementation, and also compares our results with related methods for standard univariate RD designs.

The analysis based on distance is not as straight forward as it may seem: standard results in the literature for univariate local polynomial regression need not be valid in this setting because the specific features of the assignment boundary can lead to a large misspecification bias (near kinks or other irregularities of $\mathscr{B}$). As a result, employing the distance-based approach can be detrimental in some applications. An alternative is to employ the bivariate location score directly. Section 4 studies this approach, and offers pointwise and uniform estimation and inference methods over $\mathscr{B}$. The pointwise results follow directly from the literature, provided an additional regularity condition on the bivariate kernel function and boundary $\mathscr{B}$, formalized in Assumption 3 below. On the other hand, the uniform inference results require some additional technical care, and appear to be new to the literature. The main potential issue is that a uniform distributional approximation is established over the lower-dimensional manifold $\mathscr{B}$, and thus its shape can affect the validity of the results. Section 4 also discusses new bandwidth selection methods based on mean square error (MSE) expansions, robust bias-corrected inference, and related implementation details. Our results provide natural generalizations of well-established results for univariate RD designs; see Calonico et al. (2020) for bandwidth selection, and Calonico et al. (2014, 2018, 2022) for robust bias correction.

Section 5 deploys our theoretical and methodological results to the SPP data, revising the main results reported in Londoño-Vélez, Rodríguez and Sánchez (2020). In addition to providing further empirical evidence in favor of their empirical findigns, we also find treatment effect heterogeneity along the assignment boundary $\mathscr{B}$. In addition, the supplemental appendix presents simulation evidence using the SPP data for calibration.

Section 6 concludes with specific recommendations for practice. The supplemental appendix presents generalizations of our theoretical results, reports their proofs, and gives other theoretical and methodological results that may be of independent interest. In particular, a new strong approximation result for empirical process with polynomial bounded moments is presented (building and extending recent work by Cattaneo and Yu, 2025).

2 Setup

Using the notation and causal inference framework in the introduction, we maintain throughout the paper the following basic conditions on the underlying data generating process.

Assumption 1 (Data Generating Process)**.** *Let* $t \in \{0, 1\}$.

(i) $(Y_1(t), \mathbf{X}_1)^\top, \ldots, (Y_n(t), \mathbf{X}_n)^\top$ *are independent and identically distributed random vectors with* $\mathcal{X} = [a_1, b_1] \times [a_2, b_2]$ *for* $-\infty < a_l < b_l < \infty$ *for* $l = 1, 2$.

(ii) *The distribution of* $\mathbf{X}_i$ *has a Lebesgue density* $f_X(\mathbf{x})$ *that is continuous and bounded away from zero on* $\mathcal{X}$.

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

This assumption is on par with the usual assumptions encountered in the classical RD literature with an univariate score. In particular, part (iii) imposes standard smoothness conditions on the bivariate conditional expectations of interest, which will play an important role in misspecification bias reduction (or lack thereof) in our upcoming results. Identification of $\tau(\mathbf{x})$ follows directly from those conditions (Reardon and Robinson, 2012; Keele and Titiunik, 2015, and references therein).

2.1 Notation

We employ standard concepts and notations from empirical process theory (Wellner et al., 2013; Giné and Nickl, 2016) and geometric measure theory (Simon et al., 1984; Federer, 2014). In particular, for a vector $\mathbf{v} \in \mathbb{R}^k$, $\|\mathbf{v}\| := (\sum_{i=1}^k \mathbf{v}_i^2)^{1/2}$. For a matrix $A \in \mathbb{R}^{m\times n}$, $\|A\| = \sup_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|$. For a Borel set $\mathcal{S} \subseteq \mathcal{X}$, the De Giorgi perimeter of $\mathcal{S}$ is $\texttt{perim}(\mathcal{S}) = \sup_{g\in\mathscr{D}_2(\mathcal{X})} \int_{\mathbb{R}^2} \mathbb{1}(\mathbf{x} \in \mathcal{S}) \operatorname{div} g(\mathbf{x}) d\mathbf{x} / \|g\|_\infty$, 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 $\mathcal{X}$. In the case $\mathcal{S}$ is connected and the boundary $\partial\mathcal{S}$ is a smooth simple closed curve, $\mathcal{S}$ simplifies to the curve length of $\partial\mathcal{S}$. 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 reals sequences $|a_n| = o(|b_n|)$ if $\limsup \frac{a_n}{b_n} = 0$, $|a_n| \lesssim |b_n|$ 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

$\limsup_{M\to\infty}\limsup_{n\to\infty}\mathbb{P}[|\frac{a_n}{b_n}|\geq M]=0$. $C^k(\mathcal{X},\mathcal{Y})$ denotes the class of k-times continuously differentiable functions from $\mathcal{X}$ to $\mathcal{Y}$, and $C^k(\mathcal{X})$ is a shorthand for $C^k(\mathcal{X},\mathbb{R})$. Let $\Phi(x)$ be the standard Gaussian cumulative distribution function.

3 Analysis based on Univariate Distance

For each unit $i=1,\dots,n$, define their scalar distance-based score $D_i(\mathbf{x})=\mathscr{d}(\mathbf{X}_i,\mathbf{x})(\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_1)-\mathbb{1}(\mathbf{X}_i\in\mathscr{A}_0))$ to the point $\mathbf{x}\in\mathscr{B}$, where $\mathscr{d}(\cdot,\cdot)$ denotes a distance function. It is customary to use 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 $\mathbf{x}=(x_1,x_2)^\top\in\mathscr{B}$ in applications, but other choices are sometimes encountered. For each $\mathbf{x}\in\mathscr{B}$, the setup thus reduces to a standard univariate RD design with $D_i(\mathbf{x})\in\mathbb{R}$ the score variable and $c=0$ the cutoff, where $D_i(\mathbf{x})\geq 0$ if unit i is assigned to treatment status and $D_i(\mathbf{x})<0$ if unit i is assigned to control status. The local polynomial treatment effect curve estimator based on distance is

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

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

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

with $\mathbf{r}_p(u)=(1,u,u^2,\cdots,u^p)^\top$ the usual univariate polynomial basis, $k_h(u)=k(u/h)/h$ for univariate kernel function $k(\cdot)$ and bandwidth parameter h, and $\mathscr{I}_0=(-\infty,0)$ and $\mathscr{I}_1=[0,\infty)$. See Cattaneo and Titiunik (2022) for a literature review of RD designs, and Cattaneo et al. (2020, 2024) for practical introductions. The univariate 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 $\mathscr{B}$.

We impose the following conditions on the underlying features of the distance-based local polynomial estimator $\widehat{\tau}_{\text{dis}}(\mathbf{x})$.

Assumption 2 (Univariate Distance-Based Kernel)**.** *Let $t\in\{0,1\}$.*

(i) *$\mathscr{d}:\mathbb{R}^2\mapsto[0,\infty)$ 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 $\mathbf{x}_1,\mathbf{x}_2\in\mathcal{X}$.*

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

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

Part (i) of this assumption requires the distance function be equivalent (up to constants) to the Euclidean distance, while Assumption 2(ii) imposes standard conditions on the (univariate) kernel function. The last part of this assumption is novel to the literature: it implicitly restricts the geometry of the boundary $\mathscr{B}$ relative to the kernel and distance functions. More precisely, it rules out settings where highly irregular boundary shapes will lead to regions with too "few" data: a necessary condition for the estimator $\widehat{\boldsymbol{\gamma}}_t(\mathbf{x})$ to be well-defined in large samples is that $\mathbb{P}[k_h(D_i(\mathbf{x}))\mathbb{1}(D_i(\mathbf{x}) \in \mathscr{I}_t)] = \int_{\mathscr{A}_t} k_h(\mathscr{d}(\mathbf{u}, \mathbf{x}))f_X(\mathbf{u})d\mathbf{u} > 0$. Our theoretical results show that Assumption 2(iii) is a simple sufficient condition, due to Assumption 1(ii). Commonly encountered treatment assignment boundaries and distance functions typically satisfy Assumption 2; a potential exception being highly irregular geographic boundaries.

3.1 Identification and Interpretation

For each treatment group $t \in \{0, 1\}$, boundary point $\mathbf{x} \in \mathscr{B}$, and distance-based score $D_i(\mathbf{x})$, the univariate distance-based local polynomial estimator $\widehat{\boldsymbol{\gamma}}_t(\mathbf{x})$ is the sample analog of the coefficients associated with the best (weighted, local) mean square approximation of the conditional expectation $\mathbb{E}[Y_i|D_i(\mathbf{x})]$ 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 $\widehat{\theta}_{\mathbf{x},t}(0) = \mathbf{e}_1^\top \widehat{\boldsymbol{\gamma}}_t(\mathbf{x})$, $\theta^*_{\mathbf{x},t}(0) = \mathbf{e}_1^\top \boldsymbol{\gamma}_t^*(\mathbf{x})$, and $\theta_{t,\mathbf{x}}(r) = \mathbb{E}[Y_i|D_i(\mathbf{x}) = r, D_i(\mathbf{x}) \in \mathscr{I}_t]$ for $r \in \mathbb{R}$, we have the standard least squares decomposition

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

for each group $t \in \{0, 1\}$, and where

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

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

In the decomposition (1), the first term is the stochastic linear representation of the centered estimator, $\widehat{\theta}_{\mathbf{x},t}(0) - \theta_{t,\mathbf{x}}(0)$, since it is an average of (unconditional) mean-zero random variables, the second term is the mean square approximation bias, and the third term is a non-linearity error arising from the convergence of the Gram matrix associated with the (weighted) least squares estimator.

Noting $\widehat{\tau}_{\text{dis}}(\mathbf{x}) = \widehat{\theta}_{\mathbf{x},1}(0) - \widehat{\theta}_{\mathbf{x},0}(0)$, the following lemma characterizes the target estimand of the distance-based local polynomial estimator.

Lemma 1 (Identification)**.** *Suppose Assumptions 1(i)–(iii) and 2(i) hold. Then,* $\tau(\mathbf{x}) = \theta_{1,\mathbf{x}}(0) - \theta_{0,\mathbf{x}}(0)$ *for all* $\mathbf{x} \in \mathscr{B}$.

This lemma is established by noting that, for each group $t \in \{0, 1\}$ and $r \in \mathbb{R}$,

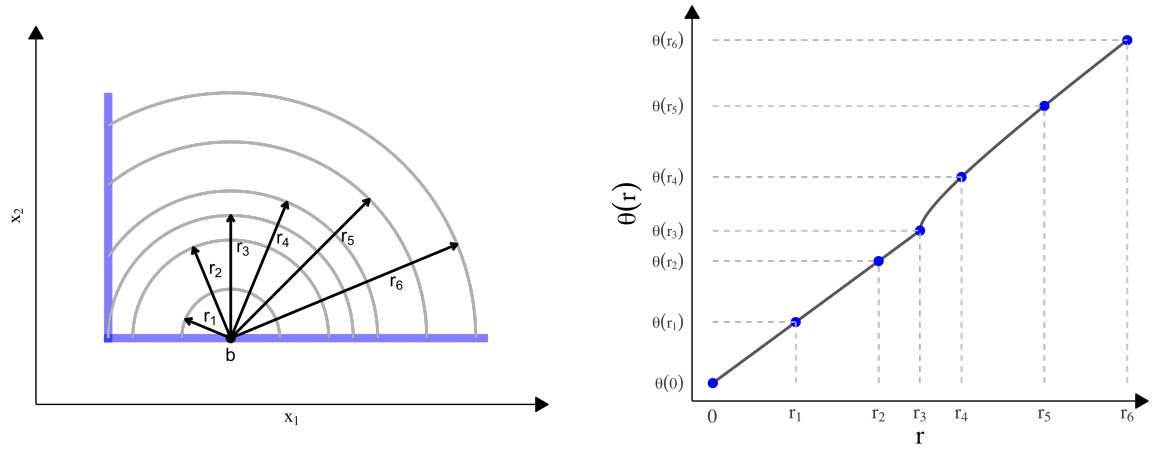
$$\begin{aligned}\theta_{t,\mathbf{x}}(r) &= \mathbb{E}\big[Y_i\big|D_i(\mathbf{x}) = r, D_i(\mathbf{x}) \in \mathscr{I}_t\big] \\ &= \mathbb{E}\big[Y_i\big|\mathscr{d}(\mathbf{X}_i, \mathbf{x}) = |r|, \mathbf{X}_i \in \mathscr{A}_t\big] = \mathbb{E}\big[Y_i(t)\big|\mathscr{d}(\mathbf{X}_i, \mathbf{x})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

and then verifying that $\lim_{u\downarrow 0}\mathbb{E}[Y_i(t)|\mathscr{d}(\mathbf{X}_i, \mathbf{x}) = u] = \mu_t(\mathbf{x})$ under the conditions imposed. Without restricting the data generating process, as well as the assignment boundary, distance function, and and kernel function, $\theta_{t,\mathbf{x}}(0)$ and $\mu_t(\mathbf{x})$ need not agree. Employing Lemma 1 and the decomposition (1), we obtain:

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

where $\mathfrak{L}_n(\mathbf{x}) = \mathbf{e}_1^\top \mathbf{\Psi}_{1,\mathbf{x}}^{-1}\mathbf{O}_{1,\mathbf{x}} - \mathbf{e}_1^\top \mathbf{\Psi}_{0,\mathbf{x}}^{-1}\mathbf{O}_{0,\mathbf{x}}$ is a mean-zero linear statistic, $\mathfrak{B}_n(\mathbf{x}) = \theta^*_{\mathbf{x},1}(0) - \theta^*_{\mathbf{x},0}(0) - \tau(\mathbf{x})$ is the bias of the estimator, and $\mathfrak{Q}_n(\mathbf{x}) = \mathbf{e}_1^\top(\widehat{\mathbf{\Psi}}_{1,\mathbf{x}}^{-1} - \mathbf{\Psi}_{1,\mathbf{x}}^{-1})\mathbf{O}_{t,\mathbf{x}} - \mathbf{e}_1^\top(\widehat{\mathbf{\Psi}}_{0,\mathbf{x}}^{-1} - \mathbf{\Psi}_{0,\mathbf{x}}^{-1})\mathbf{O}_{0,\mathbf{x}}$ is the higher-order linearization error.

In standard local polynomial regression settings, $\mathfrak{L}_n(\mathbf{x})$ is approximately Gaussian, $\mathfrak{B}_n(\mathbf{x})$ is of order h^{p+1}, and $\mathfrak{Q}_n(\mathbf{x})$ is neglected. However, because estimation is conducted along the boundary $\mathscr{B}$, not all of those standard results in the literature remain valid in the context of boundary



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

(b) Distance-based Conditional Expectation.

Figure 2: 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. The induced univariate conditional expectation $r \mapsto \theta_{1,\mathbf{b}}(r)$ is continuous but not differentiable at $r = r_3$.

discontinuity designs.

3.2 Bias Along the Boundary

Unlike the case of standard one-dimensional local polynomial estimation, in boundary discontinuity designs the smoothness of $\theta_{t,\mathbf{x}}(r) = \mathbb{E}[Y_i|D_i(\mathbf{x}) = r, D_i(\mathbf{x}) \in \mathscr{T}_t]$ depends on smoothness of the boundary $\mathscr{B}$. More specifically, the bias of the distance-based local polynomial estimator can be affected by the shape of the boundary $\mathscr{B}$, regardless of the polynomial order p used when constructing the estimator. Figure 2 demonstrates the problem graphically: for a point $\mathbf{x} \in \mathscr{B}$ that is close enough to a kink point $(0, 0)$ on the boundary, the conditional expectation $\theta_{1,\mathbf{x}}(r)$ is not differentiable for all $r \geq 0$. This 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 for a "large" r the arc is truncated by the boundary. As a result, for the example in Figure 2, $\theta_{1,\mathbf{x}}(r)$ 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, and provides further numerical evidence.

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 choice of polynomial order p. The following lemma formalizes this result, and also shows that this bias order cannot be improved by increasing $p \geq 0$.

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

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

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

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

If $nh^2 \to \infty$ and for any $p \geq 1$, then the uniform bias of the distance-based local polynomial estimator $\widehat{\tau}_{\text{dis}}(\mathbf{x})$ satisfies

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

The upper bound in Lemma 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 as Figure 2a, having a 90° kink at $\mathbf{x} = (0, 0)$. It follows that the conditions of Lemma 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 to establish a lower bound.

As a point of contrast, note that $\mathfrak{B}_n(\mathbf{x}) \lesssim_{\mathbb{P}} h^{p+1}$ pointwise in $\mathbf{x} \in \mathscr{B}$, for small enough h, provided that the kinks on the boundary $\mathscr{B}$ are sufficient far apart of each other relative to the bandwidth. However, Lemma 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 $\widehat{\tau}_{\text{dis}}(\mathbf{x})$ is at most of the

order h, regardless of the polynomial order p employed. The problem arises when the boundary $\mathscr{B}$ changes non-smoothly, leading to a non-differentiable regression function $\theta_{t,\mathbf{x}}(r) = \mathbb{E}[Y_i|D_i(\mathbf{x}) = r, D_i(\mathbf{x}) \in \mathscr{I}_t]$, $t \in \{0,1\}$, as illustrated in Figure 2. On the other hand, a better smoothing bias can be established if the boundary $\mathscr{B}$ is smooth enough.

Lemma 3 (Approximation Error: Smooth Boundary)**.** *Suppose Assumptions 1(i)-(iii) and 2 hold, with $\mathscr{d}(\cdot,\cdot)$ the Euclidean distance.*

(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, $\theta_{\mathbf{x},0}(\cdot)$ and $\theta_{\mathbf{x},1}(\cdot)$ are $(\kappa \wedge (p+1))$-times continuously differentiable on $[0,\varepsilon]$. Therefore, there exists a positive constant C such that $|\mathfrak{B}_n(\mathbf{x})| \leq Ch^{\kappa \wedge (p+1)}$ 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 $L > 0$. Then, there exists $\delta > 0$ such that for all $\mathbf{x} \in \gamma([\delta, L-\delta])$, $\theta_{\mathbf{x},0}(\cdot)$ and $\theta_{\mathbf{x},1}(\cdot)$ are $(\iota \wedge (p+1))$-times continuously differentiable on $[0,\delta]$, and $\lim_{r \downarrow 0} \frac{d^v}{dr^v}\theta_{\mathbf{x},t}(r)$ 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 $\sup_{\mathbf{x} \in \mathscr{B}} |\mathfrak{B}_n(\mathbf{x})| \leq Ch^{\iota \wedge (p+1)}$ for all $h \in [0,\delta]$.*

This lemma gives sufficient conditions in terms of smoothness of the boundary $\mathscr{B}$ to achieve a smaller order of misspecification bias of $\widehat{\tau}_{\text{dis}}(\mathbf{x})$, when compared to the minimal guarantee given by Lemma 2. In particular, for a uniform bound on misspecification bias, the sufficient condition in Lemma 3 requires the boundary $\mathscr{B}$ be *uniformly smooth* in that it can be parameterized by *one* smooth function. The example discussed after Lemma 2 already showed that the sufficient condition is also necessary, in the sense that even a piecewise smooth $\mathscr{B}$ with only one kink gives a uniform bias of order h.

In Appendix A, we also show that if we consider the class of rectifiable boundaries, then any estimator based on univariate distance can not achieve a uniform rate of consistency better than $n^{-1/4}$, that is, the non-parametric mean square minimax rate for estimating bivariate Lipschitz functions (Stone, 1982). And the local polynomial estimator based on univariate distance gives a uniform rate of consistency no larger than $n^{-1/4}\log n$, implying it is *nearly* minimax optimal (up to the $\log n$ term). This minimax result means that no matter which estimator based on distance is used, it is always possible to find a boundary among rectifiable curves and a member of the data

generating process class defined in the Appendix such that the estimation error is no less than $n^{-1/4}$. The instance that we construct for the proof involve countable many kinks.

3.3 Treatment Effect Estimation and Inference

Due to the bias issues discussed, this section develops estimation and inference results for $\widehat{\tau}_{\text{dis}}(\mathbf{x})$ under high-level conditions on $\mathfrak{B}_n(\mathbf{x})$. The next section discusses implementation, and compares our results to methods developed for standard RD designs.

In the supplemental appendix, we provide precise bounds on the linear error $\mathfrak{L}_n(\mathbf{x})$ and the non-linear error $\mathfrak{Q}_n(\mathbf{x})$, both pointwise and uniformly over $\mathscr{B}$. Those results give the following convergence rates for the distance-based local polynomial treatment effect estimator.

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

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

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

This theorem establishes the poinwise and uniform (over $\mathscr{B}$) convergence rates for the distance-based treatment effect estimator. By Lemma 2, $\widehat{\tau}_{\text{dis}}(\mathbf{x}) \to_{\mathbb{P}} \tau(\mathbf{x})$, both pointwise and uniformly, if $h \to 0$. However, $\widehat{\tau}_{\text{dis}}(\mathbf{x})$ has a variance convergence rate of order $n^{-1}h^2$ despite being a "univariate" local polynomial estimator, which would have nïvely suggested a variance convergence of order $n^{-1}h$ instead. In addition to exhibiting a bivariate curse of dimensionality, the convergence rate of the treatment effect estimator $\widehat{\tau}_{\text{dis}}(\mathbf{x})$ along the boundary $\mathscr{B}$ is determined by the smoothness of the $\mathscr{B}$, that is, the presence of kinks. It is not difficult to establish valid pointwise and integrated over $\mathscr{B}$ MSE convergence rates analogous to those in Theorem 1(i). See the supplemental appendix for details. Section 3.4 discusses the implications of these results for implementation, leveraging standard methods from univariate RD designs.

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

$$\widehat{\mathsf{T}}_{\text{dis}}(\mathbf{x}) = \frac{\widehat{\tau}_{\text{dis}}(\mathbf{x}) - \tau(\mathbf{x})}{\sqrt{\widehat{\boldsymbol{\Xi}}_{\mathbf{x},\mathbf{x}}}},$$

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

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

and

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

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

$$\widehat{I}_{\text{dis}}(\mathbf{x};\alpha) = \left[\, \widehat{\tau}_{\text{dis}}(\mathbf{x}) - \mathscr{q}_\alpha \sqrt{\widehat{\mathbf{\Xi}}_{\mathbf{x},\mathbf{x}}} \, , \, \widehat{\tau}_{\text{dis}}(\mathbf{x}) + \mathscr{q}_\alpha \sqrt{\widehat{\mathbf{\Xi}}_{\mathbf{x},\mathbf{x}}} \, \right], \qquad \mathbf{x} \in \mathscr{B},$$

for any $\alpha \in [0,1]$, and where $\mathscr{q}_\alpha$ denotes the appropriate quantile depending on the desired inference procedure.

For pointwise inference, it is straightforward to show that $\sup_{t\in\mathbb{R}} |\mathbb{P}[\widehat{\mathsf{T}}_{\text{dis}}(\mathbf{x}) \leq t] - \Phi(t)| \to 0$ for each $\mathbf{x} \in \mathscr{B}$, under standard regularity conditions, and provided that $\sqrt{nh^2}|\mathfrak{B}_n(\mathbf{x})| \to 0$. Thus, in this case, $\mathscr{q}_\alpha = \Phi^{-1}(1-\alpha/2)$ is an asymptotically valid choice. For uniform inference, we first establish a novel strong approximation for the entire stochastic process $(\widehat{\mathsf{T}}_{\text{dis}}(\mathbf{x}) : \mathbf{x} \in \mathscr{B})$, assuming that $\sqrt{nh^2}\sup_{\mathbf{x}\in\mathscr{B}} |\mathfrak{B}_n(\mathbf{x})| \to 0$. With some technical work, we can then choose the appropriate $\mathscr{q}_\alpha$ for uniform inference because

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

and the distribution of $\sup_{\mathbf{x}\in\mathscr{B}} \big|\widehat{\mathsf{T}}_{\text{dis}}(\mathbf{x})\big|$ can be deduced from the strong approximation of the stochastic process $(\widehat{\mathsf{T}}_{\text{dis}}(\mathbf{x}) : \mathbf{x} \in \mathscr{B})$.

Theorem 2 (Statistical Inference)**.** *Suppose Assumptions 1 and 2 hold. Let $\mathscr{W}$ be the σ-algebra generated by $((Y_i, (D_i(\mathbf{x}) : \mathbf{x} \in \mathscr{B})) : 1 \leq i \leq n)$.*

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

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

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

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

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

for $\mathcal{q}_\alpha = \inf\{c>0 : \mathbb{P}[\sup_{\mathbf{x}\in\mathscr{B}} |\widehat{Z}_n(\mathbf{x})| \geq c|\mathscr{W}] \leq \alpha\}$, *where* $(\widehat{Z}_n : \mathbf{x} \in \mathscr{B})$ *is a Gaussian process conditional on* $\mathscr{W}$, *with* $\mathbb{E}[\widehat{Z}_n(\mathbf{x}_1)|\mathscr{W}] = 0$ *and* $\mathbb{E}[\widehat{Z}_n(\mathbf{x}_1)\widehat{Z}_n(\mathbf{x}_2)|\mathscr{W}] = \widehat{\mathbf{\Xi}}_{\mathbf{x}_1,\mathbf{x}_2}/\sqrt{\widehat{\mathbf{\Xi}}_{\mathbf{x}_1,\mathbf{x}_1}\widehat{\mathbf{\Xi}}_{\mathbf{x}_2,\mathbf{x}_2}}$, *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 $\widehat{\tau}_{\mathrm{dis}}(\mathbf{x})$. For uniform inference, an additional mild restriction on the assignment boundary $\mathscr{B}$ is imposed: the De-Giorgi perimeter condition can be verified when the boundary of $\{\mathbf{y} \in \mathscr{A}_t : \mathscr{d}(\mathbf{y},\mathbf{x})/h \in \mathrm{Supp}(k)\}$ 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".

3.4 Implementation and Discussion

Although the distance-based estimator $\widehat{\tau}_{\mathrm{dis}}(\mathbf{x})$ looks like an univariate local polynomial estimation procedure, based on the scalar score variable $D_i(\mathbf{x})$, Theorem 1 shows that its pointwise and uniform variance convergence rate is equal to that of a bivariate nonparametric estimator, which are sharp. Nevertheless, Theorem 2 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 from the fact that $\widehat{\mathsf{T}}_{\mathrm{dis}}(\mathbf{x})$ is constructed as a self-normalizing statistic, and therefore it is *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). It follows that standard estimation and inference methods from the univariate RD design literature remain valid, provided an appropriate bandwidth is chosen and the induced bias due to the shape of the assignment boundary $\mathscr{B}$ is small (or accounted for).

For implementation, consider first the case that $\mathfrak{B}_n(\mathbf{x}) \lesssim h^{p+1}$, that is, the assignment boundary

$\mathscr{B}$ is smooth. Establishing a precise MSE expansion for $\widehat{\tau}_{\text{dis}}(\mathbf{x})$ is cumbersome due to the added complexity introduced by the distance transformation, but the convergence rates can be deduced from Theorem 1. The *incorrect* univariate MSE-optimal bandwidth is $h_{1d} \asymp n^{-1/(3+2p)}$, while the *correct* MSE-optimal bandwidth is $h_{\text{dis}} \asymp n^{-1/(4+2p)}$, implying that taking the distance variable as the univariate covariate, and thus ignoring its intrinsic bivariate dimension, implies that the resulting bandwidth choice will lead to a smaller bandwidth because $n^{-1/(3+2p)} < n^{-1/(4+2p)}$, that is, undersmoothing the point estimator (relative to the optimal MSE 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. An obvious solution is to rescale the *incorrect* univariate $\frac{n^{1/(3+2p)}}{n^{1/(4+2p)}} h_{1d}$, but this may not be necessary if the empirical implementation of h_{1d} employs a pre-asymptotic variance estimator, as in the software package `rdrobust` (`https://rdpackages.github.io/rdrobust/`). See Calonico et al. (2020), and reference therein, for details on bandwidth selection for standard univariate RD designs.

When the assignment boundary $\mathscr{B}$ exhibits kinks, no matter how large the sample size or the polynomial order chosen, there will always be a region near each kink where the bias will be large, that is, of irreducible order h, by Lemma 2. In this case, near each kink the *incorrect* univariate MSE-optimal bandwidth is $h_{1d} \asymp n^{-1/3}$, while the *correct* MSE-optimal bandwidth is $h_{\text{dis}} \asymp n^{-1/4}$. Away from each kink, the MSE-optimal bandwidths are as discussed in the preceeding 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 $h_{\text{dis}} \asymp n^{-1/4}$ for all $\mathbf{x} \in \mathscr{B}$, 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. In our companion software, `rd2d`, we employ the software package `rdrobust` for bandwidth selection with $p = 0$ as a simple rule-of-thumb, and leave for future work developing a bandwidth selection procedure that is adaptive to the potential presence of kinks in $\mathscr{B}$. Importantly, as discussed in Section 4, estimation and inference using directly the bivariate $\mathbf{X}_i$ solves this problem automatically.

Given a choice of (MSE-optimal) 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). On the other hand, 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 $\mathbf{x} \in \mathscr{B}$).

Having discussed the role of bandwidth choices, and the related issue of bias induced by the lack of smoothness of $\mathscr{B}$, it remains to explain how uniform inference is implemented based on Theorem 5(ii). In practice, we discretize along the boundary with point of evaluations $\mathbf{x}_1, \cdots, \mathbf{x}_M \in \mathscr{B}$, as in Figure 1a, 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 $\widehat{\mathbf{Z}}_n = (\widehat{Z}_n(\mathbf{x}_1), \ldots, \widehat{Z}_n(\mathbf{x}_M))$ with covariance matrix having a typical element $\mathbb{E}[\widehat{Z}_n(\mathbf{x}_1)\widehat{Z}_n(\mathbf{x}_2)|\mathscr{W}]$. Finding $\mathscr{q}_\alpha$ boils down to finding the α-quantile of the distribution of $\max_{1\leq l\leq M} |\widehat{Z}_n(\mathbf{x}_l)|$, which can be easily simulated. Section 5 illustrates this approach empirically.

4 Analysis based on Bivariate Location

The previous section demonstrated the potential detrimental aspects of employing distance-based local polynomial regression to analyze boundary discontinuity designs. When the boundary has kinks, as it is common in many applications based on sharp or geographic assignment rules, univariate methods based on distance to the boundary can lead to large biases near kinks or other irregularities on the assignment boundary. This section shows that an easy way to avoid the distance-based bias is to employ bivariate local polynomial regression methods.

The location-based treatment effect curve estimator of $\tau(\mathbf{x})$ is

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

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

$$\widehat{\boldsymbol{\beta}}_t(\mathbf{x}) = \underset{\boldsymbol{\beta}\in\mathbb{R}^{\mathfrak{p}_p+1}}{\arg\min}\, \mathbb{E}_n\Big[\big(Y_i - \mathbf{R}_p(\mathbf{X}_i - \mathbf{x})^\top \boldsymbol{\beta}\big)^2 K_h(\mathbf{X}_i - \mathbf{x}) \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\Big],$$

with $\mathfrak{p}_p = (2 + p)(1 + p)/2 - 1$, $\mathbf{R}_p(\mathbf{u}) = (1, u_1, u_2, u_1^2, u_2^2, u_1u_2, \cdots, u_1^p, u_2^p)^\top$ denotes the pth order polynomial expansion of the bivariate vector $\mathbf{u} = (u_1, u_2)^\top$, $K_h(\mathbf{u}) = K_h(u_1/h, u_2/h)/h^2$ for a

bivariate kernel function $K(\cdot)$ and a bandwidth parameter h. We employ the same bandwidth for both dimensions of $\mathbf{X}_i$ only for simplicity, and because it is common practice to first standardize each dimension of the bivariate score.

We impose the following assumption on the bivariate kernel function and assignment boundary.

Assumption 3 (Kernel Function)**.** *Let* $t \in \{0, 1\}$.

(i) *Either* $K : \mathbb{R} \to [0, \infty)$ *is compact supported and Lipschitz continuous, or* $K(\mathbf{u}) = \mathbb{1}(\mathbf{u} \in [-1, 1]^2)$.

(ii) $\liminf_{h \downarrow 0} \inf_{\mathbf{x} \in \mathscr{B}} \int_{\mathscr{A}_t} K_h(\mathbf{u} - \mathbf{x}) d\mathbf{u} \gtrsim 1$.

The first condition is standard in the literature, while the second condition is new but rather minimal. Similar to the case of distance-based estimation (Assumption 2(iii)), the goal of Assumption 3(ii) is to ensure enough data is available in large samples for each point $\mathbf{x} \in \mathscr{B}$ because $\mathbb{P}[K_h(\mathbf{X}_i - \mathbf{x})\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)] \gtrsim \int_{\mathscr{A}_t} K_h(\mathbf{u} - \mathbf{x}) d\mathbf{u}$ under Assumption 1. Given these conditions, pointwise and uniform estimation results, as well as valid MSE expansions, follow from standard local polynomial calculations and empirical process theory. On the other hand, uniform distribution theory needs to be established with some extra technical care: the supplemental appendix establishes a key new strong approximation result that takes into account the specific features of the estimator and manifold $\mathscr{B}$.

4.1 Treatment Effect Estimation and Inference

Using standard concentration techniques from empirical process theory, we obtain the pointwise and uniform convergence rate of $\widehat{\tau}(\mathbf{x})$.

Theorem 3 (Rate of Convergence)**.** *Suppose Assumptions 1 and 3 hold. If* $nh^2/\log(n) \to \infty$*, then*

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

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

This theorem immediately establishes consistency of the treatment effect estimator based on the bivariate location score, provided that $h \to 0$. More importantly, the theorem shows that the bias is of order h^{p+1} regardless of whether there are kinks or other irregularities in $\mathscr{B}$, under the

assumptions imposed. Comparing Theorems 1 and 3, it follows that both estimators can achieve the same (optimal) convergence rate when the boundary $\mathscr{B}$ is smooth, but otherwise $\widehat{\tau}(\mathbf{x})$ can achieve a faster (indeed, optimal) convergence rate, while $\widehat{\tau}_{\text{dis}}(\mathbf{x})$ cannot because of the large order-h bias when $\mathscr{B}$ is not smooth documented in Lemma 2.

Given its more standard structure, it is possible to establish a pointwise and integrated (conditional) MSE expansion for the estimator $\widehat{\tau}(\mathbf{x})$. Using standard multi-index notation, define the leading conditional bias $\mathbf{B}_{\mathbf{x}} = \mathbf{B}_{1,\mathbf{x}} - \mathbf{B}_{0,\mathbf{x}}$ with

$$\mathbf{B}_{t,\mathbf{x}} = \mathbf{e}_1^\top \widehat{\mathbf{\Gamma}}_{t,\mathbf{x}}^{-1} \sum_{|\mathbf{k}|=p+1} \frac{\mu_t^{(\mathbf{k})}(\mathbf{x})}{\mathbf{k}!} \mathbb{E}_n\Big[\mathbf{r}_p\Big(\frac{\mathbf{X}_i-\mathbf{x}}{h}\Big)\Big(\frac{\mathbf{X}_i-\mathbf{x}}{h}\Big)^{\mathbf{k}} K_h(\mathbf{X}_i-\mathbf{x})\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\Big]$$

and $\widehat{\mathbf{\Gamma}}_{t,\mathbf{x}} = \mathbb{E}_n\big[\mathbf{r}_p\big(\frac{\mathbf{X}_i-\mathbf{x}}{h}\big)\mathbf{r}_p\big(\frac{\mathbf{X}_i-\mathbf{x}}{h}\big)^\top K_h(\mathbf{X}_i-\mathbf{x})\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\big]$, for $t \in \{0,1\}$. Similarly, the leading conditional variance is $\mathbf{V}_{\mathbf{x}} = \mathbf{V}_{1,\mathbf{x}} + \mathbf{V}_{0,\mathbf{x}}$ with $\mathbf{V}_{t,\mathbf{x}} = \mathbf{e}_1^\top \widehat{\mathbf{\Gamma}}_{t,\mathbf{x}}^{-1} \mathbf{\Sigma}_{t,\mathbf{x},\mathbf{x}} \widehat{\mathbf{\Gamma}}_{t,\mathbf{x}}^{-1} \mathbf{e}_1$ and

$$\mathbf{\Sigma}_{t,\mathbf{x},\mathbf{x}} = h^d \mathbb{E}_n\Big[\mathbf{r}_p\Big(\frac{\mathbf{X}_i-\mathbf{x}}{h}\Big)\mathbf{r}_p\Big(\frac{\mathbf{X}_i-\mathbf{x}}{h}\Big)^\top K_h(\mathbf{X}_i-\mathbf{x})^2 \sigma_t^2(\mathbf{X}_i)\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\Big],$$

for $t \in \{0,1\}$. The following theorem gives the MSE expansions.

Theorem 4 (MSE Expansions)**.** *Suppose Assumptions 1 and 3 hold, and let $w(\mathbf{x})$ be a non-negative continuous function on $\mathscr{B}$ such that $\int_{\mathscr{B}} w(\mathbf{x})d\mathbf{x} < \infty$. If $nh^2/\log(n) \to \infty$, then*

(i) $\mathbb{E}[(\widehat{\tau}(\mathbf{x}) - \tau(\mathbf{x}))^2|\mathbf{X}] = h^{2(p+1)}\mathbf{B}_{\mathbf{x}}^2 + \frac{1}{nh^2}\mathbf{V}_{\mathbf{x}} + o_{\mathbb{P}}(r_n)$, *and*

(ii) $\int_{\mathscr{B}} \mathbb{E}[(\widehat{\tau}(\mathbf{x}) - \tau(\mathbf{x}))^2|\mathbf{X}]w(\mathbf{x})d\mathbf{x} = h^{2(p+1)} \int_{\mathscr{B}} \mathbf{B}_{\mathbf{x}}^2 dw(\mathbf{x}) + \frac{1}{nh^2}\int_{\mathscr{B}} \mathbf{V}_{\mathbf{x}} w(\mathbf{x})d\mathbf{x} + o_{\mathbb{P}}(r_n)$,

with $r_n = h^{2p+2} + n^{-1}h^{-2} + n^{-\frac{2(1+v)}{2+v}}h^{-4}$.

Ignoring the asymptotically constant and higher-order terms, the approximate MSE-optimal and IMSE-optimal bandwidth choices are

$$h_{\text{MSE},p} = \Big(\frac{2\mathbf{V}_{\mathbf{x}}}{(2p+2)\mathbf{B}_{\mathbf{x}}^2}\frac{1}{n}\Big)^{1/(2p+4)} \quad \text{and} \quad h_{\text{IMSE},p} = \Big(\frac{2\int_{\mathscr{B}} \mathbf{V}_{\mathbf{x}} w(\mathbf{x})d\mathbf{x}}{(2p+2)n\int_{\mathscr{B}} \mathbf{B}_{\mathbf{x}}^2 w(\mathbf{x})d\mathbf{x}}\frac{1}{n}\Big)^{1/(2p+4)},$$

provided that $\mathbf{B}_{\mathbf{x}} \neq 0$ and $\int_{\mathscr{B}} \mathbf{B}_{\mathbf{x}}^2 \omega(\mathbf{x})d\mathbf{x} \neq 0$, respectively.

4.2 Uncertainty Quantification

Given a bandwidth choice, the feasible t-statistic at each boundary point $\mathbf{x} \in \mathscr{B}$ is

$$\widehat{\mathsf{T}}(\mathbf{x}) = \frac{\widehat{\tau}(\mathbf{x}) - \tau(\mathbf{x})}{\sqrt{\widehat{\Omega}_{\mathbf{x},\mathbf{x}}}},$$

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

$$\widehat{\Omega}_{\mathbf{x}_1,\mathbf{x}_2} = \widehat{\Omega}_{0,\mathbf{x}_1,\mathbf{x}_2} + \widehat{\Omega}_{1,\mathbf{x}_1,\mathbf{x}_2}, \qquad \widehat{\Omega}_{t,\mathbf{x}_1,\mathbf{x}_2} = \frac{1}{nh^2} \mathbf{e}_1^\top \widehat{\mathbf{\Gamma}}_{t,\mathbf{x}_1}^{-1} \widehat{\mathbf{\Sigma}}_{t,\mathbf{x}_1,\mathbf{x}_2} \widehat{\mathbf{\Gamma}}_{t,\mathbf{x}_2}^{-1} \mathbf{e}_1$$

with

$$\widehat{\mathbf{\Sigma}}_{t,\mathbf{x}_1,\mathbf{x}_2} = h^2 \mathbb{E}_n\Big[\mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}_1}{h}\Big)\mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}_2}{h}\Big)^\top K_h(\mathbf{X}_i - \mathbf{x}_1) K_h(\mathbf{X}_i - \mathbf{x}_2) \varepsilon_i^2 \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\Big]$$

and $\varepsilon_i = Y_i - \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_0)\mathbf{e}_1^\top \widehat{\boldsymbol{\beta}}_0(\mathbf{X}_i) - \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_1)\mathbf{e}_1^\top \widehat{\boldsymbol{\beta}}_1(\mathbf{X}_i)$.

Feasible confidence intervals and confidence bands over the treatment boundary $\mathscr{B}$ take the form:

$$\widehat{I}(\mathbf{x}; \alpha) = \Big[\ \widehat{\tau}(\mathbf{x}) - \mathscr{q}_\alpha \sqrt{\widehat{\Omega}_{\mathbf{x}}}\ ,\ \widehat{\tau}(\mathbf{x}) + \mathscr{q}_\alpha \sqrt{\widehat{\Omega}_{\mathbf{x}}}\ \Big], \qquad \mathbf{x} \in \mathscr{B},$$

for any $\alpha \in [0, 1]$, and where $\mathscr{q}_\alpha$ denotes the appropriate quantile depending on the targeted inference procedure. For pointwise inference, as in the previous section, it is a textbook exercise to show that $\sup_{t \in \mathbb{R}} |\mathbb{P}[\widehat{\mathsf{T}}(\mathbf{x}) \leq t] - \Phi(t)| \to 0$ for each $\mathbf{x} \in \mathscr{B}$, under standard regularity conditions, and provided that $\sqrt{nh^2}|\mathfrak{B}_n(\mathbf{x})| \to 0$. For uniform inference, as in the previous section, we approximate the distribution of the entire stochastic process $(\widehat{\mathsf{T}}(\mathbf{x}) : \mathbf{x} \in \mathscr{B})$ first, and then deduce an approximation for $\sup_{\mathbf{x} \in \mathscr{B}} |\widehat{\mathsf{T}}(\mathbf{x})|$. This approach enable us to construct asymptotically valid confidence bands because, as noted previously,

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

See the supplemental appendix for omitted technical details.

Theorem 5 (Inferences)**.** *Suppose Assumptions 1 and 3 hold. Let* $\mathbf{W} = ((Y_1, \mathbf{X}_1), \ldots, (Y_n, \mathbf{X}_n))$.

(i) *For all* $\mathbf{x} \in \mathscr{B}$*, if* $n^{\frac{v}{2+v}} h^2 \to \infty$ *and* $\sqrt{nh^2} h^{p+1} \to 0$*, then*

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

for $\Phi^{-1}(1-\alpha/2)$*.*

(ii) *If* $n^{\frac{v}{2+v}} h^2/\log n \to \infty$ *and* $\sqrt{nh^2} h^{p+1} \to 0$*, and* $\texttt{perim}(\{\mathbf{y} \in \mathscr{A}_t : (\mathbf{y}-\mathbf{x})/h \in \mathrm{Supp}(K)\}) \lesssim h$ *for all* $\mathbf{x} \in \mathscr{B}$ *and* $t \in \{0,1\}$*, then*

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

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

This theorem gives valid pointwise and inference procedures for the boundary treatment effect $\tau(\mathbf{x})$ based on the bivariate local polynomial estimator. Unlike the case of distance-based estimation, the bias condition is the same for $\mathbf{x} \in \mathscr{B}$, making implementation substantially easier. However, as in the case of Theorem 2, and additional technical restriction on $\mathscr{B}$ is needed in order to avoid overly "wiggly" boundary designs that would lead to invalid statistical inference.

4.3 Implementation

The bivariate local polynomial estimator $\widehat{\tau}(\mathbf{x})$ and associated t-statistic $\widehat{\mathsf{T}}(\mathbf{x})$ is fully adaptive to kinks and other irregularities of the boundary $\mathscr{B}$. Therefore, it is straightforward to implement local and global bandwidth selectors based on Theorem 4 and the discussion given above. In particular, replacing the (asymptotic) bias and variance constants, $\mathbf{B}_\mathbf{x}$ and $\mathbf{V}_\mathbf{x}$, with preliminary estimators, we obtain the feasible plug-in bandwidth selectors

$$\widehat{h}_{\mathrm{MSE},p} = \Big(\frac{2\widehat{\mathbf{V}}_\mathbf{x}\mathbf{x}}{(2p+2)\widehat{\mathbf{B}}_\mathbf{x}^2}\frac{1}{n}\Big)^{1/(2p+4)} \quad \text{and} \quad \widehat{h}_{\mathrm{IMSE},p} = \Big(\frac{2\int_\mathscr{B}\widehat{\mathbf{V}}_\mathbf{x} w(\mathbf{x})d\mathbf{x}}{(2p+2)\int_\mathscr{B}\widehat{\mathbf{B}}_\mathbf{x}^2 w(\mathbf{x})d\mathbf{x}}\frac{1}{n}\Big)^{1/(2p+4)},$$

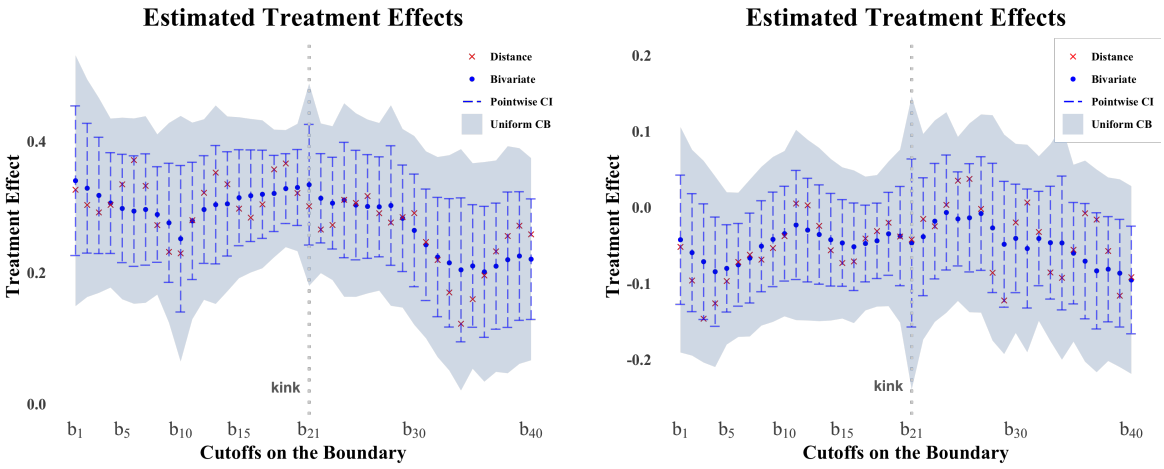
where $\widehat{\mathbf{B}}_\mathbf{x}$ is an appropriate estimator of $\mathbf{B}_\mathbf{x}$, and $\widehat{\mathbf{V}}_\mathbf{x} = \widehat{\mathbf{\Sigma}}_{\mathbf{x},\mathbf{x}}$. Omitted details are given in the supplemental appendix. These bandwidth choices can now be used to implement (I)MSE-optimal $\widehat{\tau}(\mathbf{x})$

point treatment effect estimators pointwise and uniformly along the boundary $\mathscr{B}$. Furthermore, leveraging the results in Theorem 5, a simple application of robust bias-corrected inference proceeds by employing the same (I)MSE-optimal bandwidth (for pth order point estimation), but then constructing the t-statistic $\widehat{\mathsf{T}}(\mathbf{x})$ with a choice of $p+1$ (instead of p). This inference approach has several theoretical advantages (Calonico et al., 2014, 2018, 2022), and has been validated empirically (Hyytinen et al., 2018; De Magalhães et al., 2025). Our companion software, `rd2d`, implements the procedures described above.

5 Empirical Applications

We illustrate our proposed methodology with the SPP application. We consider both distance-based and bivariate estimation and inference. The data set has $n = 363,096$ complete observations for the first cohort of the program (2014). Each observation corresponds to one student, and the bivariate score $\mathbf{X}_i = (X_{1i}, X_{2i})^\top = (\texttt{SABER11}_i, \texttt{SISBEN}_i)^\top$ is composed by the student's SABER11 test score (ranging from 1 to 100) and SISBEN wealth index (ranging from -43.84 to 56.23). **TO DO: DOUBLE CHECK RANGE IS CORRECT**. As discussed in the introduction, and without loss of generality, the score is recentered at the corresponding cutoff for program eligibility, so that the treatment assignment boundary is $\mathscr{B} = \{(\texttt{SABER11}, \texttt{SISBEN}) : (\texttt{SABER11}, \texttt{SISBEN}) \in \{\texttt{SABER11} \geq 0 \text{ and } \texttt{SISBEN} = 0\} \cup \{\texttt{SABER11} = 0 \text{ and } \texttt{SISBEN} \geq 0\}\}$, as shown in Figure 1a. Furthermore, also without loss of generality, each dimension of $\mathbf{X}_i$ is standardized in order to more naturally accommodate a common bandwidth h. The main outcome of interest is an indicator variable equal to one if the student enrolled in a higher education institution after receiving the SPP subsidy, or equal to zero otherwise.

Recall that Figure 1a in the introduction plotted the bivariate score (centered, but without standardization), as well as the assignment boundary determining treatment assignment. Figure 3a presents the empirical results using our proposed methods, both pointwise across the 40 boundary points depicted in 1a and uniformly over $\mathscr{B}$. The fidings are highlight statistically significant, indicating some degree of heterogeneity in treatment effects across the boundary. On the one hand, as students become less wealthy (going from $\mathbf{b}_1$ to $\mathbf{b}_{21}$), the treatment effects appear to be fairly stable. On the other hand, as students exhibit higher academic performance (going from $\mathbf{b}_{21}$ to



(a) Outcome: College enrollment

(b) Outcome: Mother's education (placebo)

Figure 3: Boundary Treatment Effect Estimation and Inference (SPP application).

$\mathbf{b}_{40}$), treatment effects appear to decrease.

To demonstrate the credibility of the boundary discontinuity design, we repeat the empirical analysis using pre-intervention covariates. This corresponds to a standard "placebo" analysis, where the treatment effect is expected to be statistically indistinguishable from zero. Figure 3b shows that this is indeed the case when Y_i is taken to be mother's education of the student; see the supplemental appendix for other examples.

6 Conclusion

We studied the pointwise and uniform statistical properties of the two most popular local polynomial methods for treatment effect estimation and inference in boundary discontinuity designs. Our theoretical and numerical results demonstrated that methods based on the univariate distance to the assignment boundary can exhibit a large bias in the presence of kinks or other irregularities of $\mathscr{B}$. In contrast, methods based on the bivariate score do not suffer from those problems, and thus can perform better in applications. We thus recommend employing bivariate local polynomial analysis whenever possible in boundary discontinuity designs. Our companion software package `rd2d` offers general-purpose implementations of all the estimation and inference methods developed in this paper, and is available at https://rdpackages.github.io/rd2d.

A Distance-based Minimax Convergence Rate

The following theorem presents a minimax result for estimation of bivariate nonparametric functions employing a class of estimators based on distance to the boundary of the support of the independent variables.

TO DO: DOUBLE CHECK EVERYTHING HERE

Theorem 6 (Distance-based Minimax Convergence Rate)**.** *For $p \in \mathbb{N}$, $L > 0$, let $\mathscr{P}(p; L)$ be the collection of probability laws $\mathscr{P}$ of $(Y_i, \mathbf{X}_i : 1 \leq i \leq n)$ i.i.d taking values in $\mathbb{R} \times \mathbb{R}^2$, and satisfying the following:*

(i) *$\mathbf{X}_i$ admits a Lebesgue density that is continuous and bounded from below on its support $\mathscr{X}$, which is a compact set in $\mathbb{R}^2$ and $\mathscr{B} = \mathtt{bd}(\mathscr{X})$ is a rectifiable curve.*

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

$$\max_{0\leq|\boldsymbol{\nu}|\leq p} \sup_{\mathbf{x}\in\mathscr{X}} |\partial_{\boldsymbol{\nu}}\mu(\mathbf{x})| + \max_{0\leq|\boldsymbol{\nu}|\leq p} \sup_{\mathbf{x},\mathbf{y}\in\mathscr{X}} \frac{|\partial_{\boldsymbol{\nu}}\mu(\mathbf{x}) - \partial_{\boldsymbol{\nu}}\mu(\mathbf{y})|}{\|\mathbf{x}-\mathbf{y}\|} \leq L.$$

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

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

$$\inf_{T\in\mathscr{T}} \sup_{\mathbb{P}\in\mathscr{P}} \mathbb{E}_{\mathbb{P}}\Big[\sup_{\mathbf{x}\in\mathscr{B}} \big|T(\mathbf{x};\mathbf{W}) - \mu(\mathbf{x})\big|\Big] \gtrsim n^{-1/4},$$

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

Letting $\widehat{\theta}_{\mathbf{x}}(0)$ be a distance-based local polynomial estimator, in particular, it is not difficult to show that by choosing $h^* \asymp n^{-1/4}$, we have

$$\sup_{\mathbb{P}\in\mathscr{P}} \sup_{\mathbf{x}\in\mathscr{B}} |\widehat{\theta}_{\mathbf{x}}(0) - \mu(\mathbf{x})| \lesssim_{\mathbb{P}} n^{-1/4}\sqrt{\log n}.$$

TO DO: BAD NOTATION $\lesssim_{\mathbb{P}}$... We should state the result precisely... that is, sup over $\mathbb{P}$ is "outside" the $\mathbb{P}$ concentration bound, right?

TO DO: ADD DISCUSSION

References

Black, S. E. (1999), "Do Better Schools Matter? Parental Valuation of Elementary Education," *Quarterly Journal of Economics*, 114, 577–599.

Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2018), "On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference," *Journal of the American Statistical Association*, 113, 767–779.

——— (2020), "Optimal Bandwidth Choice for Robust Bias Corrected Inference in Regression Discontinuity Designs," *Econometrics Journal*, 23, 192–210.

——— (2022), "Coverage Error Optimal Confidence Intervals for Local Polynomial Regression," *Bernoulli*, forthcoming.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014), "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 82, 2295–2326.

Cattaneo, M. D., Idrobo, N., and Titiunik, R. (2020), *A Practical Introduction to Regression Discontinuity Designs: Foundations*, Cambridge University Press.

——— (2024), *A Practical Introduction to Regression Discontinuity Designs: Extensions*, Cambridge University Press.

Cattaneo, M. D., and Titiunik, R. (2022), "Regression Discontinuity Designs," *Annual Review of Economics*, 14, 821–851.

Cattaneo, M. D., and Yu, R. R. (2025), "Strong Approximations for Empirical Processes Indexed by Lipschitz Functions," *Annals of Statistics*.

De Magalhães, L., Hangartner, D., Hirvonen, S., Meriläinen, J., Ruiz, N. A., and Tukiainen, J. (2025), "When Can We Trust Regression Discontinuity Design Estimates from Close Elections? Evidence from Experimental Benchmarks," *Political Analysis*.

Dell, M. (2010), "The Persistent Effects of Peru's Mining Mita," *Econometrica*, 78, 1863–1903.

Diaz, J. D., and Zubizarreta, J. R. (2023), "Complex Discontinuity Designs Using Covariates for Policy Impact Evaluation," *Annals of Applied Statistics*, 17, 67–88.

Federer, H. (2014), *Geometric measure theory*, Springer.

Galiani, S., McEwan, P. J., and Quistorff, B. (2017), "External and Internal Validity of a Geographic Quasi-Experiment Embedded in a Cluster-Randomized Experiment," in *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, volume 38)*, eds. M. D. Cattaneo and J. C. Escanciano, Emerald Group Publishing, pp. 195–236.

Giné, E., and Nickl, R. (2016), *Mathematical Foundations of Infinite-dimensional Statistical Models*, New York: Cambridge University Press.

Hahn, J., Todd, P., and van der Klaauw, W. (2001), "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 69, 201–209.

Hyytinen, A., Meriläinen, J., Saarimaa, T., Toivanen, O., and Tukiainen, J. (2018), "When does regression discontinuity design work? Evidence from random election outcomes," *Quantitative Economics*, 9, 1019–1051.

Jardim, E., Long, M. C., Plotnick, R., Vigdor, J., and Wiles, E. (2024), "Local minimum wage laws, boundary discontinuity methods, and policy spillovers," *Journal of Public Economics*, 234, 105131.

Keele, L., and Titiunik, R. (2016), "Natural experiments based on geography," *Political Science Research and Methods*, 4, 65–95.

Keele, L. J., Lorch, S., Passarella, M., Small, D., and Titiunik, R. (2017), "An Overview of Geographically Discontinuous Treatment Assignments with an Application to Children's Health Insurance," in *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, volume 38)*, eds. M. D. Cattaneo and J. C. Escanciano, Emerald Group Publishing, pp. 147–194.

Keele, L. J., and Titiunik, R. (2015), "Geographic Boundaries as Regression Discontinuities," *Political Analysis*, 23, 127–155.

Londoño-Vélez, J., Rodríguez, C., and Sánchez, F. (2020), "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, 193–227.

Reardon, S. F., and Robinson, J. P. (2012), "Regression Discontinuity Designs with Multiple Rating-Score Variables," *Journal of Research on Educational Effectiveness*, 5, 83–104.

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

Stone, C. J. (1982), "Optimal Global Rates of Convergence for Nonparametric Regression," *Annals of Statistics*, 10, 1040–1053.

Wellner, J. et al. (2013), *Weak convergence and empirical processes: with applications to statistics*, Springer Science & Business Media.