rd2d: Causal Inference in Boundary Discontinuity Designs

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

June 10, 2025

Abstract

Boundary discontinuity designs—also known as Multi-Score Regression Discontinuity (RD) designs, with Geographic RD designs as a prominent example—are often used in empirical research to learn about causal treatment effects along a continuous assignment boundary defined by a bivariate score. This article introduces the `R` package `rd2d`, which implements and extends the methodological results developed in Cattaneo et al. [2025] for boundary discontinuity designs. The package employs local polynomial estimation and inference using either the bivariate score or a univariate distance-to-boundary metric. It features novel data-driven bandwidth selection procedures, and offers both pointwise and uniform estimation and inference along the assignment boundary. The numerical performance of the package is demonstrated through a simulation study.

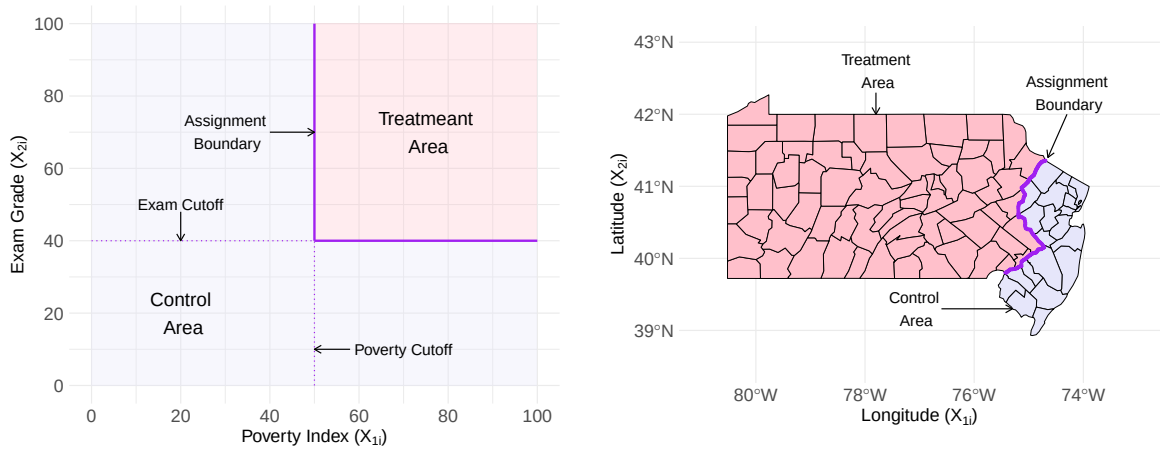
Keywords: treatment effects, regression discontinuity designs, nonparametric regression.

1 Introduction

Regression Discontinuity (RD) designs are commonly used for treatment effect estimation and causal inference in quantitative sciences [see Cattaneo and Titiunik, 2022, and references therein]. In their canonical form, each unit $i \in \{1, 2, \cdots, n\}$ is assigned to control ($T_i = 0$) or treatment ($T_i = 1$) according to the discontinuous rule $T_i = \mathbb{1}(X_i \geq c)$, where X_i denotes a scalar score variable, c denotes a scalar cutoff, and $\mathbb{1}(\cdot)$ is the indicator function. The key idea underlying all RD designs is that units with a score near the cutoff determining treatment assignment are comparable in terms of all pretreatment observables and unobservable characteristics, the only difference being that some units are assigned to control ($X_i < c$) while other are assigned to treatment ($X_i \geq c$). Therefore, in the absence of score manipulation, units having a score near (but on different sides of) the cutoff can be used as counterfactual of each other to learn about causal treatment effects.

Boundary discontinuity designs generalize the canonical RD design to allow for a multi-dimensional score variable [Papay et al., 2011, Reardon and Robinson, 2012, Keele and Titiunik, 2015]. The

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



(a) Two-Score RD Design.

(b) Geographic RD Design.

Figure 1: Two Examples of Boundary Discontinuity Designs

most common case is a bivariate score $\mathbf{X}_i = (X_{1i}, X_{2i})^\top$ with an assignment boundary curve $\mathscr{B}$ on its support. For example, Londoño-Vélez et al. [2020] study the effect of a Colombian social program where X_{1i} is a poverty index and X_{2i} is an exam grade, for each student i, and where eligibility for receiving the treatment required either a minimum poverty index ($X_{1i} \geq c_1$) or a minimum exam grade ($X_{2i} \geq c_2$). Therefore, in their application, the boundary determining treatment is $\mathscr{B} = \{(X_{1i}, X_{2i}) : \{X_{1i} \geq c_1 \text{ and } X_{2i} = c_2\} \cup \{X_{1i} = c_1 \text{ and } X_{2i} \geq c_2\}\}$. This type of bivariate RD designs, where each score component has its own cutoff, is illustrated in Figure 1a. Another prototypical class of boundary discontinuity designs are the Geographic RD designs: for instance, Keele and Titiunik [2015] study the effect of political advertisements on voter turnout during a presidential campaign by leveraging sharp discontinuities in exposure to presidential ads induced across geographic media market boundaries. Figure 1b illustrates a generic example of a Geographic RD design employing the boundary separating two US states. Jardim et al. [2024] gives another recent empirical application of a boundary discontinuity design in the context of labor markets, and provides further references. See Cattaneo et al. [2020, 2024] for a two-part practical introductory monograph.

While classical RD designs based on a scalar score are well-understood in the literature, boundary discontinuity designs are surprisingly less studied. Methodological developments have lagged empirical practice for a while, leading to different approaches in practice, but without a foundational understanding of their relative merits. Cattaneo et al. [2025] address this gap in the literature by studying the properties of two leading approaches often used in empirical research leveraging boundary discontinuity designs:

- the *location-based approach* employs bivariate local polynomial regression analysis based on the bivariate location score relative to each point on the boundary $\mathscr{B}$; and
- the *distance-based approach* employs univariate local polynomial regression analysis based on

a scalar score constructed as distance to each point on the boundary $\mathscr{B}$.

Cattaneo et al. [2025] study the two methodologies, and establish novel identification, estimation and inference results, both pointwise for each point on $\mathscr{B}$ and uniformly over $\mathscr{B}$. Importantly, they demonstrate that the distance-based approach can exhibit a large bias near kinks or other irregularities in the assignment boundary $\mathscr{B}$, while the location-based approach remains valid even in those cases. Based on their findings, it is recommended to employ the bivariate location-based approach whenever possible, but their results provide foundational theoretical guidance for both empirical approaches.

This article introduces the `R` software package `rd2d`, which expands and implements the methodological results in Cattaneo et al. [2025], thereby offering data-driven general-purpose methods for the analysis and interpretation of boundary discontinuity designs. The package includes the following four functions.

- `rd2d()`. This function implements location-based local polynomial regression analysis for estimation and inference of causal treatment effects in boundary discontinuity designs. For the n units of analysis, the function takes as inputs their outcomes $\mathbf{Y} = (Y_i : i = 1, \cdots, n)^\top$, their bivariate location scores $\mathbf{X} = [\mathbf{X}_i = (X_{1i}, X_{2i})^\top : i = 1, \cdots, n]^\top$, their treatment assignment indicators $\mathbf{T} = (T_i : i = 1, \cdots, n)^\top$, and a collection of cutoffs $\mathbf{b} = [\mathbf{b}_j = (b_{1j}, b_{2j})^\top : j = 1, \cdots, J]^\top$ on the boundary $\mathscr{B}$ determining treatment assignment. The function then implements estimation and (robust bias-corrected) inference via bivariate local polynomial regression, both pointwise and uniformly over the cutoffs $\mathbf{b}_j \in \mathscr{B}$. As it is customary in nonparametric regression settings, the function requires specifying a bandwidth (localization) parameter: if not provided by the user, it is selected via the companion function `rdbw2d()` for data-driven bandwidth selection.

- `rdbw2d()`. This function employs mean square error (MSE) approximations to implement (approximate) MSE-optimal bandwidth selection for treatment effect estimation and inference in boundary discontinuity designs. It provides second-generation direct plug-in (DPI) rules [Härdle et al., 2004, Wand and Jones, 1994], incorporating several regularization schemes.

- `rd2d.dist()`. This function implements distance-based local polynomial regression analysis for estimation and inference of causal treatment effects in boundary discontinuity designs. For the n units of analysis, the function takes as inputs their outcomes $\mathbf{Y} = (Y_i : i = 1, \cdots, n)^\top$, their scalar distance scores $\mathbf{D} = [[D_i(\mathbf{b}_j) : j = 1, \cdots, J]^\top : i = 1, \cdots, n]^\top$, and a collection of cutoffs $\mathbf{b} = [\mathbf{b}_j = (b_{1j}, b_{2j})^\top : j = 1, \cdots, J]^\top$ on the boundary $\mathscr{B}$. That is, $T_i(\mathbf{b}_j) = \mathbb{1}(D_i(\mathbf{b}_j) \geq 0)$ denotes the treatment assignment for unit i relative to cutoff $\mathbf{b}_j \in \mathscr{B}$. This function also implements estimation and (robust bias-corrected) inference via univariate local polynomial regression, both pointwise and uniformly over the cutoffs $\mathbf{b}_j \in \mathscr{B}$. If the bandwidth (localization) parameter is not provided by the user, then it is chosen via the companion function `rdbw2d.dist()` for data-driven bandwidth selection.

- `rdbw2d.dist()`. This function implements rule-of-thumb (ROT) bandwidth selection rules, depending on the specific assumptions on $\mathscr{B}$ imposed. More precisely, depending on whether the assignment boundary $\mathscr{B}$ is assumed to be smooth or not, different ROT bandwidth selectors are implemented as supported by the underlying theoretical results in Cattaneo et al. [2025]. Unfortunately, when $\mathscr{B}$ exhibits kinks or other irregularities, it is difficult to develop MSE-optimal bandwidth selection, in which case the function reverts back to a simple ROT implementation based on a rate-optimality criteria.

The methods `print()` and `summary()` are supported for objects returned by `rd2d()`, `rdbw2d()`, `rd2d.dist()`, and `rdbw2d.dist()`. We also demonstrate how to use the outputs to generate useful plots for empirical work, depicting treatment effect estimation, confidence intervals, and confidence bands, along the treatment assignment boundary $\mathscr{B}$.

In addition, the four functions in the package `rd2d` offer several practically relevant features, including (i) heteroskedasticity-robust and cluster-robust variance estimation, (ii) explicit regularization for the presence of mass points in the bivariate location score $\mathbf{X}_i$ or the univariate distance score $D_i(\mathbf{b}_j)$, and (iii) explicit regularization accounting for specific extreme shape features of the underlying unknown conditional expectation functions for bandwidth selection. In the case of `rd2d()`, the bivariate location-based approach, the data-driven point estimator is approximately MSE-optimal, while in the case of `rd2d.dist()`, the univariate distance-based approach, the rate-optimality of the data-driven point estimator depends on the underlying geometry of the assignment boundary $\mathscr{B}$. Putting aside the induced bias by the possibly non-smooth $\mathscr{B}$, both functions offer pointwise (for each $\mathbf{b}_j \in \mathscr{B}$) and uniform (over $\mathscr{B}$) robust bias-corrected inference [Calonico et al., 2018, 2022].

The main contribution of this article is to introduce and discuss the first general-purpose software implementation of (MSE-optimal) treatment effect estimation and (pointwise and uniform) uncertainty quantification methods for boundary discontinuity designs, given by the `R` package `rd2d`. To this end, the article develops second-generating DPI rules for bandwidth selection, along with principled regularization schemes for specific empirically relevant settings (e.g., mass points in $\mathbf{X}_i$) and practically relevant variance estimators (e.g., cluster-robust).

The rest of the paper proceeds as follows. Section 2 reviews the main methodological contributions in Cattaneo et al. [2025], and presents additional results related to bandwidth selection and regularized implementation. Section 3 demonstrates the performance of the `R` package `rd2d` using a simulation study with a data generated process calibrated using the data in Londoño-Vélez et al. [2020]. Section 4 concludes. Replication codes, background references, and other information related the software package `rd2d` can be found at: `https://rdpackages.github.io/rd2d/`.

2 Methods and Implementation

We employ standard potential outcomes notation. Suppose that $(Y_i(0), Y_i(1), \mathbf{X}_i^\top)^\top$, $i = 1, 2, \ldots, n$, is a random sample, where $Y_i(0)$ and $Y_i(1)$ denote the scalar potential outcomes for unit i under con-

trol and treatment assignment, respectively. Units are assigned to control group or treatment group according to their bivariate location score $\mathbf{X}_i = (X_{1i}, X_{2i})^\top$ relative to a known one-dimensional boundary $\mathscr{B}$ splitting the support of $\mathbf{X}_i$ in two disjoint regions: $\mathscr{A}_0$ denotes the control region, and $\mathscr{A}_1$ denotes the treatment region. Thus, $\mathscr{B} = \texttt{bd}(\mathscr{A}_0) \cap \texttt{bd}(\mathscr{A}_1)$, where $\texttt{bd}(\mathscr{A}_t)$ denotes the boundary of the set $\mathscr{A}_t$. The observed response variable is $Y_i = (1 - T_i) \cdot Y_i(0) + T_i \cdot Y_i(1)$, where $T_i = \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_1)$. Without loss of generality, we assume that the boundary belongs to the treatment group, that is, $\texttt{bd}(\mathscr{A}_1) \subset \mathscr{A}_1$ and $\mathscr{B} \cap \mathscr{A}_0 = \emptyset$. Figure 1 gives two graphical examples.

The causal parameter of interest is the *average treatment effect curve along the boundary*:

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

Identification follows directly by the usual continuity assumptions invoked in canonical RD designs [Hahn et al., 2001]:

$$\tau(\mathbf{x}) = \lim_{\mathbf{u} \to \mathbf{x}, \mathbf{u} \in \mathscr{A}_1} \mu_1(\mathbf{u}) - \lim_{\mathbf{u} \to \mathbf{x}, \mathbf{u} \in \mathscr{A}_0} \mu_0(\mathbf{u}), \qquad \mathbf{x} \in \mathscr{B},$$

where $\mu_t(\mathbf{x}) = \mathbb{E}[Y_i(t) | \mathbf{X}_i = \mathbf{x}] = \mathbb{E}[Y_i | \mathbf{X}_i = \mathbf{x}, T_i = t]$, $t \in \{0, 1\}$, are assumed to be smooth functions. See Papay et al. [2011], Reardon and Robinson [2012], Keele and Titiunik [2015], and references therein.

For implementation, the continuous assignment boundary $\mathscr{B}$ is first discretized into J cutoff points $\mathbf{b} = (\mathbf{b}_1, \cdots, \mathbf{b}_J)^\top$ with $\mathbf{b}_j \in \mathscr{B}$ for all $j = 1, \ldots, J$. Then, the empirical analysis is conducted pointwise for each cutoff or uniformly over all cutoffs, employing either the bivariate location score $\mathbf{X}_i$ directly, or an induced univariate distance to each cutoff point. See Cattaneo et al. [2024, Section 5] for an introductory discussion.

Following the methodological recommendations in Cattaneo et al. [2025], most of the discussion focuses on location-based methods via bivariate local polynomial regression based on the data $(Y_i, \mathbf{X}_i^\top)^\top$, $i = 1, \cdots, n$, which are implemented in the functions `rd2d()` and `rdbw2d()`. However, given their predominance in empirical work, Section 2.3 also discusses distance-based methods via univariate local polynomial regression, which require a user-chosen scalar distance score to each cutoff point on the assignment boundary, and are implemented in the functions `rd2d.dist()` and `rdbw2d.dist()`. We omit assumptions and other technical details in the remaining of the paper, which can be found in the references given.

2.1 Location-Based 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 $t \in \{0,1\}$,

$$\widehat{\boldsymbol{\beta}}_t(\mathbf{x}) = \underset{\boldsymbol{\beta}\in\mathbb{R}^{\mathfrak{p}_p+1}}{\arg\min} \sum_{i=1}^{n} \left(Y_i - \mathbf{R}_p(\mathbf{X}_i - \mathbf{x})^\top \boldsymbol{\beta}\right)^2 K_h(\mathbf{X}_i - \mathbf{x}) \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t),$$

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 $\mathbf{u} = (u_1, u_2)^\top$, $K_h(\mathbf{u}) = K(u_1/h, u_2/h)/h^2$ for a bivariate kernel function $K(\cdot)$, and a bandwidth parameter h.

In practice, it is often important to first standardize each dimension of the bivariate location score $\mathbf{X}_i = (X_{1i}, X_{2i})^\top$, and in some applications it may also be useful to allow for different bandwidths for control and treatment groups. Thus, the function `rd2d()` allows for four different bandwidths: h_{01} is used for X_{1i} and control units, h_{02} is used for X_{2i} and control units, h_{11} is used for X_{1i} and treatment units, and h_{02} is used for X_{2i} and treatment units. The data-driven bandwidth selection function `rdbw2d()` also allows for both standardization of each dimension of $\mathbf{X}_i$ (via the option `stdvar=TRUE`) and different bandwidth selection for control and treatment regions (via the options `bwselect="msetwo"` or `bwselect="imsetwo"`). Thus, the function `rdbw2d()` can report up to four distinct estimated bandwidths: $(\widehat{h}_{01}, \widehat{h}_{02}, \widehat{h}_{11}, \widehat{h}_{11})$ corresponding to $(h_{01}, h_{02}, h_{11}, h_{11})$. The discussion in this article focuses on a single common bandwidth h for simplicity, but we explain how different MSE-optimal bandwidth are estimated as appropriate. In addition, while the notation does not explicitly reflect clustering, which is a common feature in geographic and other multidimensional RD designs, the package `rd2d` allows for cluster-robust inference as explained below.

2.1.1 Point Estimation, MSE Expansions, and Bandwidth Choices

Under minimal regularity conditions, the bivariate location-based treatment effect estimator $\widehat{\tau}(\mathbf{x})$ is pointwise and uniform consistent for the treatment effect curve $\tau(\mathbf{x})$ along the assignment boundary: $\widehat{\tau}(\mathbf{x}) \to_{\mathbb{P}} \tau(\mathbf{x})$ for each $\mathbf{x} \in \mathscr{B}$, and $\sup_{\mathbf{x}\in\mathscr{B}} |\widehat{\tau}(\mathbf{x}) - \tau(\mathbf{x})| \to_{\mathbb{P}} 0$, where $\to_{\mathbb{P}}$ denotes convergence in probability as $h \to 0$ and $nh^2 \to \infty$. Furthermore, under similar regularity conditions, precise (conditional) MSE expansions can be established along $\mathscr{B}$, which can then be used for principled bandwidth selection.

The pointwise (conditional) MSE expansion is

$$\mathbb{E}[(\widehat{\tau}(\mathbf{x}) - \tau(\mathbf{x}))^2 | \mathbf{X}] \approx_{\mathbb{P}} \frac{1}{nh^2} V_{\mathbf{x}} + h^{2(p+1)} B_{\mathbf{x}}^2,$$

where $V_{\mathbf{x}} = V_{1,\mathbf{x}} + V_{0,\mathbf{x}}$ and $B_{\mathbf{x}} = B_{1,\mathbf{x}} - B_{0,\mathbf{x}}$ denote the fixed-n conditional variance and the leading conditional bias of the treatment effect estimator, respectively, and $\approx_{\mathbb{P}}$ denotes equality in probability up to vanishing higher-order terms. More precisely, using standard multi-index notation and least squares algebra, the fixed-n conditional variance and leading conditional bias for each

group $t \in \{0,1\}$ are

$$V_{t,\mathbf{x}} = \mathbf{e}_1^\top \widehat{\mathbf{\Gamma}}_{t,\mathbf{x}}^{-1} \mathbf{\Sigma}_{t,\mathbf{x},\mathbf{x}} \widehat{\mathbf{\Gamma}}_{t,\mathbf{x}}^{-1} \mathbf{e}_1 \qquad \text{and} \qquad B_{t,\mathbf{x}} = \sum_{|\mathbf{k}|=p+1} \frac{\mu_t^{(\mathbf{k})}(\mathbf{x})}{\mathbf{k}!} \mathbf{e}_1^\top \widehat{\mathbf{\Gamma}}_{t,\mathbf{x}}^{-1} \widehat{\boldsymbol{\vartheta}}_{t,\mathbf{x}}(\mathbf{k}),$$

respectively, where

$$\begin{aligned}
\widehat{\mathbf{\Gamma}}_{t,\mathbf{x}} &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{r}_p\big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\big)\mathbf{r}_p\big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\big)^\top K_h(\mathbf{X}_i - \mathbf{x}) \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t), \\
\mathbf{\Sigma}_{t,\mathbf{x},\mathbf{x}} &= \frac{h^2}{n}\sum_{i=1}^{n} \mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\Big)\mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\Big)^\top K_h(\mathbf{X}_i - \mathbf{x})^2 \varepsilon_{i,t,\mathbf{x}}^2 \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t), \\
\widehat{\boldsymbol{\vartheta}}_{t,\mathbf{x}}(\mathbf{k}) &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{r}_p\Big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\Big)\Big(\frac{\mathbf{X}_i - \mathbf{x}}{h}\Big)^{\mathbf{k}} K_h(\mathbf{X}_i - \mathbf{x}) \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t),
\end{aligned}$$

and $\varepsilon_{i,t,\mathbf{x}} = Y_i - \mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)\mu_t(\mathbf{x})$.

Therefore, for each $\mathbf{x} \in \mathscr{B}$, and noting that $\widehat{\mathbf{\Gamma}}_{t,\mathbf{x}}$, $\mathbf{\Sigma}_{t,\mathbf{x},\mathbf{x}}$, and $\widehat{\boldsymbol{\vartheta}}_{t,\mathbf{x}}(\mathbf{k})$ converge in probability (as $nh^2 \to \infty$) to well-defined limits independent on the bandwidth h, an MSE-optimal bandwidth choice is

$$h_{\mathrm{MSE},\mathbf{x}} = \Big(\frac{2V_{\mathbf{x}}}{(2p+2)B_{\mathbf{x}}^2}\frac{1}{n}\Big)^{1/(2p+4)},$$

provided that $B_{\mathbf{x}} \neq 0$.

Similarly, given a weighting function $w(\mathbf{x})$, an integrated MSE expansion is

$$\int_{\mathscr{B}} \mathbb{E}[(\widehat{\tau}(\mathbf{x}) - \tau(\mathbf{x}))^2 | \mathbf{X}] w(\mathbf{x}) d\mathbf{x} \approx_{\mathbb{P}} h^{2(p+1)} \int_{\mathscr{B}} B_{\mathbf{x}}^2 w(\mathbf{x}) d\mathbf{x} + \frac{1}{nh^2}\int_{\mathscr{B}} V_{\mathbf{x}} w(\mathbf{x}) d\mathbf{x},$$

using the notation already introduced. Therefore, along $\mathscr{B}$, an integrated MSE-optimal bandwidth choice is

$$h_{\mathrm{IMSE}} = \Big(\frac{2\int_{\mathscr{B}} V_{\mathbf{x}} w(\mathbf{x}) d\mathbf{x}}{(2p+2)\int_{\mathscr{B}} B_{\mathbf{x}}^2 w(\mathbf{x}) d\mathbf{x}}\frac{1}{n}\Big)^{1/(2p+4)},$$

provided that $\int_{\mathscr{B}} B_{\mathbf{x}}^2 w(\mathbf{x}) d\mathbf{x} \neq 0$.

The basic MSE-optimal and IMSE-optimal bandwidth choices, $h_{\mathrm{MSE},\mathbf{x}}$ and h_{IMSE}, can be extended to accommodate different selections for each coordinate of the bivariate location score and/or for control and treatment groups separately. Different bandwidths for each coordinate in $\mathbf{X}_i = (X_{1i}, X_{2i})^\top$ are obtained by first standardizing each component, then applying the basic bandwidth rules, and finally removing the standardization: letting $\tilde{\mathbf{X}}_i = (X_{1i}/\sigma_{X_1}, X_{2i}/\sigma_{X_2})^\top$, where $\sigma_{X_l}^2 = \mathbb{V}[X_{li}]$ for $l \in \{1,2\}$, then for each coordinate $l \in \{1,2\}$ the bandwidth selectors are

$$h_{l,\mathrm{MSE},\mathbf{x}} = \sigma_{X_l} \cdot h_{\mathrm{MSE},\mathbf{x}} \qquad \text{and} \qquad h_{l,\mathrm{IMSE}} = \sigma_{X_l} \cdot h_{\mathrm{IMSE}},$$

where $h_{\text{MSE},\mathbf{x}}$ and h_{IMSE} are computed using the standardized bivariate location score $\tilde{\mathbf{X}}_i$ (instead of using the original score $\mathbf{X}_i$). Different bandwidth selection for control and treatment groups are obtained by implementing $h_{\text{MSE},\mathbf{x}}$ and h_{IMSE} for each group $t \in \{0,1\}$ separately:

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

assuming the denominators are not zero.

A combination of the two ideas gives four distinct MSE-optimal and IMSE-optimal bandwidth selection rules: $(\sigma_{X_1}\cdot h_{\text{MSE},0,\mathbf{x}}, \sigma_{X_2}\cdot h_{\text{MSE},0,\mathbf{x}}, \sigma_{X_1}\cdot h_{\text{MSE},1,\mathbf{x}}, \sigma_{X_2}\cdot h_{\text{MSE},1,\mathbf{x}})$ and $(\sigma_{X_1}\cdot h_{\text{IMSE},0}, \sigma_{X_2}\cdot h_{\text{IMSE},0}, \sigma_{X_1}\cdot h_{\text{IMSE},1}, \sigma_{X_2}\cdot h_{\text{IMSE},1})$, respectively. For implementation, the package `rd2d` employs the following options:

Option		Bandwidth Selection		
`stdvar`	`bwselect`	Score Coordinates	Treatment Groups	Default
`TRUE`	`mserd`	Distinct	Same	Yes
`FALSE`	`mserd`	Same	Same	No
`TRUE`	`msetwo`	Distinct	Distinct	No
`FALSE`	`msetwo`	Same	Distinct	No

To implement the bandwidth selection procedures it is necessary to (i) estimate the unconditional variances $\sigma_{X_l}^2$ for standardization of each coordinate of $\mathbf{X}_i = (\mathbf{X}_{li} : l = 1,2)^\top$, as needed; (ii) estimate residuals $\varepsilon_{i,t,\mathbf{x}}$ entering the variances $V_{t,\mathbf{x}}$ for each group $t \in \{0,1\}$; (iii) estimate higher-order derivatives $(\mu_t^{(\mathbf{k})}(\mathbf{x}) : |\mathbf{k}| = p+1)$ entering the bias $B_{t,\mathbf{x}}$ quantities for each group $t \in \{0,1\}$; and (iv) select (preliminary) bandwidth entering the matrices $\widehat{\mathbf{\Gamma}}_{t,\mathbf{x}}$, $\mathbf{\Sigma}_{t,\mathbf{x},\mathbf{x}}$, and $\widehat{\boldsymbol{\vartheta}}_{t,\mathbf{x}}(\mathbf{k})$ for each group $t \in \{0,1\}$. (A data-driven regularization term is also added to the denominators to avoid near-zero bias, as explained below.) These unknown quantities are estimated as follows.

- The unconditional variances used for standardization are estimated using their sample variance counterparts: $\widehat{\sigma}_{X_l}^2 = \frac{1}{n-1}\sum_{i=1}^n (X_{li} - \frac{1}{n}\sum_{i=1}^n X_{li})^2$, for each coordinate $l \in \{1,2\}$.

- Given a (preliminary) bandwidth, the package `rd2d` implements several variance estimators for each group $t \in \{0,1\}$: $V_{t,\mathbf{x}}$ is replaced by $\widehat{V}_{t,\mathbf{x}} = \mathbf{e}_1^\top \widehat{\mathbf{\Gamma}}_{t,\mathbf{x}}^{-1}\widehat{\mathbf{\Sigma}}_{t,\mathbf{x},\mathbf{x}}\widehat{\mathbf{\Gamma}}_{t,\mathbf{x}}^{-1}\mathbf{e}_1$, where $\widehat{\mathbf{\Sigma}}_{t,\mathbf{x},\mathbf{x}}$ is either a heteroskedasticity-consistent (HC) or cluster-consistent (CR) variance estimator based on replacing the unknown residuals $\varepsilon_{i,t,\mathbf{x}}$ with the plug-in residuals estimates

$$\widehat{\varepsilon}_{i,t,\mathbf{x}} = Y_i - \mathbf{R}_p(\mathbf{X}_i - \mathbf{x})^\top \widehat{\boldsymbol{\beta}}_t(\mathbf{x}).$$

More precisely, the package allows for the following options:

Option		Variance Implementation
`vce`	`cluster`	HC-estimate (`cluster=NULL`) or CR-estimate (`cluster=`*varname*)
`hc0`	`NULL` or *varname*	no weighting of estimated residuals
`hc1`	`NULL` or *varname*	degrees-of-freedom weighting of estimated residuals
`hc2`	`NULL` or *varname*	inverse-diagonal-projection weighting of estimated residuals
`hc3`	`NULL` or *varname*	inverse-diagonal-projection-squared weighting of estimated residuals

See Zeileis [2004] and Zeileis et al. [2020] for more discussion.

- Given a (preliminary) bandwidth, the package `rd2d` estimate the higher-order curvature of the unknown conditional expectations, $\mu_t^{(\mathbf{k})}(\mathbf{x})$ and $\mu_1^{(\mathbf{k})}(\mathbf{x})$ for each $\mathbf{k} \in \mathbb{N}^2$ such that $|\mathbf{k}| = p+1$, using a higher-order polynomial approximation. By default, a local polynomial regression of order $q = p+1$ is used.

- The preliminary bandwidth(s) needed to construct the matrices $\widehat{\mathbf{\Gamma}}_{t,\mathbf{x}}$, $\mathbf{\Sigma}_{t,\mathbf{x},\mathbf{x}}$, and $\widehat{\boldsymbol{\vartheta}}_{t,\mathbf{x}}(\mathbf{k})$, as well as the plug-in residuals and higher-order derivative estimates, are selected using a combination of ROT and DPI-2 methods. Specifically, two bandwidth are sequentially constructed as follows:

Step 1. Using a Gaussian distribution reference model, construct a plug-in IMSE-optimal ROT bandwidth selector for the canonical kernel density estimator of $f_{\mathbf{X}}(\mathbf{x})$, the Lebesgue density of score. The resulting data-driven bandwidth choice is:

$$\widehat{c} = \widehat{C} n^{-1/6},$$

where $\widehat{C}$ is a function of the variance of $\mathbf{X}_i$ and known constants determined by the kernel function used. This preliminary bandwidth choice is motivated by the fact that $\widehat{\mathbf{\Gamma}}_{t,\mathbf{x}} \to_{\mathbb{P}} f_{\mathbf{X}}(\mathbf{x})\mathbf{\Gamma}_{t,\mathbf{x}}$ and $\widehat{\boldsymbol{\vartheta}}_{t,\mathbf{x}}(\mathbf{k}) \to_{\mathbb{P}} f_{\mathbf{X}}(\mathbf{x})\boldsymbol{\vartheta}_{t,\mathbf{x}}(\mathbf{k})$, where $\mathbf{\Gamma}_{t,\mathbf{x}}$ and $\boldsymbol{\vartheta}_{t,\mathbf{x}}(\mathbf{k})$ are non-random matrices, only function of $\mathscr{B}$, K, p, $|\mathbf{k}| = p+1$, and $t \in \{0,1\}$.

Step 2. Construct an MSE-optimal bandwidth choice for estimating the linear combination given by $\sum_{|\mathbf{k}|=p+1} \frac{\mu_t^{(\mathbf{k})}(\mathbf{x})}{\mathbf{k}!} \mathbf{e}_1^\top \widehat{\mathbf{\Gamma}}_{t,\mathbf{x}}^{-1} \widehat{\boldsymbol{\vartheta}}_{t,\mathbf{x}}(\mathbf{k})$, using a qth order local polynomial estimator and the bandwidth $\widehat{c}$ for the coefficients $\mathbf{e}_1^\top \widehat{\mathbf{\Gamma}}_{t,\mathbf{x}}^{-1} \widehat{\boldsymbol{\vartheta}}_{t,\mathbf{x}}(\mathbf{k})$. To implement this bandwidth choice, $\widehat{c}$ is used for variance estimation, and a preliminary nearest-neighbor-based polynomial regression approximation is used for bias estimation. The resulting data-driven bandwidth choice is:

$$\widehat{b}_t = \widehat{C}_t n^{-1/(2q+4)}, \qquad t \in \{0,1\},$$

where $\widehat{C}_t$ depends on the variance and bias estimates for the target linear combination, and K and q.

Step 3. Construct the final variance and bias constants using the preliminary bandwidth estimates $(\widehat{c}, \widehat{b}_0, \widehat{b}_1)$. Specifically, the basic MSE-optimal (or IMSE-optimal) bandwidth choice is implemented as follows: for each $t \in \{0, 1\}$,

- $V_{t,\mathbf{x}}$ is replaced by $\widehat{V}_{t,\mathbf{x}}$, where $\widehat{c}$ is used to construct the matrices, and a pth order local polynomial regression is used for residual estimation; and
- $B_{t,\mathbf{x}}$ is replaced by $\widehat{B}_{t,\mathbf{x}}$, where $\widehat{c}$ is used to construct the matrices, and $\widehat{b}_t$ is used to estimate the derivatives of the regression function for each group.

See Wand and Jones [1994] and Härdle et al. [2004] for technical details.

2.1.2 Statistical Inference

To assess uncertainty in the estimation of the causal effects $\tau(\mathbf{x})$ along the boundary $\mathscr{B}$, we consider the usual Wald-type test statistic

$$\mathrm{T}(\mathbf{x}) = \frac{\widehat{\tau}(\mathbf{x}) - \tau(\mathbf{x})}{\sqrt{\widehat{V}_{\mathbf{x}}/(nh^2)}}, \qquad \mathbf{x} \in \mathscr{B},$$

where $\widehat{\tau}(\mathbf{x})$ and $\widehat{V}_{\mathbf{x}}$ are constructed using the polynomial order p and bandwidth h, after choosing the appropriate HC and CR variance estimator, as explained above in the context of bandwidth selection.

In practice, the bandwidth is chosen to be (I)MSE-optimal, and thus the sampling distribution of the statistic satisfies the following pointwise distributional approximation:

$$\mathrm{T}(\mathbf{x}) - \mathrm{Bias}(\mathbf{x}) \overset{a}{\sim} \mathsf{N}(0,1), \qquad \mathrm{Bias}(\mathbf{x}) = \frac{h^{p+1} B_{\mathbf{x}}}{\sqrt{\widehat{V}_{\mathbf{x}}/(nh^2)}}, \qquad \mathbf{x} \in \mathscr{B},$$

where $\overset{a}{\sim}$ denote an approximation in distribution as $nh^2 \to \infty$, $\mathsf{N}(0,1)$ denotes the standard Gaussian distribution, and $\mathrm{Bias}(\mathbf{x})$ denotes the standardized leading bias emerging whenever a "large" bandwidth is used (i.e., when the (I)MSE-optimal bandwidth is used, or any other bandwidth choice such that $\mathrm{Bias}(\mathbf{x}) \not\to_{\mathbb{P}} 0$). The standard confidence interval estimator with $100(1-\alpha)\%$ nominal coverage is

$$\mathrm{CI}(\mathbf{x}) = \left[\widehat{\tau}(\mathbf{x}) \pm \Phi_{1-\alpha/2}\sqrt{\widehat{V}_{\mathbf{x}}/(nh^2)}\right], \qquad \mathbf{x} \in \mathscr{B},$$

where Φ_α is the αth quantile of the standard Gaussian distribution. However, for "large" bandwidths such as the (I)MSE-optimal choice, $\mathrm{CI}(\mathbf{x})$ will be invalid due to the bias $\mathrm{Bias}(\mathbf{x})$, thereby delivering empirical coverage well below its nominal target. A solution to this problem is to employ ad-hoc undersmoothing, that is, to implement $\mathrm{CI}(\mathbf{x})$ with a "smaller" bandwidth h relative to the (I)MSE-optimal choice. Calonico et al. [2018, 2022] showed that undersmoothing is sub-optimal (possibly invalid) under standard assumptions, while the robust bias-correction (RBC) methodology

introduced by Calonico et al. [2014] enjoys validity and better (in some cases optimal) higher-order distributional properties. The core idea behind the RBC approach can be summarized as follows: (i) employ the (I)MSE-optimal bandwidth for constructing the point estimator $\widehat{\tau}(\mathbf{x})$, (ii) de-bias (bias correct) the numerator of the statistic $\mathrm{T}(\mathbf{x})$, and (iii) adjust the variance estimate to account for the variability introduced by the debiasing of $\widehat{\tau}(\mathbf{x})$.

The implementation of the RBC inference methodology is straightforward: given a chosen (I)MSE-optimal bandwidth for the pth order local polynomial point estimator $\widehat{\tau}(\mathbf{x})$, an adjsted test statistic is constructed using a qth order local polynomial point estimator $\widehat{\tau}_q(\mathbf{x})$ and its associated variance estimate $\widehat{V}_{\mathbf{x},q}$, with $q > p$. Thus, the approach employs the RBC statistic $T_q(\mathbf{x}) = (\widehat{\tau}_q(\mathbf{x}) - \tau(\mathbf{x}))/\sqrt{\widehat{V}_{\mathbf{x},q}/(nh^2)}$, instead of the original statistic $\mathrm{T}(\mathbf{x})$ above. The resulting RBC confidence interval estimator is

$$\mathrm{CI}_q(\mathbf{x}) = \left[\widehat{\tau}_q(\mathbf{x}) \pm \Phi_{1-\alpha/2}\sqrt{\widehat{V}_{\mathbf{x},q}/(nh^2)}\right], \qquad \mathbf{x} \in \mathscr{B},$$

which is constructed with an (I)MSE-optimal bandwidth choice for the pth order local polynomial point estimator $\widehat{\tau}(\mathbf{x})$. (This amounts to a form of robust bias correction because $\widehat{\tau}_q(\mathbf{x}) = \widehat{\tau}(\mathbf{x}) - h^{p+1}\widehat{B}_{\mathbf{x}}$, where $\widehat{B}(\mathbf{x})$ is an "estimate" of $B_{\mathbf{x}}$.) The package `rd2d` employs $p = 1$ and $q = 2$ as defaults, which are standard choices for implementation.

Uniform inference and confidence bands along the boundary $\mathscr{B}$ also employ RBC methodology. Specifically, given an (I)MSE-optimal bandwidth for the pth order local polynomial point estimator $\widehat{\tau}(\mathbf{x})$, the associated RBC confidence band estimate is

$$\mathrm{CB}_q(\mathscr{B}) = \left\{\left[\widehat{\tau}_q(\mathbf{x}) \pm \mathsf{q}_\alpha\sqrt{\widehat{V}_{\mathbf{x},q}/(nh^2)}\right] : \mathbf{x} \in \mathscr{B}\right\},$$

where q_α denotes a suitably chosen quantile to control false rejections uniformly over $\mathscr{B}$. In practice, the continuous assignment boundary is discretized to consider the J cutoff in $\mathbf{b} = (\mathbf{b}_1, \cdots, \mathbf{b}_J)^\top$ jointly. Then, a feasible quantile choice is

$$\mathsf{q}_\alpha = \inf\left\{u \geq 0 : \ \mathbb{P}\Big[\max_{1\leq j\leq J}\big|\mathbf{e}_j^\top \widehat{\mathbf{C}}_{\mathbf{b},q}^{1/2}\mathbf{Z}\big| \geq u\Big|\mathrm{Data}\Big] \leq \alpha\right\},$$

where the J-dimensional standard Gaussian vector $\mathbf{Z} \sim \mathsf{N}(\mathbf{0}_J, \mathbf{I}_J)$ is independent of the data, and the $(J \times J)$ covariance matrix $\widehat{\mathbf{C}}_{\mathbf{b},q} = \big[\widehat{V}_{\mathbf{b}_j,\mathbf{b}_k}/(\widehat{V}_{\mathbf{b}_j,\mathbf{b}_j}\widehat{V}_{\mathbf{b}_k,\mathbf{b}_k})^{1/2} : 1 \leq j,k \leq J\big]$ where $\widehat{V}_{\mathbf{b}_j,\mathbf{b}_k} = \widehat{V}_{1,\mathbf{b}_j,\mathbf{b}_k} + \widehat{V}_{0,\mathbf{b}_j,\mathbf{b}_k}$ with

$$V_{t,\mathbf{b}_j,\mathbf{b}_k} = \mathbf{e}_1^\top \widehat{\mathbf{\Gamma}}_{t,\mathbf{b}_j}^{-1}\widehat{\mathbf{\Sigma}}_{t,\mathbf{b}_j,\mathbf{b}_k}\widehat{\mathbf{\Gamma}}_{t,\mathbf{b}_k}^{-1}\mathbf{e}_1$$

and

$$\widehat{\mathbf{\Sigma}}_{t,\mathbf{b}_j,\mathbf{b}_k} = \frac{h^2}{n}\sum_{i=1}^{n}\mathbf{r}_q\Big(\frac{\mathbf{X}_i - \mathbf{b}_j}{h}\Big)\mathbf{r}_q\Big(\frac{\mathbf{X}_i - \mathbf{b}_k}{h}\Big)^\top K_h(\mathbf{X}_i - \mathbf{b}_j)K_h(\mathbf{X}_i - \mathbf{b}_k)\widehat{\varepsilon}_{i,\mathbf{b}_j}\widehat{\varepsilon}_{i,\mathbf{b}_k}\mathbb{1}(\mathbf{X}_i \in \mathscr{A}_t)$$

with $\widehat{\varepsilon}_{i,t,\mathbf{x}} = Y_i - \mathbf{R}_q(\mathbf{X}_i - \mathbf{x})^\top \widehat{\boldsymbol{\beta}}_t(\mathbf{x})$, for $t \in \{0, 1\}$. This formulation is robust to unknown conditional heteroskedasticity, while the clustered-robust analogue formula is omitted to save space; see Zeileis [2004] and Zeileis et al. [2020].

The RBC method produces confidence intervals/bands that are not centered at the treatment effect point estimator because different polynomial orders are used for estimation and inference. As a result, the point estimates may lie outside the RBC confidence intervals/bands, particularly if the underlying treatment effect curve $\tau(\mathbf{x})$ exhibits high curvature at certain evaluation points. One possible solution is to increase the polynomial orders p and q, or to use a bandwidth smaller than the (I)MSE-optimal one.

2.1.3 Regularization Strategies

The package `rd2d` implements several regualization schemes to ensure robustness in applications.

- *Small bias regularization.* Ignoring the asymptotically constant and higher-order terms, the approximate MSE-optimal and IMSE-optimal bandwidth choices require $B_{\mathbf{x}} \neq 0$ and $\int_{\mathscr{B}} B_{\mathbf{x}}^2 \omega(\mathbf{x}) d\mathbf{x} \neq 0$, respectively. Thus, a small estimated bias can result in a bandwidth that is too large. To avoid this problem, a regularization term is added to the term of estimated bias, leading to the regularized MSE-optimal bandwidth choice,

$$h_{\mathrm{MSE},\mathbf{x}} = \Big(\frac{2\widehat{V}_{\mathbf{x}}}{(2p+2)(\widehat{B}_{\mathbf{x}}^2 + s \cdot \mathbb{V}[\widehat{B}_{\mathbf{x}}])} \frac{1}{n} \Big)^{1/(2p+4)}, \qquad \mathbf{x} \in \mathscr{B},$$

  and the regularized IMSE-optimal bandwidth choice,

$$h_{\mathrm{IMSE}} = \Big(\frac{2 \int_{\mathscr{B}} \widehat{V}_{\mathbf{x}} d\mathbf{x}}{(2p+2)(\int_{\mathscr{B}} \widehat{B}_{\mathbf{x}}^2 + s \cdot \mathbb{V}[\widehat{B}_{\mathbf{x}}] d\mathbf{x})} \frac{1}{n} \Big)^{1/(2p+4)}.$$

  where the regularization terms account for variance of the bias estimator, and are estimated as discussed previously. The factor s, defaulted to 3, controls the degree of regularization

- *Minimum sample size.* A sample size of at least `bwcheck` is required by (possibly) enlarging the selected/provided bandwidth until `bwcheck` number of observations are included in the estimation region. The default is `bwcheck` $= 50 + (2+p)(1+p)/2 - 1$. When kernel type is `"prod"`, a smallest rectangle centered at the evaluation point with two edges proportional to $(\sigma(X_{i1}), \sigma(X_{i2}))$, σ stands for the standard deviation, is found, and the bandwidth for local polynomial fitting should allow the smallest rectangle to be contained in its resulting kernel. When kernel type is `"rad"`, a smallest ball centered at evaluation point with `bwcheck` number of data points is found, and the bandwidth is increased until its resulting kernel contains the smallest ball.

- *Mass points in* $\mathbf{X}_i$. The `masspoint` option checks for unique number of points in the data. The default is `masspoint = "check"`, where unique number of data points is reported, and

a warning is issued if duplication exceeds 20% of the data. When `masspoint = "adjust"`, bandwidths are regularized so that the resulting kernels contain a minimal number of unique observations (see *minimum sample size*). When `masspoint = "off"`, the potential presence of mass points is ignored.

2.2 Aggregated Average Treatment Effects Along the Boundary

For a weight function $w : \mathscr{B} \to [0, \infty)$, the *aggregated average treatment effect* (AATE) along the boundary $\mathscr{B}$ is

$$\tau_{\text{AATE},\mathscr{B}} = \frac{\int_{\mathscr{B}} \tau(\mathbf{b}) w(\mathbf{b})\, d\mathbf{b}}{\int_{\mathscr{B}} w(\mathbf{b})\, d\mathbf{b}}$$

A choice of weight function recovers a scalar causal effect, aggregating potential heterogeneous treatment effects along the boundary. Estimation and inference methods for this class of causal parameters can be deduced from our results. For example, consider the generic plug-in "estimator" of $\tau_{\text{AATE},\mathscr{B}}$ given by

$$\widehat{\tau}_{\text{AATE},\mathbf{b}} = \frac{\sum_{j=1}^{J} \widehat{\tau}(\mathbf{b}_j) w(\mathbf{b}_j)}{\sum_{j=1}^{J} w(\mathbf{b}_j)},$$

where the cutoff points $\mathbf{b} = (\mathbf{b}_1, \ldots, \mathbf{b}_J)^\top$ along the assignment boundary are assumed to be evenly-spaced over $\mathscr{B}$, and $\widehat{\tau}(\mathbf{b}_j)$ is constructed using pth order local polynomial fits. It follows that the IMSE-optimal bandwidth choice is more natural. Feasible RBC confidence intervals are

$$\text{CI}_q(\text{AATE}) = \left[\widehat{\tau}_{\text{AATE},\mathbf{b},q} \pm \Phi_{1-\alpha/2} \sqrt{\frac{\mathbf{w}^\top \widehat{\mathbf{V}}_{\mathbf{b},q} \mathbf{w}}{nh^2}} \right],$$

where $\widehat{\tau}_{\text{AATE},\mathbf{b},q}$ is the point estimator $\widehat{\tau}_{\text{AATE},\mathbf{b}}$ constructed using a qth order polynomial basis, $\widehat{\mathbf{V}}_{\mathbf{b},q} = \left[\widehat{V}_{\mathbf{b}_j,\mathbf{b}_k} : 1 \le j,k \le J\right]$ denotes its associated variance estimator, as explained previously, and $\mathbf{w} = (w(\mathbf{b}_1), \ldots, w(\mathbf{b}_J))^\top / (\sum_{j=1}^{J} w(\mathbf{b}_j))$ is the user-chosen vector of weights. In `rd2d`, AATE estimation and inference are implemented via the optional argument `AATE = w` in the `summary()` method for `rd2d` objects.

2.3 Distance-Based Methods

For each unit $i = 1, \ldots, n$, their scalar distance-based score to the boundary point $\mathbf{x} = (x_1, x_2)^\top \in \mathscr{B}$ is $D_i(\mathbf{x}) = (2T_i - 1)\mathscr{d}(\mathbf{X}_i, \mathbf{x})$, where $\mathscr{d}(\cdot,\cdot)$ denotes a distance function such as the Euclidean distance $\mathscr{d}(\mathbf{X}_i, \mathbf{x}) = \|\mathbf{X}_i - \mathbf{x}\| = \sqrt{(X_{1i} - x_1)^2 + (X_{2i} - x_2)^2}$. Therefore, for each $\mathbf{x} \in \mathscr{B}$, the setup reduces to a standard univariate RD design with distance score $D_i(\mathbf{x}) \in \mathbb{R}$ and cutoff $c = 0$, the observed data now being $(Y_1, D_1(\mathbf{x})), \cdots, (Y_n, D_n(\mathbf{x}))$ for each point on the assignment boundary $\mathbf{b} = (\mathbf{b}_1, \cdots, \mathbf{b}_J)^\top$.

The distance-based local polynomial treatment effect curve estimator of $\tau(\mathbf{x})$ 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 $t \in \{0,1\}$,

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

with $\mathbf{r}_p(u) = (1, u, u^2, \cdots, u^p)^\top$ the usual univariate polynomial basis, $k_h(u) = k(u/h)/h^2$ for univariate kernel function $k(\cdot)$ and bandwidth parameter h, and $\mathscr{I}_0 = (-\infty, 0)$ and $\mathscr{I}_1 = [0, \infty)$. Cattaneo et al. [2025] studied the statistical properties of the distance-based approach in boundary discontinuity designs, and obtained the following main results (under regularity conditions).

1. *Consistency.* As $h \to 0$ and $nh^2 \to \infty$, $\widehat{\tau}_{\text{dis}}(\mathbf{x}) \to_{\mathbb{P}} \tau_{\text{dis}}(\mathbf{x})$ for all $\mathbf{x} \in \mathscr{B}$, and $\sup_{\mathbf{x}\in\mathscr{B}} |\widehat{\tau}_{\text{dis}}(\mathbf{x}) - \tau_{\text{dis}}(\mathbf{x})| \to_{\mathbb{P}} 0$, where $\tau_{\text{dis}}(\mathbf{x}) = \lim_{r\downarrow 0} \theta_{1,\mathbf{x}}(r) - \lim_{r\uparrow 0} \theta_{0,\mathbf{x}}(r)$ with

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

   for $\mathbf{x} \in \mathscr{B}$ and $t \in \{0,1\}$. The functions $\theta_{t,\mathbf{x}}(r)$ are the univariate induced conditional expectations based on distance to the boundary point $\mathbf{x} \in \mathscr{B}$ for each group $t \in \{0,1\}$.

2. *Identification.* $\tau(\mathbf{x}) = \tau_{\text{dis}}(\mathbf{x})$ for all $\mathbf{x} \in \mathscr{B}$, thereby showing that the distance-based estimator is a valid treatment effect estimator.

3. *Bias.* If the assignment boundary $\mathscr{B}$ is non-smooth, then the uniform bias of the estimator $\widehat{\tau}_{\text{dis}}(\mathbf{x})$ along the boundary is no better than of order h, regardless of the polynomial order p used. In other words, the distance-based estimator exhibits a "large" bias near kinks or other irregularities of the assignment boundary $\mathscr{B}$. On the other hand, if the assignment boundary $\mathscr{B}$ is smooth enough, then the bias of $\widehat{\tau}_{\text{dis}}(\mathbf{x})$ is of order h^{p+1} as expected in local polynomial regression settings.

4. *Mean Square Convergence and Bandwidth Choice.* Due to unknown form of distance function $\mathscr{d}(\cdot,\cdot)$ and the assignment boundary $\mathscr{B}$, it is not possible to obtain valid (I)MSE expansions and precise bandwidth selection rules. At this level of generality, only bandwidth selection in terms of rates can be established:

   - If $\mathscr{B}$ is smooth, then $h \asymp n^{-1/(2p+4)}$ is (I)MSE rate-optimal, where $\asymp$ denotes up to a proportional constant.
   - If $\mathscr{B}$ is non-smooth, then $h \asymp n^{-1/4}$ is (I)MSE rate-optimal, regardless of the polynomial order p used in constructing $\widehat{\tau}_{\text{dis}}(\mathbf{x})$.

5. *Statistical Inference.* Putting aside the issue of "large" bias whenever $\mathscr{B}$ is non-smooth, valid confidence intervals/bans can be developed based on the distance-based estimator. The same

inference results outlined for the location-based approach are available for the distance based approach, with some important caveats:

- If $\mathscr{B}$ is smooth, then RBC inference is possible. Thus, first the (I)MSE-rate-optimal bandwidth $h \asymp n^{-1/(2p+4)}$ is used for point estimation (i.e., $\widehat{\tau}_{\text{dis}}(\mathbf{x})$) using pth order local polynomial regression, and then inference proceeds using qth order local polynomial regression. This is implemented using the option `kink = "off"`, and is the default for `rd2d.dist()`.
- If $\mathscr{B}$ is non-smooth, then the RBC inference is not possible because the leading bias is unknown and increasing the polynomial order does not reduce bias. In this case, point estimation employs the (I)MSE-rate-optimal bandwidth $h \asymp n^{-1/4}$, and then inference employs the undersmoothed bandwidth choice $h \asymp n^{-1/3}$ following the results in Calonico et al. [2018, 2022]. As a result, point estimation and inference employ the same polynomial order ($p = q$).

Other implementation and regularization methods follow the same logic as for bivariate location-based estimation, taking into account the distance variable explicitly. In particular, `bwcheck` $= 50 + p + 1$ is used as default. See Cattaneo et al. [2025] for omitted technical and methodological details.

3 Numerical Illustrations

We illustrate the capabilities of the general-purpose `R` software package `rd2d` with a synthetic dataset of size $n = 20,000$ calibrated using the *Ser Pilo Paga* (SPP) dataset [Londoño-Vélez et al., 2020]. We set $T_i = \mathbb{1}(X_{1i} \geq 0, X_{2i} \geq 0)$, where $T_i = 1$ indicates unit i is in the treatment group, and $T_i = 0$ indicates unit i is in the control group. Covariates $\mathbf{X}_i = (X_{1i}, X_{2i})^\top$ are drawn from the product distribution $(100\mathsf{Beta}(3,4) - 25, \mathsf{B}_2 = 100\mathsf{Beta}(3,4) - 25)$ with independent components. Potential outcomes are generated by

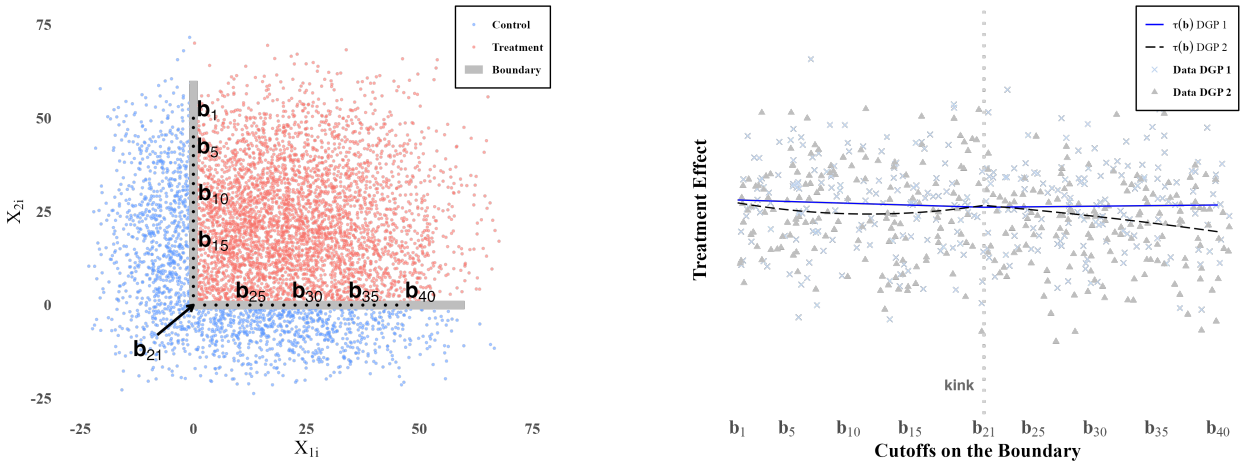
$$Y_i(t) = \beta_{t,0} + X_{1i}\beta_{t,11} + X_{2i}\beta_{t,12} + X_{1i}^2\beta_{t,21} + X_{2i}^2\beta_{t,22} + X_{1i}X_{2i}\beta_{t,23} + \varepsilon_{t,i},$$

where $\mathbf{X}_i$, $\varepsilon_{0,i}$ and $\varepsilon_{1,i}$ are mutually independent, and $\varepsilon_{t,i} \sim \mathsf{N}(0, \sigma_t^2)$, for $i = 1, 2, \cdots, n$ and $t = 0, 1$. We consider two DGPs as in Table 1, where coefficients are estimated from the *Ser Pilo Paga* (SPP) dataset [Londoño-Vélez et al., 2020], and scaled by a factor of 2 to enhance signal strength.

Figure 2a presents a scatterplot of a synthetic dataset constructed using the data generating process described above, which we use for numerical illustration of the main capabilities of the package `rd2d`. The gray assignment boundary separates the control (blue) and treatment (red) groups. The plot also includes forty grid points along the boundary, where the cutoff $\mathbf{b}_{21}$ is a kink point. Figure 2b presents the population treatment effects corresponding to the two data generating processes. To demonstrate the variability in the outcome variable, the figure includes synthetic data points derived from 300 uniform draws along the boundary.

	DGP 1 (Linear)		DGP 2 (Quadratic)	
	$t=0$	$t=1$	$t=0$	$t=1$
$\beta_{t,0}$	$\mathbf{2}\times 3.35\times 10^{-1}$	$\mathbf{2}\times 6.98\times 10^{-1}$	$\mathbf{2}\times 3.72\times 10^{-1}$	$\mathbf{2}\times 7.435\times 10^{-1}$
$\beta_{t,11}$	$\mathbf{2}\times 2.52\times 10^{-3}$	$\mathbf{2}\times 2.74\times 10^{-3}$	$\mathbf{2}\times 4.23\times 10^{-3}$	$\mathbf{2}\times 2.29\times 10^{-3}$
$\beta_{t,12}$	$-\mathbf{2}\times 1.72\times 10^{-3}$	$-\mathbf{2}\times 6.05\times 10^{-4}$	$-\mathbf{2}\times 2.45\times 10^{-3}$	$-\mathbf{2}\times 5.85\times 10^{-3}$
$\beta_{t,21}$	0	0	$\mathbf{2}\times 1.25\times 10^{-5}$	$-\mathbf{2}\times 1.33\times 10^{-7}$
$\beta_{t,22}$	0	0	$-\mathbf{2}\times 4.92\times 10^{-6}$	$\mathbf{2}\times 2.14\times 10^{-5}$
$\beta_{t,23}$	0	0	$\mathbf{2}\times 3.12\times 10^{-5}$	$\mathbf{2}\times 1.04\times 10^{-4}$
σ_t	3.32×10^{-1}	4.35×10^{-1}	3.31×10^{-1}	4.35×10^{-1}

Table 1: True parameter values for DGP 1 (Linear) and DGP 2 (Quadratic) under treatment statuses $t=0$ and $t=1$, fitted from the *Ser Pilo Paga* (SPP) dataset [Londoño-Vélez et al., 2020] and coefficients boosted by 2.



(a) Scatter Plot of (X_{1i}, X_{2i})

(b) Treatment Effects Along the Boundary

Figure 2: Estimation and Inference (Simulations).

3.1 Function `rd2d()`

The function `rd2d()` provides point estimation, robust confidence intervals, and robust uniform confidence bands for (derivatives of) treatment effect function $\tau(\mathbf{x})$ based on bivariate location-based local polynomial regression. It takes as input an outcome vector `y`, a bivariate location score matrix `X`, a treatment indicator vector `t`, and a grid of evaluation points `b` along the boundary $\mathscr{B}$.

Optional arguments include bandwidth choices `h`, degrees of polynomial for point estimation (`p`) and inference (`q`), the partial derivative of treatment effect to be estimated `deriv`, and confidence level `level`. When optional arguments are not provided, the function defaults to estimate the value of treatment effect `deriv = c(0,0)` using the MSE-optimal bandwidth, with `p = 1` degree polynomial for point estimation and `q = 2` degree polynomial for robust bias-corrected confidence interval (and bands if requested), and `level = 95` percentage points confidence level. Additionally, `kernel_type` indicates whether a product kernel (`"prod"`) or a radial kernel (`"rad"`) is used for

weighting. The default is `kernel_type = "prod"`.

Below is a demonstration of `rd2d()` applied on the synthetic dataset, with results for selected indices printed using the `summary()` method.

```
> result.rd2d <- rd2d(y, X, t, eval)
> summary(result.rd2d, subset = c(1,5,10,15,21,25,30,35,40))
rd2d

Number of Obs.          20000
BW type.                mserd-dpi-std
Kernel                  triangular-prod
VCE method              hc1
Masspoints              check

Number of Obs.          6191          13809
Estimand (deriv)        0             0
Order est. (p)          1             1
Order rbc. (q)          2             2
Unique Obs.             6191          13809

====================================================================
  ID        b1        b2      Est.         z     P>|z|              95% CI
====================================================================
   1     0.000    50.000    0.8143    9.8761    0.0000    [0.6191, 0.9257]
   5     0.000    40.000    0.7298   11.8383    0.0000    [0.6125, 0.8556]
  10     0.000    27.500    0.6752   11.3927    0.0000    [0.5487, 0.7767]
  15     0.000    15.000    0.6456   13.4458    0.0000    [0.5448, 0.7307]
  21     0.000     0.000    0.6890    8.8464    0.0000    [0.5294, 0.8307]
  25    10.000     0.000    0.6647   14.1815    0.0000    [0.5492, 0.7254]
  30    22.500     0.000    0.6116   11.5752    0.0000    [0.5049, 0.7107]
  35    35.000     0.000    0.5722    8.6064    0.0000    [0.4110, 0.6534]
  40    47.500     0.000    0.5120    6.2842    0.0000    [0.3354, 0.6393]
====================================================================
```

Listing 1: Functionality of `rd2d()`

The first part of the output provides basic information on the options specified in the function. For example, the default estimand is the value of treatment effect, indicated by `deriv = (0,0)`. The rest of the output gives estimation results, including (i) `b1` and `b2`: First and second coordinate of the evaluation points; (ii) `Coef.`: Point estimation of (derivative) of treatment effect using $p =$ `p` polynomial order; and (iii) t-statistics, (iv) p-value, and (v) `level%` confidence intervals using $q =$ `q` polynomial order. When $q > p$, the resulting inference procedures correspond to robust bias correction [Calonico et al., 2014, 2018, 2022], which is the default and recommended method; $p = q$ corresponds to standard least squares methods. Point estimates, standard errors, and other information can be easily extracted for further statistical analysis. The output is stored in a standard matrix, and can be accessed with the following command,

```
> result.rd2d$results
```

```
> result.rd2d$results.A0
> result.rd2d$results.A1
```

Listing 2: Output matrices of `rd2d()`

where `result.rd2d$main.A0` contains results for the control group, `result.rd2d$main.A1` contains results for the treatment group, and `result.rd2d$main` contains results for both.

The `summary()` method allows for three optional arguments. First, the option `subset` takes the indices of evaluation points to be presented, which should be a subset of `c(1:nrow(eval))`. The default is `NULL`, and thus all evaluation points are presented. Second, `CBuniform` is boolean variable for confidence bands construction, where `FALSE` indicates that pointwise confidence intervals are provided, and `TRUE` indicates uniform confidence bands are provided. The default is `CBuniform = FALSE`.

```
> summary(result.rd2d, subset = c(1,5,10,15,21,25,30,35,40), CBuniform = TRUE)
rd2d

Number of Obs.          20000
BW type.                mserd-dpi-std
Kernel                  triangular-prod
VCE method              hc1
Masspoints              check

Number of Obs.          6191          13809
Estimand (deriv)        0             0
Order est. (p)          1             1
Order rbc. (q)          2             2
Unique Obs.             6191          13809

=====================================================================
  ID        b1        b2      Est.         z     P>|z|        95% Unif. CB
=====================================================================
   1     0.000    50.000    0.8143    9.8761    0.0000    [0.5322, 1.0126]
   5     0.000    40.000    0.7298   11.8383    0.0000    [0.5436, 0.9245]
  10     0.000    27.500    0.6752   11.3927    0.0000    [0.4841, 0.8414]
  15     0.000    15.000    0.6456   13.4458    0.0000    [0.4920, 0.7834]
  21     0.000     0.000    0.6890    8.8464    0.0000    [0.4439, 0.9161]
  25    10.000     0.000    0.6647   14.1815    0.0000    [0.4993, 0.7753]
  30    22.500     0.000    0.6116   11.5752    0.0000    [0.4465, 0.7690]
  35    35.000     0.000    0.5722    8.6064    0.0000    [0.3423, 0.7221]
  40    47.500     0.000    0.5120    6.2842    0.0000    [0.2492, 0.7255]
=====================================================================
```

Listing 3: Functionality of `rd2d()` with `CBuniform = TRUE`

When the optional argument `AATE` $= \left(w_j\right)_{j=1}^{J}$, representing the weights for the evaluation points $\{\boldsymbol{b}_j\}_{j=1}^{J}$, is supplied to `summary()`, the output gains an extra row labeled "AATE." This row reports

the aggregated average treatment effect estimator

$$\widehat{\tau}_{\text{AATE}} = \frac{\sum_{j=1}^{J} w_j \, \widehat{\tau}(\boldsymbol{b}_j)}{\sum_{j=1}^{J} w_j},$$

along with its inference statistics.

```
> summary(result.rd2d, subset = c(1,5,10,15,21,25,30,35,40), AATE = rep(1,neval))
rd2d

Number of Obs.          20000
BW type.                mserd-dpi-std
Kernel                  triangular-prod
VCE method              hc1
Masspoints              check

Number of Obs.          6191          13809
Estimand (deriv)        0             0
Order est. (p)          1             1
Order rbc. (q)          2             2
Unique Obs.             6191          13809

=====================================================================
  ID        b1        b2      Est.         z     P>|z|              95% CI
=====================================================================
   1     0.000    50.000    0.8143    9.8761    0.0000    [0.6191, 0.9257]
   5     0.000    40.000    0.7298   11.8383    0.0000    [0.6125, 0.8556]
  10     0.000    27.500    0.6752   11.3927    0.0000    [0.5487, 0.7767]
  15     0.000    15.000    0.6456   13.4458    0.0000    [0.5448, 0.7307]
  21     0.000     0.000    0.6890    8.8464    0.0000    [0.5294, 0.8307]
  25    10.000     0.000    0.6647   14.1815    0.0000    [0.5492, 0.7254]
  30    22.500     0.000    0.6116   11.5752    0.0000    [0.5049, 0.7107]
  35    35.000     0.000    0.5722    8.6064    0.0000    [0.4110, 0.6534]
  40    47.500     0.000    0.5120    6.2842    0.0000    [0.3354, 0.6393]
---------------------------------------------------------------------
AATE                       0.6534   28.9598    0.0000    [0.5987, 0.6856]
=====================================================================
```

Listing 4: Functionality of `rd2d()` for AATE

Finally, the `summary()` method allows for presenting the underlying bandwidths used and associated effective sample sizes via the option `output = "bw"`.

```
> summary(result.rd2d, subset = c(1,5,10,15,21,25,30,35,40), output = "bw")
rd2d

Number of Obs.          20000
BW type.                mserd-dpi-std
Kernel                  triangular-prod
VCE method              hc1
```

```
Masspoints              check

Number of Obs.          6191          13809
Estimand (deriv)        0             0
Order est. (p)          1             1
Order rbc. (q)          2             2
Unique Obs.             6191          13809

=====================================================================
      Bdy Points          BW Control      BW Treatment        Eff. N
  ID       b1       b2      h01      h02      h11      h12      Nh0      Nh1
=====================================================================
   1    0.000   50.000   16.687   16.717   16.687   16.717      689     1295
   5    0.000   40.000   14.782   14.808   14.782   14.808     1019     1810
  10    0.000   27.500   12.736   12.759   12.736   12.759     1252     2108
  15    0.000   15.000   12.045   12.067   12.045   12.067     1295     2167
  21    0.000    0.000   14.740   14.766   14.740   14.766     2169     1552
  25   10.000    0.000   12.455   12.477   12.455   12.477     1467     1980
  30   22.500    0.000   11.476   11.496   11.476   11.496     1107     1843
  35   35.000    0.000   11.595   11.615   11.595   11.615      824     1418
  40   47.500    0.000   14.961   14.987   14.961   14.987      638     1242
=====================================================================
```

Listing 5: Functionality of `rd2d()` with `output = "bw"`

3.2 Function `rdbw2d()`

The function `rdbw2d()` is used for MSE (or IMSE) optimal bandwidth implementation, and is used internally in `rd2d()` when user does not specify bandwidth choices manually. The function takes the same input data as `rd2d()`, that is, an outcome vector `y`, a bivariate location score matrix `X`, a treatment indicator vector `t`, and a grid of evaluation points `b` along the boundary $\mathscr{B}$. In addition, the option `bwselect` encodes four options of bandwidth type: (i) `"mserd"` finds the MSE-optimal bandwidth for estimating (derivatives of) treatment effect, (ii) `"imserd"` finds the integreated MSE optimal bandwidth for estimating (derivatives of) treatment effect, (iii) `"msetwo"` finds the MSE optimal bandwidth for estimation (the derivatives of) conditional means of two potential outcome variables, (iv) `"imsetwo"` finds the integrated MSE optimal bandwidth for estimation (the derivatives of) conditional means of two potential outcome variables. The default is `bwselect = "mserd"`. An additional Boolean argument `stdvar` indicates whether the covariates are first standardized to unit standard deviation in each coordinate, in which case the optimal bandwidth is estimated and then converted back to the original scale. The default is `stdvar = TRUE`.

```
> bws.rd2d <- rdbw2d(y, X, t, eval)
> summary(bws.rd2d,subset = c(1,5,10,15,21,25,30,35,40))
Call: rdbw2d

Number of Obs.          20000
```

```
BW type.                mserd-dpi
Kernel                  triangular-prod
VCE method              hc1
Masspoints              check
Standardization         on

Number of Obs.          6191          13809
Estimand (deriv)        0             0
Order est. (p)          1             1
Unique Obs.             6191          13809

Bandwidth Selection
=====================================================
      Bdy Points          BW Control      BW Treatment
  ID       b1       b2      h01      h02      h11      h12
=====================================================
   1    0.000    2.855   20.552   20.588   20.552   20.588
   5    0.000    2.284   18.668   18.701   18.668   18.701
  10    0.000    1.570   16.476   16.505   16.476   16.505
  15    0.000    0.856   14.098   14.123   14.098   14.123
  21    0.000    0.000   19.115   19.149   19.115   19.149
  25    0.572    0.000   15.737   15.765   15.737   15.765
  30    1.287    0.000   14.000   14.025   14.000   14.025
  35    2.002    0.000   13.562   13.586   13.562   13.586
  40    2.717    0.000   18.798   18.832   18.798   18.832
=====================================================
```

Listing 6: Functionality of `rdbw2d()`

The first part of the `summary` output lists the options used for bandwidth selection. The second part of the `summary` output gives bandwidth selection results, including: (i) Boundary points, `b1` for the first coordinate and `b2` for the second coordinate; (ii) Bandwidths for control group, `h01` for the first coordinate and `h02` for the second coordinate; (iii) Bandwidths for treatment group, `h11` for the first coordinate and `h12` for the second coordinate.

3.3 Function `rd2d.dist()`

The function `rd2d.dist()` provides point estimation and inference for boundary treatment effects using distance-based univariate local polynomial regression. It takes as input an outcome vector `y`, and a signed distance matrix `D` of distance to each boundary point, where each column of `D` corresponds to the signed distance from all observations to one evaluation point, with a positive sign indicating the unit is in the treatment group and a negative sign indicating the unit is in the control group.

Optional arguments include evaluation points `b`, bandwidth choices `h`, degrees of polynomial for point estimation (`p`) and inference (`q`), option for kink adjustment `kink`, and confidence level `level`, among other options. When not provided, the function defaults to estimate the value

of treatment effect using the MSE-optimal bandwidth without kink adjustment (`kink = "off"`), using `p = 1` degree polynomial for point estimation and `q = 2` degree polynomial for robust bias-corrected confidence intervals and bands, providing `level = 95`% confidence interval and uniform confidence bands, and without displaying of evaluation points (`b = NULL`).

```
> result.dist <- rd2d.dist(y,D, b = eval)
> summary(result.dist, subset = c(1,5,10,15,21,25,30,35,40))
rd2d.dist

Number of Obs.          20000
BW type                 mserd-rot
Kernel                  triangular-rad
VCE method              hc1
Masspoints              check

Number of Obs.          6191          13809
Estimand (deriv)        0             0
Order est. (p)          1             1
Order rbc. (q)          2             2

=====================================================================
  ID        b1        b2      Est.         z   P > |z|              95% CI
=====================================================================
   1     0.000    50.000    0.8371   11.5847    0.0000    [0.6514, 0.9167]
   5     0.000    40.000    0.7307    9.4650    0.0000    [0.5807, 0.8840]
  10     0.000    27.500    0.6572    9.7763    0.0000    [0.5123, 0.7692]
  15     0.000    15.000    0.6272    7.7001    0.0000    [0.4344, 0.7311]
  21     0.000     0.000    0.7426    6.8180    0.0000    [0.5056, 0.9136]
  25    10.000     0.000    0.6301   11.2414    0.0000    [0.5474, 0.7785]
  30    22.500     0.000    0.6109    7.3862    0.0000    [0.4337, 0.7470]
  35    35.000     0.000    0.5768    6.4730    0.0000    [0.3668, 0.6855]
  40    47.500     0.000    0.4654    6.7582    0.0000    [0.3602, 0.6544]
=====================================================================
```

Listing 7: Functionality of `rd2d.dist()`

The first part of the `summary` output provides basic information on the options specified in the function. The rest of the `summary` output gives estimation and inference results, including: (i) `b1` and `b2` (when `b` is provided) report first and second coordinates of the evaluation points; (ii) `Coef.` reports the treatment effect estimate using a pth order polynomial regression; and (iii) the last three columns correspond to t-statistic, p-value and `level`% confidence intervals using qth order polynomial regression. The `summary()` method also has the option of displaying the uniform confidence bands (instead of the confidence intervals) in the last two columns as follows (numerical results omitted to conserve space).

```
> summary(result.dist, subset = c(1,5,10,15,21,25,30,35,40), CBuniform = TRUE)
```

Listing 8: Functionality of `rd2d.dist()` with `CBuniform = TRUE`

When the optional argument `AATE` $= \left(w_j\right)_{j=1}^{J}$, representing the weights for the evaluation points $\{\boldsymbol{b}_j\}_{j=1}^{J}$, is supplied to `summary()`, the output gains an extra row labeled "AATE" for estimation and inference of aggregated average treatment effect.

```
> summary(result.dist, subset = c(1,5,10,15,21,25,30,35,40), AATE = rep(1,neval))
```

Listing 9: Functionality of `rd2d()` for AATE

In addition, The `summary()` method can also display the underlying bandwidths and effective samples sizes as follows (numerical results omitted to conserve space).

```
> summary(result.dist, subset = c(1,5,10,15,21,25,30,35,40), output = "bw")
```

Listing 10: Functionality of `rd2d.dist()` with `output = "bw"`

Point estimates, standard errors, and other information can be easily extracted for further statistical analysis. The output is stored in a standard matrix, and can be accessed with the following command.

```
> result.dist$results
> result.dist$results.A0
> result.dist$results.A1
```

Listing 11: Output matrices of `rd2d.dist()`

`result.dist$main.A0` contains results for the control group, `result.dist$main.A1` contains results for the treatment group, and `result.dist$main` contains results for both.

3.4 Function `rdbw2d.dist()`

The function `rdbw2d.dist()` is used to implement MSE (or IMSE) rate-optimal ROT bandwidth selectors, and is used internally in `rd2d.dist()` when user does not provide bandwidths manually. As for `rdbw2d()`, four bandwidth types are allowed, that is, `bwselect` can be `"mserd"`, `"imserd"`, `"msetwo"` or `"imsetwo"`.

```
> bws.dist <- rdbw2d.dist(y,D, b = eval)
> summary(bws.dist,subset = c(1,5,10,15,21,25,30,35,40))
Call: rdbw2d.dist

Number of Obs.          20000
BW type                 mserd-rot
Kernel                  triangular-rad
Kink                    off
VCE method              hc1
Masspoints              check

Number of Obs.          6191          13809
Estimand (deriv)        0             0
Order est. (p)          1             1

Bandwidth Selection
```

```
=================================================
      Bdy Points         BW Control   BW Treatment
  ID       b1       b2             h0             h1
=================================================
   1    0.000   50.000         32.918         32.918
   5    0.000   40.000         22.071         22.071
  10    0.000   27.500         20.899         20.899
  15    0.000   15.000         14.455         14.455
  21    0.000    0.000         14.381         14.381
  25   10.000    0.000         19.712         19.712
  30   22.500    0.000         14.041         14.041
  35   35.000    0.000         16.722         16.722
  40   47.500    0.000         27.499         27.499
=================================================
```

Listing 12: Functionality of `rdbw2d.dist()`

An additional argument `kink`, taking values `"off"` or `"on"`, indicates whether a kink adjustment is made for estimation and inference. The default is `kink = "off"`, but when `kink = "on"` is specified then the bandwidth is shrank to account for lack of smoothness of the assignment boundary $\mathscr{B}$.

```
> bws.dist <- rdbw2d.dist(y,D, kink = "on")
> summary(bws.dist,subset = c(1,5,10,15,21,25,30,35,40))
Call: rdbw2d.dist

Number of Obs.          20000
BW type                 mserd-rot
Kernel                  triangular-rad
Kink                    on
VCE method              hc1
Masspoints              check

Number of Obs.          6191          13809
Estimand (deriv)        0             0
Order est. (p)          1             1

Bandwidth Selection
================================
        BW Control   BW Treatment
  ID             h0             h1
================================
   1         14.421         14.421
   5          9.670          9.670
  10          9.156          9.156
  15          6.333          6.333
  21          6.300          6.300
  25          8.636          8.636
  30          6.152          6.152
  35          7.326          7.326
  40         12.048         12.048
```

```
==============================
```

Listing 13: Functionality of `rdbw2d.dist()` with `kink = "on"`

3.5 Graphical Presentation

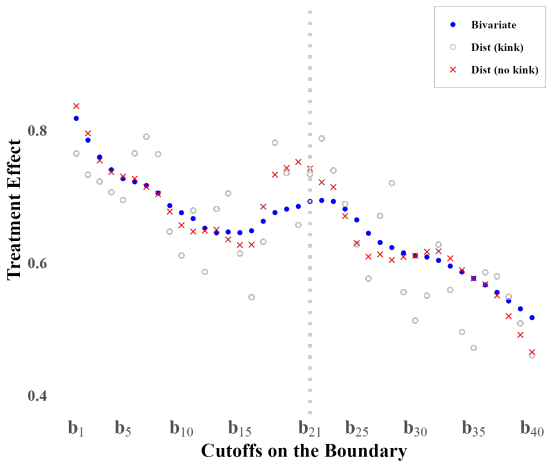
The package `rd2d` provides an array of estimation and inference results that can be used for graphical presentation. Figure 3a compares three point estimation methods: (i) bivariate method via `rd2d`; (ii) distance-based method via `rd2d.dist`, ignoring the kink (default argument `kink = "off"`); and (iii) distance-based method via `rd2d.dist`, adjusting for kink (`kink = "on"`). Figure 3b plots point estimation using the bivariate method, along with its associated robust bias-corrected confidence intervals and confidence band. Figure 3c presents a heatmap of treatment effects along the boundary, with high to low point estimation indicated by red to blue colors. Finally, Figure 3d presents a heatmap of p-values along the boundary, with five colors assigned to five ranges of values. The codes for generating the graphical presentations are given in the replication `R` file.

It can be seen that distance-based estimation with `kink = "off"` overshoots compared to the bivariate estimation before the kink, and undershoots after the kink. This corresponds to the phenomena of getting a first-order bias using distance based method in the presence of a kink, despite using local polynomial regression of degree greater than or equal to 1. See Cattaneo et al. [2025] for more methodological and theoretical discussions.

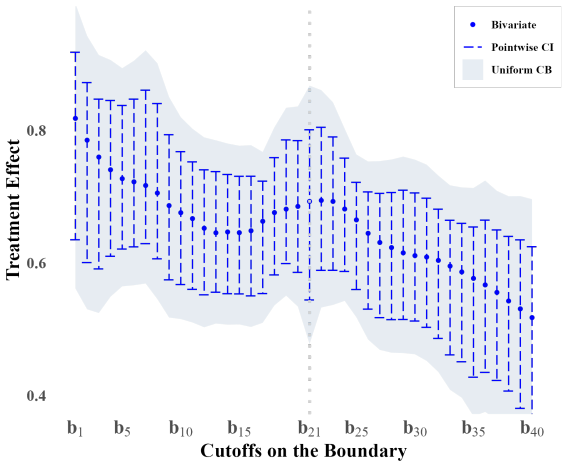
3.6 Simulation Evidence

The discussion so far employed one realization of the data generating process to illustrate the main features of the package `rd2d`. In this final section, we conduct a Monte Carlo experiment to assess the performance of the package in repeated sampling. We consider $m = 1,000$ simulations of the two data generating processes defined at the beginning of this section, and report the simulation results in Table 2 (DGP 1: linear model) and Table 3 (DGP 2: quadratic model). Three methods are used and compared: the bivariate method `rd2d`, the distance-based method `rd2d.dist` ignoring the presence of the kink in the boundary (`kink = "off"`), and the distance-based method `rd2d.dist` adjusting for the kink (`kink = "on"`). Bandwidths are chosen automatically by the package, and their average across simulations is reported. We also report diagnostic measures including: bias, standard deviation of point estimator, root mean-squared error of point estimator, pointwise empirical coverage, pointwise interval length, uniform empirical coverage, and uniform interval length.

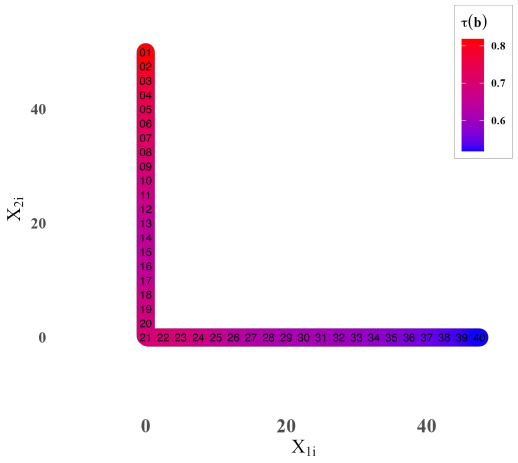
For both DPG 1 and DGP 2, all of the three methods give pointwise coverage around 95%, while the pointwise interval length for `rd2d` and `rd2d.dist (kink = "off")` are smaller compared to `rd2d.dist (kink = "on")`. This is likely due to bandwidth shrinkage for kink adjustment, which does results in a significantly smaller bias compared to the one ignoring the kink. Both `rd2d` and `rd2d.dist (kink = "off")` give around 95% uniform coverage and a relatively shorter interval length compared to `rd2d.dist (kink = "on")`, likely due to the same reason.



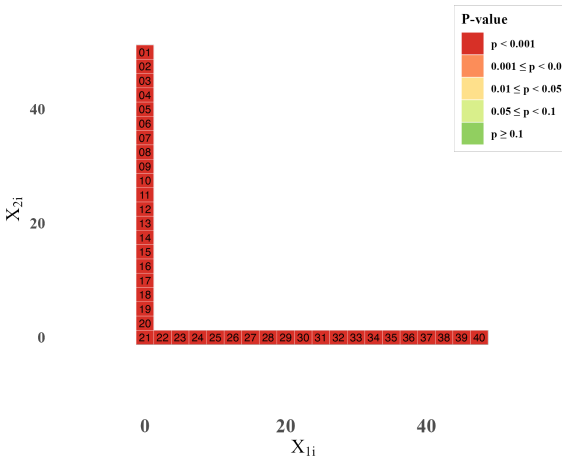
(a) Point Estimation



(b) Confidence Interval and Bands



(c) Treatment Effects Heatmap



(d) p-values Heatmap

Figure 3: Estimation and Inference (Simulations).

4 Conclusion

This paper introduced the `R` software package `rd2d` for causal inference in Boundary Discontinuity designs. The package provides pointwise and uniform (over the treatment assignment boundary) estimation and inference methods employing either a bivariate location score or a univariate distance score. In addition, the methods can be used for graphical presentation. From a methodological perspective, this paper introduced second generation bandwidth selection methods complementing the main results in Cattaneo et al. [2025]. Simulation evidence demonstrated a good performance of the package `rd2d`. Replication codes and related information are available at: https://rdpackages.github.io/rd2d/.

5 Acknowledgments

We thank Xinwei Ma and Filippo Palomba for insightful discussions. Cattaneo and Titiunik gratefully acknowledge financial support from the National Science Foundation (SES-2019432 and SES-2241575). Cattaneo gratefully acknowledge financial support from the Data-Driven Social Science initiative at Princeton University.

Method	Index	h	Bias	SD	RMSE	EC	IL
rd2d	1	15.764	0.005	0.053	0.054	0.953	0.313
	5	13.885	0.003	0.042	0.042	0.958	0.243
	10	11.997	0.002	0.038	0.038	0.952	0.218
	15	12.706	0.001	0.033	0.033	0.949	0.191
	21	13.788	-0.003	0.049	0.049	0.957	0.329
	25	12.908	0.001	0.031	0.031	0.950	0.177
	30	11.688	-0.001	0.036	0.036	0.948	0.212
	35	13.012	0.001	0.039	0.039	0.957	0.228
	40	15.255	0.001	0.049	0.049	0.960	0.289
	Uniform					0.948	0.348
rd2d.dist kink = "off"	1	34.136	0.031	0.038	0.049	0.940	0.261
	5	26.747	0.016	0.035	0.038	0.950	0.240
	10	19.478	0.003	0.035	0.035	0.953	0.258
	15	16.905	0.001	0.038	0.038	0.957	0.275
	21	21.151	0.002	0.038	0.038	0.960	0.283
	25	18.524	-0.021	0.039	0.044	0.955	0.255
	30	16.986	-0.000	0.037	0.037	0.948	0.280
	35	24.532	-0.005	0.033	0.033	0.956	0.233
	40	33.070	-0.015	0.035	0.038	0.953	0.244
	Uniform					0.946	0.408
rd2d.dist kink = "on"	1	14.955	0.012	0.079	0.080	0.954	0.729
	5	11.718	0.003	0.072	0.072	0.948	0.646
	10	8.533	0.001	0.078	0.078	0.954	0.698
	15	7.406	-0.002	0.083	0.083	0.948	0.753
	21	9.266	-0.003	0.083	0.083	0.944	0.810
	25	8.116	0.005	0.076	0.077	0.955	0.697
	30	7.442	-0.001	0.079	0.079	0.951	0.761
	35	10.748	0.002	0.067	0.067	0.948	0.626
	40	14.488	-0.001	0.075	0.075	0.948	0.667
	Uniform					0.89	1.073

Table 2: Simulation results for DGP 1 (Linear)

Method	Index	h	Bias	SD	RMSE	EC	IL
rd2d	1	15.730	0.004	0.052	0.053	0.958	0.315
	5	13.803	0.005	0.041	0.041	0.965	0.244
	10	11.954	0.003	0.038	0.038	0.943	0.218
	15	12.600	0.001	0.033	0.033	0.944	0.193
	21	13.716	-0.008	0.049	0.049	0.962	0.331
	25	12.833	-0.003	0.031	0.031	0.952	0.177
	30	11.679	0.000	0.035	0.034	0.962	0.212
	35	12.966	-0.001	0.041	0.041	0.954	0.230
	40	15.312	-0.007	0.050	0.050	0.942	0.288
	Uniform					0.942	0.350
rd2d.dist kink = "off"	1	34.303	0.040	0.038	0.055	0.927	0.260
	5	24.471	0.016	0.037	0.041	0.960	0.262
	10	18.392	-0.001	0.036	0.036	0.951	0.273
	15	14.179	-0.003	0.042	0.042	0.944	0.324
	21	21.170	-0.004	0.037	0.038	0.959	0.283
	25	16.036	-0.020	0.045	0.049	0.943	0.293
	30	13.771	-0.002	0.043	0.043	0.947	0.341
	35	19.017	-0.016	0.043	0.046	0.954	0.297
	40	31.541	-0.071	0.042	0.083	0.935	0.260
	Uniform					0.949	0.461
rd2d.dist kink = "on"	1	15.028	0.015	0.079	0.080	0.950	0.718
	5	10.721	0.005	0.074	0.074	0.958	0.710
	10	8.058	0.000	0.082	0.082	0.937	0.746
	15	6.212	-0.008	0.097	0.097	0.942	0.903
	21	9.275	-0.001	0.084	0.084	0.945	0.811
	25	7.026	0.001	0.088	0.088	0.950	0.810
	30	6.033	0.006	0.102	0.102	0.944	0.947
	35	8.331	-0.001	0.090	0.090	0.957	0.817
	40	13.818	-0.020	0.079	0.081	0.952	0.708
	Uniform					0.822	1.199

Table 3: Simulation results for DGP 2 (Quadratic)

References

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

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

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

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

Matias D Cattaneo, Rocio Titiunik, and Ruiqi Rae Yu. Estimation and inference in boundary discontinuity designs. *arXiv preprint arXiv:2505.05670*, 2025.

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, Axel Werwatz, et al. *Nonparametric and semiparametric models*, volume 1. Springer, 2004.

Ekaterina Jardim, Mark C Long, Robert Plotnick, Jacob Vigdor, and Emma Wiles. Local minimum wage laws, boundary discontinuity methods, and policy spillovers. *Journal of Public Economics*, 234:105131, 2024.

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

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.

Matt P Wand and M Chris Jones. *Kernel smoothing*. CRC press, 1994.

Achim Zeileis. Econometric computing with hc and hac covariance matrix estimators. *Journal of Statistical Software*, 11(10):1–17, 2004.

Achim Zeileis, Susanne Köll, and Nathaniel Graham. Various versatile variances: An object-oriented implementation of clustered covariances in r. *Journal of Statistical Software*, 95:1–36, 2020.