rdrobust: An **R** Package for Robust Nonparametric Inference in Regression-Discontinuity Designs

Sebastian Calonico
University of Michigan

Matias D. Cattaneo
University of Michigan

Rocio Titiunik
University of Michigan

Abstract

This article gives an introduction to the R package **rdrobust**. This package includes three main functions to conduct robust data-driven statistical inference in regression-discontinuity (RD) designs. The first and main function, `rdrobust`, implements conventional nonparametric RD treatment-effect point estimators and confidence intervals, as well as the robust bias-corrected confidence intervals proposed in Calonico, Cattaneo, and Titiunik (2013c) for local average treatment effects. This function covers sharp RD, sharp kink RD, fuzzy RD and fuzzy kink RD designs, among other possibilities. The second function, `rdbwselect`, implements several bandwidth selectors proposed in the RD literature. Finally, the third function, `rdbinselect`, implements a novel data-driven optimal choice of evenly-spaced bins, employs the resulting optimal bins to approximate the underlying regression functions by local sample means, and constructs the familiar RD plots usually found in empirical applications. A companion `Stata` package is described in Calonico, Cattaneo, and Titiunik (2013b).

Preliminary and incomplete, comments welcome

This draft: December 9, 2013

Keywords: regression discontinuity, local polynomial, robust inference, R.

1. Introduction

The regression-discontinuity (RD) design is a widely employed empirical research design in social, behavioral and related sciences.[1] In this research design units are assigned to treatment based on whether their value of an observed covariate is above or below a known cutoff, and the probability of receiving treatment conditional on this covariate jumps discontinuously at

---

[1] For review see, among others, Cook (2008), Imbens and Lemieux (2008), van der Klaauw (2008), Lee and Lemieux (2010) and Dinardo and Lee (2011).

the cutoff. This jump induces "variation" in treatment assignment that may be regarded, under appropriate assumptions, as being unrelated to potential confounders. Thus, inference in RD designs is typically conducted using only observations near the cutoff or threshold, where the discontinuous change in the probability of treatment assignment occurs. Due to its local nature, RD average treatment effects estimators are usually constructed using local-polynomial nonparametric regression, and statistical inference is based on large-sample approximations.[2]

This article gives an introduction to the R package **rdrobust**, which offers an array of data-driven (i.e., fully automatic) local-polynomial-based inference procedures in the RD design. We introduce three main functions allowing for a variety of data-driven nonparametric point and confidence intervals estimators, as well as bandwidth selectors and plotting procedures useful for RD empirical applications. The first and main function is `rdrobust`, which implements the bias-corrected robust (to "large" bandwidth choices) inference procedure proposed by Calonico, Cattaneo, and Titiunik (2013c). This function also implements many other RD inference procedures employing local-polynomial regression; see, e.g., references in footnote 1. The function **rdrobust** offers bias-corrected confidence intervals for average treatment effects at the cutoff for sharp RD, sharp kink RD, fuzzy RD and fuzzy kink RD designs, among other possibilities.

The second function is `rdbwselect`. This function implements several data-driven bandwidth selectors for RD designs based on the recent work of Imbens and Kalyanaraman (2012) and Calonico *et al.* (2013c). Although this command may be used as a stand-alone bandwidth selector in RD applications, its main purpose is to provide fully data-driven bandwidth choices to be used by our main function `rdrobust`.

Finally, our third function is `rdbinselect`. This function implements a data-driven optimal length choice of equally-spaced bins, which are useful to approximate the regression function by local sample averages of the outcome variable. Following Cattaneo and Farrell (2013), this optimal choice is based on an integrated mean-square error expansion of the appropriate estimators. We discuss how these binned sample means, and hence the bin-length choice, may be used to construct the plots commonly found in RD applications. Employing this optimal choice, `rdbinselect` offers an automatic way of constructing RD plots.

The rest of this article is organized as follows. Section 2 provides a brief review of all the methods implemented in our commands. Section 3 illustrates some of the main features of our functions `rdrobust`, `rdbwselect` and `rdbinselect`, by employing the RD research design in Cattaneo *et al.* (2013) and analyzing the corresponding data on party advantages in the U.S. Senate. Section 4 concludes and briefly describes some upcoming extensions. A full description of the capabilities of the R package **rdrobust** may be found in the R help files. A companion Stata package is described in Calonico, Cattaneo, and Titiunik (2013b).

2. Methods

We now present the basic RD framework, describe the population parameters of interest, introduce the local-polynomial based estimators, review the different inference procedures and briefly summarize the RD plotting capabilities available in our R package **rdrobust**. We do

---

[2]An alternative approach based on exact finite-sample inference, under different assumptions, is discussed in Cattaneo, Frandsen, and Titiunik (2013).

not, however, discuss regularity conditions or other technical aspects underlying the estimands and estimators – these may be found in the references given throughout. In addition to the references given in footnote 1, recent results and further details on RD designs are discussed in Imbens and Kalyanaraman (2012, IK hereafter), (Calonico, Cattaneo, and Titiunik 2013c, CCT hereafter) and its supplemental appendix (Calonico, Cattaneo, and Titiunik (2013d)), and references therein.

2.1. Setup and Notation

We adopt the potential outcomes framework commonly employed in the treatment effects literature (e.g., Heckman and Vytlacil (2007) and Imbens and Wooldridge (2009) for reviews). Let $\{(Y_i(0), Y_i(1), T_i(0), T_i(1), X_i)' : i = 1, 2, \ldots, n\}$ be a random sample from $(Y(0), Y(1), T(0), T(1), X)'$, where $Y(1)$ and $Y(0)$ denote the potential outcomes with and without treatment, respectively, $T(0)$ and $T(1)$ denote potential actual treatment status, and the scalar regressor $X_i \in \mathbb{R}$ is the so-called "running variable" or "score" determining treatment assignment based on whether it exceeds a known cutoff. In particular, unit i is assigned treatment if $X_i > \bar{x}$ and not assigned treatment if $X_i < \bar{x}$, for some known fixed value $\bar{x} \in \mathbb{R}$.

This setup allows for imperfect compliance, which in the RD literature is known as the *fuzzy RD design*. The case of perfect treatment compliance is usually called the *sharp RD design*. In either case, the observed outcome and treatment status are

$$Y_i = \begin{cases} Y_i(0) & \text{if } X_i < \bar{x} \\ Y_i(1) & \text{if } X_i \geq \bar{x} \end{cases} \qquad \text{and} \qquad T_i = \begin{cases} T_i(0) & \text{if } X_i < \bar{x} \\ T_i(1) & \text{if } X_i \geq \bar{x} \end{cases},$$

respectively. Notice that $\mathbb{P}[T_i = 0 | X_i < \bar{x}] = 1 = \mathbb{P}[T_i = 1 | X_i \geq \bar{x}]$ in the sharp RD design, while this need not to be the case in the fuzzy RD design.

The observed data is $\{(Y_i, T_i, X_i)' : i = 1, 2, \ldots, n\}$, a random sample from a large population, allowing for (but not requiring that) $T_i = \mathbb{1}(X_i \geq \bar{x})$ with $\mathbb{1}(\cdot)$ denoting the indicator function. Therefore, for each unit i, the scalar random variable $Y_i \in \mathbb{R}$ denotes the outcome of interest, and $T_i \in \{0, 1\}$ denotes actual treatment take-up ($T_i = 1$ treatment taken, $T_i = 0$ treatment not taken).

We also introduce additional notation that will be useful throughout the paper. For $\nu \in \mathbb{Z}_+ = \{0, 1, 2, \cdots\}$, define

$$\mu_{Y+}^{(\nu)}(\bar{x}) = \lim_{x \to \bar{x}^+} \frac{\partial^\nu}{\partial x^\nu} \mu_{Y+}(x), \qquad \mu_{Y-}^{(\nu)}(\bar{x}) = \lim_{x \to \bar{x}^-} \frac{\partial^\nu}{\partial x^\nu} \mu_{Y-}(x),$$

with $\mu_{Y+}(x) = \mathbb{E}[Y(1)|X = x]$ and $\mu_{Y-}(x) = \mathbb{E}[Y(0)|X = x]$, and

$$\mu_{T+}^{(\nu)}(\bar{x}) = \lim_{x \to \bar{x}^+} \frac{\partial^\nu}{\partial x^\nu} \mu_{T+}(x), \qquad \mu_{T-}^{(\nu)}(\bar{x}) = \lim_{x \to \bar{x}^-} \frac{\partial^\nu}{\partial x^\nu} \mu_{T-}(x),$$

with $\mu_{T+}(x) = \mathbb{E}[T(1)|X = x]$ and $\mu_{T-}(x) = \mathbb{E}[T(0)|X = x]$. We also define

$$\sigma_{Y+}^2(\bar{x}) = \lim_{x \to \bar{x}^+} \sigma_Y^2(x), \qquad \sigma_{Y-}^2(\bar{x}) = \lim_{x \to \bar{x}^-} \sigma_Y^2(x), \qquad \sigma_Y^2(x) = \mathbb{V}[Y_i | X_i = x],$$

$$\sigma_{T+}^2(\bar{x}) = \lim_{x \to \bar{x}^+} \sigma_T^2(x), \qquad \sigma_{T-}^2(\bar{x}) = \lim_{x \to \bar{x}^-} \sigma_T^2(x), \qquad \sigma_T^2(x) = \mathbb{V}[Y_i | X_i = x],$$

and let $f(x)$ be the continuous (Lebesgue) density of X_i.

Whenever there is no confusion, we drop the subindex denoting the dependent variable or the point of evaluation in the conditional expectations and other functions.

2.2. Population Parameters of Interest

We focus on mean treatment effects. In particular, we consider local average treatment effects at the cutoff in the sharp RD, fuzzy RD, sharp kink RD and fuzzy kink RD designs. Our package also covers higher-order derivative estimation in the context of both sharp and fuzzy RD designs. For other extensions of interest see Section 4. For further details on the interpretation of these estimands and regularity conditions see, among others, Hahn, Todd, and van der Klaauw (2001), Lee (2008), Card, Lee, Pei, and Weber (2012), Dong (2012), Dong and Lewbel (2012) and the references mentioned in footnote 1.

Sharp RD Designs

In the sharp RD design, two popular parameters of interest are the sharp RD average treatment effect at the threshold, denoted by τ_0, and the the sharp kink RD average treatment effect at the threshold, denoted by τ_1, with the notation

$$\tau_\nu := \tau_{Y,\nu}(\bar{x}) = \left.\frac{\partial^\nu}{\partial x^\nu}\mathbb{E}[Y_i(1) - Y_i(0)|X_i = x]\right|_{x=\bar{x}}, \qquad \nu \in \mathbb{Z}_+,$$

where the definition drops the subindex denoting the dependent random variable and evaluation point for notation simplicity. These popular estimands in the RD literature are nonparametrically identifiable under mild continuity conditions. Specifically, these parameters can be written as a function of observed data as follows:

$$\tau_\nu = \mu_{Y+}^{(\nu)} - \mu_{Y-}^{(\nu)},$$

where here we drop the the evaluation point $(\bar{x})$ for simplicity. Notice that $\mu_{Y+}^{(\nu)} = \mu_{Y+}^{(\nu)}(\bar{x})$ and $\mu_{Y-}^{(\nu)} = \mu_{Y-}^{(\nu)}(\bar{x})$ are estimable from the observed data: $\tau_\nu = \tau_{Y,\nu}$ is a difference of two (one-sided) nonparametric regression functions at $\bar{x}$.

Fuzzy RD Designs

In the fuzzy RD designs, the two main population parameters are ς_0 and ς_1, where

$$\varsigma_\nu := \varsigma_\nu(\bar{x}) = \frac{\left.\frac{\partial^\nu}{\partial x^\nu}\mathbb{E}[Y_i(1) - Y_i(0)|X_i = x]\right|_{x=\bar{x}}}{\left.\frac{\partial^\nu}{\partial x^\nu}\mathbb{E}[T_i(1) - T_i(0)|X_i = x]\right|_{x=\bar{x}}}, \qquad \nu \in \mathbb{Z}_+.$$

These parameters are the instrumental variable analogue estimands of τ_0 and τ_1, respectively. In the RD literature, ς_0 is the fuzzy RD average treatment effect at the cutoff and ς_1 is the fuzzy kink RD average treatment effect at the cutoff. As in the case of the sharp RD design, the fuzzy population parameters are (under appropriate regularity conditions) nonparametrically identifiable as

$$\varsigma_\nu = \frac{\tau_{Y,\nu}}{\tau_{T,\nu}} = \frac{\mu_{Y+}^{(\nu)} - \mu_{Y-}^{(\nu)}}{\mu_{T+}^{(\nu)} - \mu_{T-}^{(\nu)}},$$

where here $\mu_{Y+}^{(\nu)} = \mu_{Y+}^{(\nu)}(\bar{x})$, $\mu_{Y-}^{(\nu)} = \mu_{Y-}^{(\nu)}(\bar{x})$, $\mu_{T+}^{(\nu)} = \mu_{T+}^{(\nu)}(\bar{x})$ and $\mu_{T-}^{(\nu)} = \mu_{T-}^{(\nu)}(\bar{x})$ are all estimable from the observed data $-\tau_{Y,\nu}$ and $\tau_{T,\nu}$ are each a difference of two (one-sided) nonparametric regression problems at $\bar{x}$, and ς_{ν} is just their ratio.

2.3. Local Polynomial Estimators

Statistical inference in the RD design reduces to nonparametric regression-based inference at the induced boundary point $\bar{x}$, and thus employs observations at either side of the threshold separately. Naturally, local-polynomial estimators have become the preferred choice of nonparametric estimator used in the RD literature (see Hahn *et al.* (2001), Porter (2003) and the references mentioned in footnote 1) because of their excellent boundary properties (see, e.g., Fan and Gijbels (1996) and Cheng, Fan, and Marron (1997)).

Our R package **rdrobust** implements local polynomial estimators of various orders for the two outcome variables (Y_i and T_i), as appropriate depending on the RD design considered, and also includes different bandwidths selectors and alternative confidence intervals estimators. All these features are briefly reviewed in the following subsections, but first we introduce the local polynomial RD estimators of order p in general. To reduce repetition, we describe the estimators using a generic outcome variable Z which either takes the value Y or T, depending on the outcome variable under consideration. For $Z \in \{Y, T\}$ and $\nu, p \in \mathbb{Z}_+$ with $\nu \leq p$,

$$\hat{\tau}_{Z,\nu,p}(x; h_n) = \hat{\mu}_{Z+,p}^{(\nu)}(x; h_n) - \hat{\mu}_{Z-,p}^{(\nu)}(x; h_n),$$

$$\hat{\mu}_{Z+,p}^{(\nu)}(x; h_n) = \mathbf{e}_{\nu}' \hat{\boldsymbol{\beta}}_{Z+,p}(x; h_n) \qquad \text{and} \qquad \hat{\mu}_{Z-,p}^{(\nu)}(x; h_n) = \mathbf{e}_{\nu}' \hat{\boldsymbol{\beta}}_{Z-,p}(x; h_n),$$

$$\hat{\boldsymbol{\beta}}_{Z+,p}(x; h_n) = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} \mathbb{1}(X_i \geq x)(Z_i - \mathbf{r}_p(X_i - x)'\boldsymbol{\beta})^2 K_{h_n}(X_i - x),$$

$$\hat{\boldsymbol{\beta}}_{Z-,p}(x; h_n) = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} \mathbb{1}(X_i < x)(Z_i - \mathbf{r}_p(X_i - x)'\boldsymbol{\beta})^2 K_{h_n}(X_i - x),$$

where here $\mathbf{r}_p(x) = (1, x, \cdots, x^p)'$, $\mathbf{e}_{\nu}$ is the conformable $(\nu + 1)$-th unit vector (e.g., $\mathbf{e}_1 = (0, 1, 0)'$ if $p = 2$), $K_h(u) = K(u/h)/h$ with $K(\cdot)$ a kernel function and h_n is a positive bandwidth sequence.

RD Estimators and Consistency

Using the generic notation above, we can easily construct the sharp RD estimators:

$$\hat{\tau}_{\nu,p}(h_n) := \hat{\tau}_{Y,\nu,p}(\bar{x}; h_n), \qquad \nu \leq p.$$

Similarly, for the fuzzy RD designs we have the RD estimators:

$$\hat{\varsigma}_{\nu,p}(h_n) := \frac{\hat{\tau}_{Y,\nu,p}(h_n)}{\hat{\tau}_{T,\nu,p}(h_n)}, \qquad \hat{\tau}_{Y,\nu,p}(h_n) := \hat{\tau}_{Y,\nu,p}(\bar{x}; h_n), \qquad \hat{\tau}_{T,\nu,p}(h_n) := \hat{\tau}_{T,\nu,p}(\bar{x}; h_n), \qquad \nu \leq p.$$

Assuming the bandwidth h_n vanishes at an appropriate rate, and other regularity conditions hold, consistency of these estimators follows easily from well-known properties of local polynomial estimators (Fan and Gijbels (1996)). Specifically, consistency follows from the fact

that local polynomial estimators satisfy, for $Z \in \{Y, T\}$,

$$\hat{\boldsymbol{\beta}}_{Z+,p}(x; h_n) \to_{\mathrm{p}} \boldsymbol{\beta}_{Z+,p}(x) \qquad \text{and} \qquad \boldsymbol{\beta}_{Z+,p}(x) = \left[\mu_{Z+}(x), \mu_{Z+}^{(1)}(x), \frac{\mu_{Z+}^{(2)}(x)}{2!}, \cdots, \frac{\mu_{Z+}^{(p)}(x)}{p!} \right]',$$

and

$$\hat{\boldsymbol{\beta}}_{Z-,p}(x; h_n) \to_{\mathrm{p}} \boldsymbol{\beta}_{Z-,p}(x) \qquad \text{and} \qquad \boldsymbol{\beta}_{Z-,p}(x) = \left[\mu_{Z-}(x), \mu_{Z-}^{(1)}(x), \frac{\mu_{Z-}^{(2)}(x)}{2!}, \cdots, \frac{\mu_{Z-}^{(p)}(x)}{p!} \right]'.$$

In applications, the most common choices are $p = 1$ for τ_0 (local-linear sharp RD estimator), $p = 2$ for τ_1 (local-quadratic sharp kink RD estimator), $p = 1$ for ς_0 (local-linear fuzzy RD estimator), and $p = 2$ for ς_1 (local-quadratic fuzzy kink RD estimator).

2.4. Bandwidth Selectors

This subsection discusses bandwidth selection, the main obstacle in the practical implementation of RD point estimators. We briefly review two approaches: (i) plug-in rules based on mean-square error (MSE) expansions, and (ii) a cross-validation procedure. IK provide a comprehensive review of these approaches, and CCT discuss some extensions and alternative ways of implementation. We define $\mathbf{X}_n = (X_1, X_2, \cdots, X_n)'$ to save notation.

Direct Plug-in Rules

The R package **rdrobust** implements several direct plug-in (DPI) approaches to select the bandwidths, all based on a mean-square error expansion of the sharp RD estimators. CCT also discuss briefly the analogous DPI rules for fuzzy designs, but these are not currently implemented. (Of course, using our package, one can still choose optimal sharp RD bandwidths separately for the numerator and denominator of the fuzzy RD estimators.) We have the following result for the (conditional) mean-square error:

$$\mathbb{E}\left[\left(\hat{\tau}_{\nu,p}(h_n) - \tau_\nu \right)^2 \Big| \mathbf{X}_n \right] \approx h_n^{2(p+1-\nu)} \mathsf{B}_{\nu,p}^2 + \frac{1}{n h_n^{1+2\nu}} \mathsf{V}_{\nu,p}, \qquad \nu \le p,$$

where the bias and variance are, respectively,

$$\mathsf{B}_{\nu,p} = \frac{\mu_+^{(p+1)} - \mu_-^{(p+1)}}{(p+1)!} \mathbf{e}_\nu' \Gamma_p^{-1} \vartheta_p, \qquad \mathsf{V}_{\nu,p} = \frac{\sigma_-^2 + \sigma_+^2}{f} \nu!^2 \mathbf{e}_\nu' \Gamma_p^{-1} \Psi_p \Gamma_p^{-1} \mathbf{e}_\nu, \qquad (1)$$

and $f = f(\bar{x})$. CCT describe the exact form of the non-random scalars $\mathbf{e}_\nu' \Gamma_p^{-1} \vartheta_p$ and $\mathbf{e}_\nu' \Gamma_p^{-1} \Psi_p \Gamma_p^{-1} \mathbf{e}_\nu$, which are functions of the kernel choice. Notice that we have dropped the dependent variables and evaluation points from the notation for simplicity – these expressions apply to both Y_i and T_i, and the functions are evaluated at $x = \bar{x}$.

It follows that, if $\mu_+^{(p+1)} \neq \mu_-^{(p+1)}$, then the (asymptotic) MSE-optimal bandwidth is

$$h_{\mathrm{MSE},\nu,p} = C_{\mathrm{MSE},\nu,p}\, n^{-\frac{1}{2p+3}}, \qquad C_{\mathrm{MSE},\nu,p} = \left(\frac{(1 + 2\nu) \mathsf{V}_{\nu,p}}{2(p + 1 - \nu) \mathsf{B}_{\nu,p}^2} \right)^{\frac{1}{2p+3}}.$$

In the context of $\nu = 0$, IK noted that $\mathsf{B}_{0,p,p+1} \propto \mu_+^{(p+1)} - \mu_-^{(p+1)}$ may be (close to) zero in some applications and thus proposed the more "robust", data-driven consistent bandwidth estimator:

$$\hat{h}_{\mathtt{IK},0,p} = \left(\frac{\hat{\mathsf{V}}_{\mathtt{IK},0,p}}{2(p+1)\hat{\mathsf{B}}_{\mathtt{IK},0,p}^2 + \hat{\mathsf{R}}_{\mathtt{IK},0,p}} \right)^{1/(2p+3)} n^{-1/(2p+3)},$$

where the additional (regularization) term $\hat{\mathsf{R}}_{\mathtt{IK},0,p}$ is introduced to avoid small denominators in finite samples. Here $\hat{\mathsf{B}}_{\mathtt{IK},0,p}$ and $\hat{\mathsf{V}}_{\mathtt{IK},0,p}$ (and $\hat{\mathsf{R}}_{\mathtt{IK},0,p}$) are nonparametric consistent estimators of their respective population counterparts, which require the choice of preliminary bandwidths, generically denoted by b_n herein. IK provide a simple, direct implementation approach for $p = 1$ where these estimators ($\hat{\mathsf{B}}_{\mathtt{IK},0,p}$, $\hat{\mathsf{V}}_{\mathtt{IK},0,p}$, $\hat{\mathsf{R}}_{\mathtt{IK},0,p}$) are consistent for their population counterparts, but the preliminary bandwidths used in their construction are not chosen optimally. As a consequence, $\hat{h}_{\mathtt{IK},\nu,p}$ may be viewed as a nonparametric first-generation plug-in rule (e.g., Wand and Jones (1995)), sometimes called a DPI-1 (direct plug-in of order 1) selector.

CCT propose a second-generation plug-in bandwidth selection approach, extending the work of IK. Specifically, CCT propose a second-order direct plug-in rule (DPI-2) of the form:

$$\hat{h}_{\mathtt{CCT},\nu,p} = \left(\frac{(1+2\nu)\hat{\mathsf{V}}_{\mathtt{CCT},\nu,p}}{2(p+1-\nu)\hat{\mathsf{B}}_{\mathtt{CCT},\nu,p}^2 + \hat{\mathsf{R}}_{\mathtt{CCT},\nu,p}} \right)^{1/(2p+3)} n^{-1/(2p+3)}, \qquad \nu \le p.$$

This alternative bandwidth estimator has two distinct features relative to $\hat{h}_{\mathtt{IK},0,p}$. First, the preliminary bandwidths used in the construction of the (consistent) estimators $\hat{\mathsf{V}}_{\mathtt{CCT},\nu,p}$ and $\hat{\mathsf{B}}_{\mathtt{CCT},\nu,p}$ are consistent estimators of the corresponding population MSE-optimal bandwidths. In this sense, $\hat{h}_{\mathtt{CCT},\nu,p}$ is a DPI-2 (direct plug-in of order 2) selector. Second, CCT construct an alternative estimator of $\mathsf{V}_{\nu,p}$ (denoted by $\hat{\mathsf{V}}_{\mathtt{CCT},\nu,p}$ above), that does not require an additional choice of bandwidth for its construction and thus may lead to better small-sample performance. This estimator relies on a fixed-matches nearest-neighbor-based "estimate" of the residuals, following the work of Abadie and Imbens (2006), rather than on estimating the residuals using conventional nonparametric methods. This approach and the traditional one are both discussed further below where we discuss the construction of confidence intervals.

All other details on implementation of the bandwidth selectors included in the R package **rdrobust** may be found in CCT.

Cross-validation Bandwidth Selector

The cross-validation bandwidth choice implemented in our package is as follows:

$$\hat{h}_{\mathtt{CV},p} = \arg\min_{h>0} \mathsf{CV}_\delta(h), \qquad \mathsf{CV}_\delta(h) = \sum_{i=1}^{n} \mathbb{1}(X_{-,[\delta]} \le X_i \le X_{+,[\delta]}) \left(Y_i - \hat{\mu}_p(X_i; h) \right)^2,$$

where

$$\hat{\mu}_p(x; h) = \begin{cases} \mathbf{e}_0' \hat{\boldsymbol{\beta}}_{Y+,p}(x, h) & \text{if } x > \bar{x} \\ \mathbf{e}_0' \hat{\boldsymbol{\beta}}_{Y-,p}(x, h) & \text{if } x < \bar{x} \end{cases},$$

and, for $\delta \in (0, 1)$, $X_{-,[\delta]}$ and $X_{+,[\delta]}$ denote the δ-th quantile of $\{X_i : X_i < \bar{x}\}$ and $\{X_i : X_i > \bar{x}\}$, respectively. Of course, this approach may also be employed using T_i as the outcome variable. See IK for further discussion on this alternative bandwidth selection approach.

2.5. Asymptotic Properties and Confidence Intervals

We briefly review the main asymptotic properties of the local polynomial RD estimators for sharp RD designs, with particular emphasis on the properties of the associated confidence interval estimators. Specifically, we discuss three type of confidence intervals (CI) based on Gaussian approximations: (i) conventional CI based on undersmoothing, (ii) bias-corrected CI (not necessarily requiring undersmoothing), and (iii) robust bias-corrected CI (not necessarily requiring undersmoothing). All the results presented herein extend immediately to the case of fuzzy RD designs, as we will discuss briefly at the end of this section.

Optimal Point Estimators

The results discussed so far lead to the following data-driven RD treatment effect point estimators.

$$
\begin{array}{llll}
\text{Sharp RD:} & \hat{\tau}_{\nu,p}(\hat{h}_{\text{IK},\nu,p}), & \hat{\tau}_{\nu,p}(\hat{h}_{\text{CCT},\nu,p}), & \hat{\tau}_{\nu,p}(\hat{h}_{\text{CV},p}), & \nu \leq p. \\
\text{Fuzzy RD:} & \hat{\varsigma}_{\nu,p}(\hat{h}_{\text{IK},\nu,p}), & \hat{\varsigma}_{\nu,p}(\hat{h}_{\text{CCT},\nu,p}), & \hat{\varsigma}_{\nu,p}(\hat{h}_{\text{CV},p}), & \nu \leq p.
\end{array}
$$

Notice that these estimators are constructed employing MSE-optimal bandwidth choices for the sharp RD case, which means that $\hat{\tau}_{\nu,p}(\hat{h}_{\text{IK},\nu,p})$, $\hat{\tau}_{\nu,p}(\hat{h}_{\text{CCT},\nu,p})$ and $\hat{\tau}_{\nu,p}(\hat{h}_{\text{CV},p})$ may be interpreted as consistent and (asymptotically) MSE-optimal point estimators of τ_ν, under appropriate regularity conditions. For the fuzzy RD cases, the bandwidth choices employed are technically optimal only for the numerator of the estimators, but since the rate of the MSE-optimal bandwidth choice does not differ from the sharp RD case, the estimators $\hat{\varsigma}_{\nu,p}(\hat{h}_{\text{IK},\nu,p})$, $\hat{\varsigma}_{\nu,p}(\hat{h}_{\text{CCT},\nu,p})$ and $\hat{\varsigma}_{\nu,p}(\hat{h}_{\text{CV},p})$ may also be viewed as consistent and (asymptotically) MSE-optimal point estimators of ς_ν, under appropriate regularity conditions.

Sharp RD Conventional Confidence Intervals

In the sharp RD design, under appropriate regularity conditions and rate-restrictions on the bandwidth sequence $h_n \to 0$, conventional confidence intervals accompanying the point estimators discussed above rely on the following distributional approximation:

$$
\sqrt{nh_n^{1+2\nu}}\left(\hat{\tau}_{\nu,p}(h_n) - \tau_\nu - h_n^{p+1-\nu}\,\mathsf{B}_{\nu,p}\right) \to_{\text{d}} \mathcal{N}(0,\mathsf{V}_{\nu,p}), \qquad \nu \leq p, \tag{2}
$$

where $\mathsf{B}_{\nu,p}$ and $\mathsf{V}_{\nu,p}$ are given in Equation (1). Therefore, an infeasible, asymptotic $100(1-\alpha)$-percent confidence interval for τ_ν is

$$
\mathsf{CI}_{1-\alpha}(h_n) = \left[\left(\hat{\tau}_{\nu,p}(h_n) - h_n^{p+1-\nu}\,\mathsf{B}_{\nu,p}\right) \pm \Phi_{1-\alpha/2}^{-1}\sqrt{\frac{\mathsf{V}_{\nu,p}}{nh_n^{1+2\nu}}}\,\right],
$$

where Φ_a^{-1} denotes the appropriate quantile of the Gaussian distribution (e.g., 1.96 for $a =$ .025). To implement this confidence interval in practice, we need to handle the leading bias $(\mathsf{B}_{\nu,p})$ and the variance $(\mathsf{V}_{\nu,p})$ of the RD estimator, because they involve unknown quantities.

Bias. In practice, the conventional approach to handling the smoothing bias is to rely on an "undersmoothing" argument: that is, to choose a "small" enough bandwidth so that the bias is negligible. Theoretically, this approach requires selecting a bandwidth sequence $h_n \to 0$ such that $\sqrt{nh_n^{1+2\nu}}h_n^{p+1-\nu}\mathsf{B}_{\nu,p} = o_p(1)$. In practice, however, this procedure may be difficult to

implement because most bandwidth selectors, such as $h_{\mathtt{MSE},\nu,p}$, will not satisfy the conditions required for undersmoothing. This fact implies that most empirical bandwidth selectors could in principle lead to a non-negligible leading bias in the distributional approximation, which in turn will bias the associated confidence intervals. Simulation evidence highlighting this potential drawback of the undersmoothing/small-bias approach is provided in CCT. Nonetheless, in applications, it is common for researchers to simply ignore the leading bias, proceeding as if $\mathsf{B}_{\nu,p} \approx 0$. This approach is justified by either assuming the bias is "small," or by shrinking the bandwidth choice by some ad-hoc factor (i.e., undersmoothing).

Standard Errors. The asymptotic variance is handled by replacing $\mathsf{V}_{\nu,p}$ with a consistent estimator. A natural approach is to construct a plug-in estimator based on the conditional (on $\mathbf{X}_n$) variance of $\hat{\tau}_{\nu,p}(h_n)$:

$$\mathsf{V}_{n,\nu,p} := n h_n^{1+2\nu} \mathbb{V}[\hat{\tau}_{\nu,p}(h_n)|\mathbf{X}_n] = \mathsf{V}_{+,n,\nu,p} + \mathsf{V}_{-,n,\nu,p}, \qquad \nu \leq p,$$

with

$$\mathsf{V}_{+,n,\nu,p} = h_n \mathbf{e}_\nu' \mathbf{\Gamma}_{+,n,p}^{-1} \mathbf{X}_{+,n,p} \mathbf{W}_{+,n,p} \mathbf{\Sigma} \mathbf{W}_{+,n,p} \mathbf{X}_{+,n,p} \mathbf{\Gamma}_{+,n,p}^{-1} \mathbf{e}_\nu,$$

$$\mathsf{V}_{-,n,\nu,p} = h_n \mathbf{e}_\nu' \mathbf{\Gamma}_{-,n,p}^{-1} \mathbf{X}_{-,n,p} \mathbf{W}_{-,n,p} \mathbf{\Sigma} \mathbf{W}_{-,n,p} \mathbf{X}_{-,n,p} \mathbf{\Gamma}_{-,n,p}^{-1} \mathbf{e}_\nu,$$

where

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma^2(X_1) & 0 & \cdots & 0 \\ 0 & \sigma^2(X_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \sigma^2(X_n) \end{bmatrix} = \mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{X}_n], \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \quad \varepsilon_i = Y_i - \mathbb{E}[Y_i|X_i],$$

and the exact form of the other matrices is given in CCT. Importantly, the only matrix including unknown quantities is $\mathbf{\Sigma}$. It is easy to show that $\mathsf{V}_{n,\nu,p} \to_{\mathrm{p}} \mathsf{V}_{\nu,p}$.

The variance $\mathsf{V}_{n,\nu,p}$ has a "sandwich" structure coming from the weighted least-squares structure of local polynomials. It looks exactly like the usual heteroskedasticity-robust standard-error formula in linear-regression models, and hence constructing standard errors requires only an estimator of $\mathbf{\Sigma}$. Thus, to construct valid standard errors we consider two types of plug-in "estimators" of $\sigma^2(X_i) = \mathbb{E}[\varepsilon_i^2|X_i]$, for control and treatment units separately: "plug-in estimated residuals" and "fixed-matches estimated residuals". Both approaches construct an estimator of $\mathsf{V}_{n,\nu,p}$ by removing the conditional expectation in $\mathbf{\Sigma}$ and replacing ε_i by some estimator of it.

- *Plug-in Estimated Residuals*. In this approach, ε_i's are replaced by

$$\check{\varepsilon}_{+,i} = Y_i - \hat{\mu}_{+,p}^{(0)}(X_i; c_n) \qquad \text{and} \qquad \check{\varepsilon}_{-,i} = Y_i - \hat{\mu}_{-,p}^{(0)}(X_i; c_n),$$

for treated and control units, respectively. In practice, the bandwidth employed, c_n, is typically set to $c_n = h_n$, although in general it could be different. In fact, the choice $c_n = h_n$ may not be optimal and could lead to poor finite-sample performance of the standard errors estimator. This approach leads to the Huber-Eicker-White standard errors estimator, which is robust to heteroskedasticity of unknown form. We denote this estimator by

$$\check{\mathsf{V}}_{n,\nu,p} = \check{\mathsf{V}}_{+,n,\nu,p} + \check{\mathsf{V}}_{-,n,\nu,p},$$

where $\check{\mathsf{V}}_{+,n,\nu,p}$ and $\check{\mathsf{V}}_{-,n,\nu,p}$ employ, respectively, $\check{\varepsilon}_{+,i}$ and $\check{\varepsilon}_{-,i}$ in their construction.

- *Fixed-matches Estimated Residuals.* CCT propose an alternative standard errors estimator employing a different construction for the residuals, motivated by the work of Abadie and Imbens (2006). This estimator is constructed using a simple nearest-neighbor (or fixed-matches) estimator for the residuals, which are in general unbiased but inconsistent. We denote these statistics by $\hat{\varepsilon}_{+,i}$ and $\hat{\varepsilon}_{-,i}$. After plugging them into Σ, the resulting standard errors estimator is denoted by

$$\hat{\mathsf{V}}_{n,\nu,p} = \hat{\mathsf{V}}_{+,n,\nu,p} + \hat{\mathsf{V}}_{-,n,\nu,p},$$

  where $\hat{\mathsf{V}}_{+,n,\nu,p}$ and $\hat{\mathsf{V}}_{-,n,\nu,p}$ employ, respectively, the fixed-matches estimators $\hat{\varepsilon}_{+,i}$ and $\hat{\varepsilon}_{-,i}$ in their construction. These estimators are shown to be consistent under appropriate regularity conditions in CCT, although we do not provide further details here to conserve space.

Recap. Our R package **rdrobust** offers two alternative consistent estimators of $\mathsf{V}_{\nu,p}$: (i) the plug-in estimated residuals estimator $\check{\mathsf{V}}_{n,\nu,p}$ and (ii) the fixed-matches estimator $\hat{\mathsf{V}}_{n,\nu,p}$. These estimators may be used for both standard error constructions and plug-in bandwidth selectors, as discussed above. As a general rule for implementation, we always employ the same estimated bandwidth used in the treatment effect estimator $\hat{\tau}_{\nu,p}(h_n)$ whenever additional bandwidth choices are required, unless explicitly noted otherwise. See CCT for all other details omitted here.

Sharp RD Bias-Corrected Confidence Intervals

As an alternative to undersmoothing, we can consider dealing with the leading bias in the distributional approximation explicitly: we can directly bias-correct the estimator by constructing an estimator of $\mathsf{B}_{\nu,p}$, which is then subtracted from the RD point estimate in an attempt to eliminate the leading bias in (2).

A simple bias estimator is constructed using a higher-order local polynomial to estimate the unknown derivatives in $\mathsf{B}_{\nu,p} \propto \mu_+^{(p+1)} - \mu_-^{(p+1)}$. For example, $\mu_+^{(p+1)}$ can be estimated by using a q-th order local polynomial ($q \geq p + 1$) with pilot bandwidth b_n, leading to the estimator $\hat{\mu}_+^{(p+1)} = e'_{p+1}\hat{\boldsymbol{\beta}}_{+,q}(b_n)$. The resulting bias-corrected estimator is

$$\hat{\tau}_{\nu,p,q}^{\mathsf{bc}}(h_n, b_n) = \hat{\tau}_{\nu,p}(h_n) - h_n^{p+1-\nu}\,\hat{\mathsf{B}}_{\nu,p,q}, \qquad \hat{\mathsf{B}}_{\nu,p,q} := \hat{\mathsf{B}}_{\nu,p,q}(h_n, b_n), \qquad \nu \leq p < q,$$

where the exact form of $\hat{\mathsf{B}}_{\nu,p,q}$ is described in CCT.

Using the bias-corrected estimator, and imposing appropriate regularity conditions and bandwidth restrictions, we obtain:

$$\sqrt{nh_n^{1+2\nu}}\left(\hat{\tau}_{\nu,p,q}^{\mathsf{bc}}(h_n, b_n) - \tau_\nu\right)$$
$$= \underbrace{\sqrt{nh_n^{1+2\nu}}\left(\hat{\tau}_{\nu,p}(h_n) - \tau_\nu - h_n^{p+1-\nu}\,\mathsf{B}_{\nu,p}\right)}_{\rightarrow_{\mathrm{d}}\,\mathcal{N}(0,\mathsf{V}_{\nu,p})} - \underbrace{\sqrt{nh_n^{1+2\nu}}h_n^{p+1-\nu}(\hat{\mathsf{B}}_{\nu,p,q} - \mathsf{B}_{\nu,p})}_{\rightarrow_{\mathrm{p}}\,0},$$

which is valid if $h_n/b_n \rightarrow 0$. This result immediately justifies bias-corrected confidence intervals, where the unknown bias in $\mathsf{CI}_{1-\alpha}(h_n)$ is replaced by the bias-estimate $\hat{\mathsf{B}}_{\nu,p,q}$. The exact

formula of the resulting confidence intervals is

$$\mathsf{CI}^{\mathsf{bc}}_{1-\alpha}(h_n, b_n) = \left[\left(\hat{\tau}_{\nu,p}(h_n) - h_n^{p+1-\nu} \, \hat{\mathsf{B}}_{\nu,p,q} \right) \pm \Phi^{-1}_{1-\alpha/2} \, \sqrt{\frac{\mathsf{V}_{\nu,p}}{n h_n^{1+2\nu}}} \right],$$

where for implementation one replaces $\mathsf{V}_{\nu,p}$ by a consistent estimator, like either of the two discussed previously.

To implement this approach, CCT discuss a choice of pilot bandwidth b_n along the lines of the MSE expansion described above. This leads to the following MSE-optimal choice of b_n for the bias-correction estimator $\hat{\mathsf{B}}_{\nu,p,q}$:

$$b_{\mathsf{MSE},p,q} = \left(\frac{(2p+3) \, \mathsf{V}_{p+1,q}}{2(q-p) \mathsf{B}^2_{p+1,q}} \right)^{1/(2q+3)} n^{-1/(2q+3)},$$

where $\mathsf{V}_{p,q}$ and $\mathsf{B}_{p,q}$ are the corresponding leading variance and bias terms arising from the MSE expansion. (Note that this choice is not necessarily optimal for $\hat{\tau}^{\mathsf{bc}}_{\nu,p,q}(h_n, b_n)$.) CCT also discuss an implementation procedure of $b_{\mathsf{MSE},p,q}$, leading to the data-driven estimator:

$$\hat{b}_{\mathsf{CCT},p,q} = \left(\frac{(2p+3)\hat{\mathsf{V}}_{\mathsf{CCT},p+1,q}}{2(q-p)\hat{\mathsf{B}}^2_{\mathsf{CCT},p+1,q} + \hat{\mathsf{R}}_{\mathsf{CCT},p+1,q}} \right)^{1/(2q+3)} n^{-1/(2q+3)},$$

where the exact form of the estimators $\hat{\mathsf{V}}_{\mathsf{CCT},p,q}$, $\hat{\mathsf{B}}_{\mathsf{CCT},p,q}$ and $\hat{\mathsf{R}}_{\mathsf{CCT},p,q}$ is given in CCT. We employ this pilot bandwidth estimator in our default implementation, but we also implement bandwidth estimators constructed following the underlying logic in IK, denoted by $\hat{b}_{\mathsf{IK},p,q}$.

Sharp RD Robust Bias-Corrected Confidence Intervals

The confidence intervals discussed so far have some unappealing properties that may affect their performance in applications. On the one hand, the confidence intervals $\mathsf{CI}_{1-\alpha}(h_n)$ require undersmoothing (or, alternatively, a "small" bias), which may lead to coverage distortions in cases where the bias is important. On the other hand, the bias-corrected confidence intervals $\mathsf{CI}^{\mathsf{bc}}_{1-\alpha}(h_n, b_n)$, while theoretically justified for a larger range of bandwidths, are usually regarded as having poor performance in empirical settings, also leading to potentially large coverage distortions in applications. Monte Carlo evidence showing some of these potential pitfalls is reported in CCT.

CCT propose an alternative, more robust confidence interval formula based on the bias-corrected RD treatment effect estimators, but employing different standard errors. Intuitively, the bias-corrected RD estimator does not perform well in finite-samples because the bias-estimate introduces additional variability in $\hat{\tau}^{\mathsf{bc}}_{\nu,p,q}(h_n, b_n) = \hat{\tau}_{\nu,p}(h_n) - h_n^{p+1-\nu} \, \hat{\mathsf{B}}_{\nu,p,q}$, $\nu \le p$, which is not accounted for when forming the associated confidence intervals $\mathsf{CI}^{\mathsf{bc}}_{1-\alpha}(h_n, b_n)$. Thus, CCT propose an alternative asymptotic approximation for $\hat{\tau}^{\mathsf{bc}}_{p,q}(h_n, b_n)$ which, provided appropriate regularity conditions hold and $h_n/b_n \to \rho \in [0, \infty)$, may be summarized as follows:

$$\sqrt{n h_n^{1+2\nu}} \left(\hat{\tau}^{\mathsf{bc}}_{\nu,p,q}(h_n, b_n) - \tau_\nu \right)$$
$$= \underbrace{\sqrt{n h_n^{1+2\nu}} \left(\hat{\tau}_{\nu,p}(h_n) - \tau_\nu - h_n^{p+1-\nu} \, \mathsf{B}_{\nu,p} \right)}_{\to_{\mathrm{d}} \, \mathcal{N}(0,\mathsf{V}_{\nu,p})} - \underbrace{\sqrt{n h_n^{1+2\nu}} h_n^{p+1-\nu} (\hat{\mathsf{B}}_{\nu,p,q} - \mathsf{B}_{\nu,p})}_{\to_{\mathrm{d}} \, \mathcal{N}(0,\mathsf{V}_{p+1,q}(\rho))},$$

where here $\mathsf{V}_{p,q}(\rho)$ could be interpreted as the contribution of the bias-correction to the variability of the bias-corrected estimator. (It can be shown that $\mathsf{V}_{p,q}(0) = 0$.)

Under weaker conditions than those typically imposed in the results summarized in the previous subsections, CCT show that

$$\sqrt{nh_n^{1+2\nu}}\left(\hat{\tau}_{\nu,p,q}^{\mathsf{bc}}(h_n, b_n) - \tau_\nu\right) \to_{\mathrm{d}} \mathcal{N}(0, \mathsf{V}_{\nu,p,q}^{\mathsf{bc}}(\rho)),$$

where $\mathsf{V}_{\nu,p,q}^{\mathsf{bc}}(\rho)$ is the asymptotic variance for the bias-corrected estimator, which is different from the usual one, $\mathsf{V}_{\nu,p}$. Indeed, it can be shown that $\mathsf{V}_{\nu,p,q}^{\mathsf{bc}}(\rho) \to \mathsf{V}_{\nu,p}$ if $\rho \downarrow 0$, but in general $\mathsf{V}_{\nu,p,q}^{\mathsf{bc}}(\rho) > \mathsf{V}_{\nu,p}$ under standard conditions. More generally,

$$\frac{\hat{\tau}_{\nu,p,q}^{\mathsf{bc}}(h_n, b_n) - \tau_\nu}{\sqrt{\mathsf{V}_{n,\nu,p,q}^{\mathsf{bc}}}} \to_{\mathrm{d}} \mathcal{N}(0, 1), \qquad \mathsf{V}_{n,\nu,p,q}^{\mathsf{bc}} := \mathsf{V}_{n,\nu,p,q}^{\mathsf{bc}}(h_n, b_n),$$

where the exact formula for $\mathsf{V}_{n,\nu,p,q}^{\mathsf{bc}}(h_n, b_n)$ is given in CCT. Intuitively, this variance formula is constructed to account for the variability of both the original RD treatment effect estimator $(\hat{\tau}_{\nu,p}(h_n))$ and the bias-correction term $(\hat{\mathsf{B}}_{\nu,p,q})$ in the distributional approximation of the studentized statistic.

Thus, this result justifies the confidence intervals:

$$\mathsf{CI}_{1-\alpha}^{\mathsf{rbc}}(h_n, b_n) = \left[\left(\hat{\tau}_{\nu,p}(h_n) - h_n^{p+1-\nu}\,\hat{\mathsf{B}}_{\nu,p,q}\right) \pm \Phi_{1-\alpha/2}^{-1}\sqrt{\mathsf{V}_{n,\nu,p,q}^{\mathsf{bc}}}\right],$$

where for implementation $\mathsf{V}_{n,\nu,p,q}^{\mathsf{bc}}(h_n, b_n)$ is replaced by an appropriate estimator. Two possible estimators are discussed in CCT: $\check{\mathsf{V}}_{n,\nu,p,q}^{\mathsf{bc}} := \check{\mathsf{V}}_{n,\nu,p,q}^{\mathsf{bc}}(h_n, b_n)$ (using plug-in estimated residuals) and $\hat{\mathsf{V}}_{n,\nu,p,q}^{\mathsf{bc}} := \hat{\mathsf{V}}_{n,\nu,p,q}^{\mathsf{bc}}(h_n, b_n)$ (using fixed-matches estimated residuals), as discussed above for the case of conventional confidence intervals.

All omitted notational, methodological and technical details are given in CCT. Further theoretical implications of this alternative approach to nonparametric bias-correction are discussed in CCT.

Sharp RD Confidence Intervals: Summary

Our R package **rdrobust** provides the following data-driven RD treatment effect confidence intervals for sharp RD designs:

(i) Undersmoothing / Small-Bias:

- *Plug-in estimated errors*: $\check{\mathsf{CI}}_{1-\alpha}(\hat{h}_n)$ where

$$\check{\mathsf{CI}}_{1-\alpha}(h_n) = \left[\hat{\tau}_{\nu,p}(h_n) \pm \Phi_{1-\alpha/2}^{-1}\sqrt{\frac{\check{\mathsf{V}}_{\nu,p}}{nh_n^{1+2\nu}}}\right].$$

- *Fixed-matches estimated errors*: $\hat{\mathsf{CI}}_{1-\alpha}(\hat{h}_n)$ where

$$\hat{\mathsf{CI}}_{1-\alpha}(h_n) = \left[\hat{\tau}_{\nu,p}(h_n) \pm \Phi_{1-\alpha/2}^{-1}\sqrt{\frac{\hat{\mathsf{V}}_{\nu,p}}{nh_n^{1+2\nu}}}\right].$$

(ii) <u>Bias-Correction</u>:

- *Plug-in estimated errors*: $\check{\mathsf{Cl}}^{\mathtt{bc}}_{1-\alpha}(\hat{h}_n, \hat{b}_n)$, where

$$\check{\mathsf{Cl}}^{\mathtt{bc}}_{1-\alpha}(h_n, b_n) = \left[\left(\hat{\tau}_{\nu,p}(h_n) - h_n^{p+1-\nu} \, \hat{\mathsf{B}}_{\nu,p,q} \right) \pm \Phi^{-1}_{1-\alpha/2} \, \sqrt{\frac{\check{\mathsf{V}}_{\nu,p}}{nh_n^{1+2\nu}}} \right].$$

- *Fixed-matches estimated errors*: $\hat{\mathsf{Cl}}^{\mathtt{bc}}_{1-\alpha}(\hat{h}_n, \hat{b}_n)$, where

$$\hat{\mathsf{Cl}}^{\mathtt{bc}}_{1-\alpha}(h_n, b_n) = \left[\left(\hat{\tau}_{\nu,p}(h_n) - h_n^{p+1-\nu} \, \hat{\mathsf{B}}_{\nu,p,q} \right) \pm \Phi^{-1}_{1-\alpha/2} \, \sqrt{\frac{\hat{\mathsf{V}}_{\nu,p}}{nh_n^{1+2\nu}}} \right].$$

(iii) <u>Robust Bias-Correction</u>:

- *Plug-in estimated errors*: $\check{\mathsf{Cl}}^{\mathtt{rbc}}_{1-\alpha}(\hat{h}_n, \hat{b}_n)$, where

$$\check{\mathsf{Cl}}^{\mathtt{rbc}}_{1-\alpha}(h_n, b_n) = \left[\left(\hat{\tau}_{\nu,p}(h_n) - h_n^{p+1-\nu} \, \hat{\mathsf{B}}_{\nu,p,q} \right) \pm \Phi^{-1}_{1-\alpha/2} \, \sqrt{\check{\mathsf{V}}^{\mathtt{bc}}_{n,\nu,p,q}} \right].$$

- *Fixed-matches estimated errors*: $\hat{\mathsf{Cl}}^{\mathtt{rbc}}_{1-\alpha}(\hat{h}_n, \hat{b}_n)$, where

$$\hat{\mathsf{Cl}}^{\mathtt{rbc}}_{1-\alpha}(h_n, b_n) = \left[\left(\hat{\tau}_{\nu,p}(h_n) - h_n^{p+1-\nu} \, \hat{\mathsf{B}}_{\nu,p,q} \right) \pm \Phi^{-1}_{1-\alpha/2} \, \sqrt{\hat{\mathsf{V}}^{\mathtt{bc}}_{n,\nu,p,q}} \right].$$

In the above displays $\hat{h}_n \in \{\hat{h}_{\mathtt{IK},\nu,p}, \hat{h}_{\mathtt{CCT},\nu,p}, \hat{h}_{\mathtt{CV},p}\}$ and $\hat{b}_n \in \{\hat{b}_{\mathtt{IK},n}, \hat{b}_{\mathtt{CCT},n}\}$, among other possibilities. The exact formulas for the different standard errors estimators and other technical details may be found in Calonico, Cattaneo, and Farrell (2013a).

Fuzzy RD Confidence Intervals

All the ideas and results presented above extend to the case of fuzzy RD designs. We do not describe this case here to conserve space, but we do note that our R package **rdrobust** implements all the fuzzy RD estimators and confidence intervals, which follow the same logic as the ones described above for the sharp RD design. See CCT for details.

2.6. RD Plots

Exploratory data analysis plays a crucial role in empirical work in RD designs. Because of the simplicity of the research design, it is customary and advisable to present the main features of the design graphically. In this section, we briefly describe the main results implemented in our function `rdbinselect`, which selects tuning parameters optimally and produces commonly used RD plots.

The main graph commonly used in RD applications presents global estimates of the regression function, $\mu_{Y-}(x)$ and $\mu_{Y+}(x)$, in an attempt to describe their shapes for control ($X_i < \bar{x}$) and treated ($X_i \geq \bar{x}$) units relative to some summary of the raw data. This RD plot gives

an idea of overall fit and, at the same time, exhibits graphically the sharp RD estimate. In most empirical applications, this figure is constructed using "dots" for local sample means over non-overlapping bins or partitions of the support of X, in addition to the two smooth "global" polynomial regression curve estimates for control and treatment units. The binned means are usually included to capture the behavior of the "cloud of points" and to show whether other discontinuities are likely to be present in the data. Figure 1 gives an example of this typical RD plot using the data from Lee (2008).

This figure has two main ingredients: (i) two polynomial regression curves (estimated for $X_i < \bar{x}$ and $X_i \geq \bar{x}$, separately), and (ii) sample means over non-overlapping regions of the support of the running variable X, for control and treatment units separately.

Constructing the first ingredient (i) of the RD plot is easy, as it requires only estimating a polynomial regression on the data. Typical choices encountered in the literature are 4th and 5th order polynomials, but our implementation allows any polynomial order. In some applications a restricted support of the running variable is used, instead of the full support of the data.

The second ingredient (ii) of the RD plot requires computing sample means over non-overlapping regions of the support of the running variable X, which are meant to provide an approximation of the population regression functions as well while also giving a sense of the dispersion of the data around them. This graphical device could also be used to highlight other potential discontinuities in the data (as a form of a visual falsification test). For implementation, one needs to decide the number of bins that will be used, denoted by $J_{-,n}$ (control) and $J_{+,n}$ (treatment), and the length of each bin. Typically, the partition is chosen to be evenly-spaced, thereby reducing the problem to two tuning parameters to be selected: $J_{-,n}$ and $J_{+,n}$.

As discussed in Calonico, Cattaneo, and Titiunik (2013b, Section 2.7), it turns out that this approach (ii) can be viewed as a simple version of a nonparametric partitioning estimator; see, e.g., Cattaneo and Farrell (2013) and references therein. This formalization, and the asymptotic (integrated) MSE expansion given in Cattaneo and Farrell (2013, Theorem 3), may be used to derive an optimal choice for $J_{-,n}$ and $J_{+,n}$. These choices take the form

$$J_{-,n}^* = \left\langle (\mathscr{C}_- \, n)^{1/3} \right\rangle, \qquad \mathscr{C}_- = \frac{2\,\mathscr{B}_-}{\mathscr{V}_-}, \qquad J_{+,n}^* = \left\langle (\mathscr{C}_+ \, n)^{1/3} \right\rangle, \qquad \mathscr{C}_+ = \frac{2\,\mathscr{B}_+}{\mathscr{V}_+},$$

where $\langle \cdot \rangle$ denotes the nearest integer, and assuming that $x_l \leq X_i < \bar{x}$ (control units) and $\bar{x} \leq X_i \leq x_u$ (treatment units),

$$\mathscr{V}_- = \frac{1}{\bar{x} - x_l} \int_{x_l}^{\bar{x}} \frac{\sigma_-^2(x)}{f(x)} \, w(x) \, \mathrm{d}x, \qquad \mathscr{B}_- = \frac{(\bar{x} - x_l)^2}{12} \int_{x_l}^{\bar{x}} \left(\mu_-^{(1)}(x) \right)^2 \, w(x) \, \mathrm{d}x,$$

$$\mathscr{V}_+ = \frac{1}{x_u - \bar{x}} \int_{\bar{x}}^{x_u} \frac{\sigma_+^2(x)}{f(x)} \, w(x) \, \mathrm{d}x, \qquad \mathscr{B}_+ = \frac{(x_u - \bar{x})^2}{12} \int_{\bar{x}}^{x_u} \left(\mu_+^{(1)}(x) \right)^2 \, w(x) \, \mathrm{d}x,$$

and where $w(x)$ is a weighting function. A feasible plug-in rule can be easily constructed by using preliminary estimators for the unknown objects in $\mathscr{C}_-$ and $\mathscr{C}_+$. In our implementation, we set $w(x) = f(x)$ to simplify the constants and propose simple polynomial-based approximations to the unknown quantities, which we then estimate in a preliminary step to construct data-driven optimal bin-length choices:

$$\hat{J}_{-,n} = \left\langle (\hat{\mathscr{C}}_- \, n)^{1/3} \right\rangle, \qquad \hat{\mathscr{C}}_- = \frac{2\,\hat{\mathscr{B}}_-}{\hat{\mathscr{V}}_-}, \qquad \hat{J}_{+,n} = \left\langle (\hat{\mathscr{C}}_+ \, n)^{1/3} \right\rangle, \qquad \hat{\mathscr{C}}_+ = \frac{2\,\hat{\mathscr{B}}_+}{\hat{\mathscr{V}}_+}.$$

The function `rdbinselect` implements these bin-length selectors and also gives all the other ingredients required to construct the RD plots discussed in this section. Further details on implementation are presented in Calonico, Cattaneo, and Titiunik (2013b, Section 2.7).

3. The rdrobust Package

To illustrate some of the features of our R package **rdrobust**, we employ an extract of the dataset constructed by Cattaneo, Frandsen, and Titiunik (2013). This dataset contains information on elections for the U.S. Senate during the period 1914–2010. We focus here on the RD effect of the Democratic party winning a U.S. Senate seat on the vote share obtained in the following election for that same seat, mimicking the analysis conducted in Lee (2008) for the U.S. House.

The dataset `rdrobust_RDsenate.dta` contains two variables: *vote* and *margin*. The variable *vote* records the state-level vote share of the Democratic party in a given election for a Senate seat, while the variable *margin* records the margin of victory of the Democratic party in the previous election for the same Senate seat (i.e., six years prior).

First, we load the database and present basic summary statistics.

```
> library(rdrobust)
> senate =  load("rdrobust_RDsenate.rda")
> vote=rdrobust_RDsenate$vote
> margin=rdrobust_RDsenate$margin

> summary(vote)

  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  0.00   42.67   50.55   52.67   61.35  100.00      93

> summary(margin)

    Min.  1st Qu.   Median    Mean 3rd Qu.     Max.
-100.000  -12.210    2.166    7.171  22.770  100.000
```

This data set has a total of 1297 complete observations. The variable *margin* ranges from -100 to 100, and records the margin of victory in a given election for a given U.S. Senate seat, defined as the vote share of the Democratic party minus the vote share of the strongest opponent. When *margin* is above zero, the Democratic party wins the election for that seat, otherwise it looses. The variable *vote* ranges from 0 to 100 because it records the outcome of the (two-periods ahead) election for that given seat. Thus, observations for years 2008 and 2010 have missing *vote*. As it is usual in the literature, we exploit the discontinuity in incumbency status that occurs at 0 on *margin* to employ an RD design.

To gain further intuition on the available data, we use `rdbinselect` to construct an automatic plot of the RD design.

```
> rdbinselect(y=vote, x=margin, title="RD Plot - Senate Elections Data",
+             y.label="Vote Share in Election at time t+1",
+             x.label="Vote Share in Election at time t")
```

	Left	Right
Number of Obs.	595.00000	702.00000
Number of Bins	8.00000	8.00000
Number of Bins Orig.	8.00000	8.00000
Bin Length	12.49013	12.49554
Bin Length Orig.	12.49013	12.49554

RD Plot – Senate Elections Data
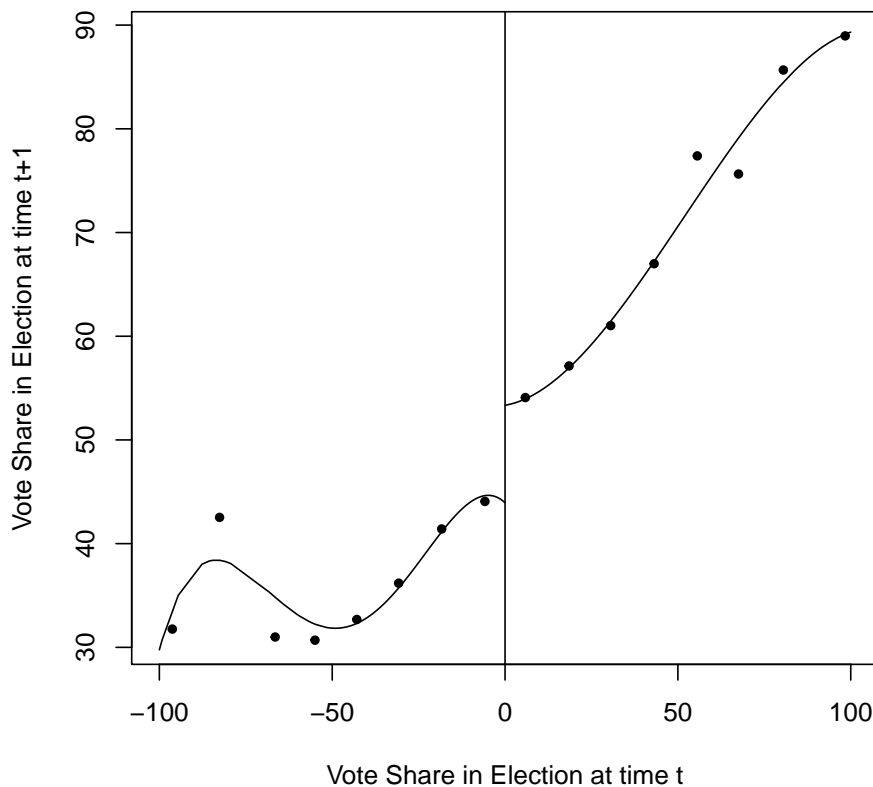
Figure 2 is constructed using the default options in the command `rdbinselect`. Using the notation introduced above, the number of optimal bins for control and treatment units are $\hat{J}_{-,n} = 8$ and $\hat{J}_{+,n} = 8$, respectively, implying a bin length of roughly 12 percentage points. The global polynomial is constructed using a $4 - th$ degree polynomial ($p = 4$ for $\hat{\mu}_{-.p,1}(x)$ and $\hat{\mu}_{+.p,1}(x)$). Figure 2 shows the resulting RD plot. The default bin choices are explicitly constructed to approximate the underlying regression function. As this figure shows, the local, binned sample means indeed seem to approximate well the underlying regression function (taking the global polynomial fit as benchmark).

While providing a natural starting point, the default number of bins will usually be too small in applications. This happens because the optimal formulas seek to balance square-bias and variance in order to approximate the underlying regression function globally. To obtain a visual "cloud of points" we need to increase the number of bins, that is, to undersmooth the estimator. In other words, in order to increase the overall variability of the plotted points, we

may reduce the bin-length -which is done by increasing the total number of bins used. This may be easily done using the option `scale(.)` as follows:

```
> rdbinselect(y=vote, x=margin, scale=5,
+               title="RD Plot - Senate Elections Data",
+               y.label="Vote Share in Election at time t+1",
+               x.label="Vote Share in Election at time t")
```

	Left	Right
Number of Obs.	595.000000	702.000000
Number of Bins	40.000000	40.000000
Number of Bins Orig.	8.000000	8.000000
Bin Length	2.498027	2.499109
Bin Length Orig.	12.490134	12.495543

RD Plot – Senate Elections Data
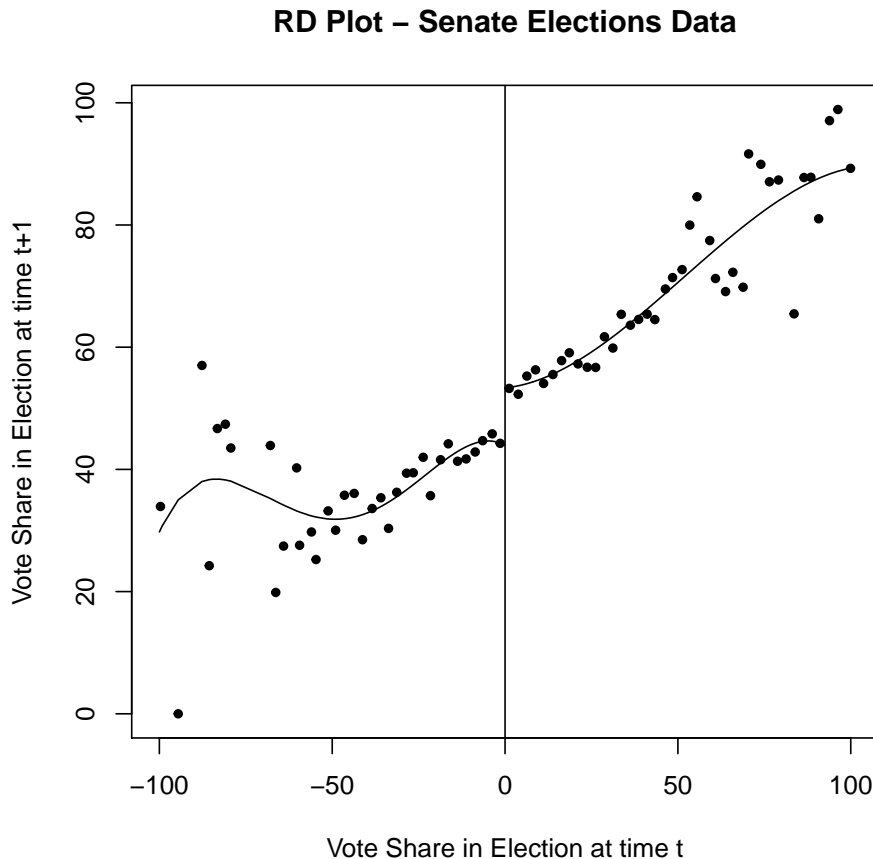

Vote Share in Election at time t

Figure 3 shows the resulting (undersmoothed) RD plot, where now the number of bins used is five times larger than the optimal choice in an integrated mean square error sense. The resulting estimator is naturally more variable than before.

Next, we conduct fully data-driven RD treatment effect estimation and inference. The command `rdrobust` using its default options leads to the following output:

```
> rdrobust(y=vote, x=margin)

$Details
          Value
Cutoff   0.0000000
N_l    332.0000000
N_r    298.0000000
p        1.0000000
q        2.0000000
h       16.0073008
b       26.0502314
rho      0.6144783

$AddDetails
          Value
BW Type    "CCT"
Kernel Type "Triangular"

$RDresults
                 Coef. Std. Err.        z       P>|z| 95% CI Left 95% CI Right
Conventional 7.429218  1.573146 4.722523 2.329362e-06    4.345909     10.51253
Robust             NA        NA       NA          NA    4.011582     11.28118
```

These results contain a variety of information, which is organized in three panels. The first two panels of the output table contain a summary of the main choices selected to construct the RD treatment effect estimators, while the lower panel includes the main estimation results. Specifically, using the notation introduced above, this table shows:

1. The total number of observations is $1,297$, with effective 332 control and 298 treated units (given the bandwidth h_n chosen; see below). The estimation is conducted using a local-linear ($p = 1$) estimator with a local-quadratic ($q = 2$) bias-correction estimate, with a triangular kernel. The standard-error estimators are the robust ones proposed by CCT, computed using 3 nearest-neighbors.

2. The bandwidth selection procedure is the one proposed by CCT, leading to $\hat{h}_{CCT,n,p} = 16.007$, ($p = 1$) and $\hat{b}_{CCT,n,q} = 26.050$, ($q = 2$)

3. The point estimator, robust standard-errors and robust confidence intervals are: $\hat{\tau}_p(\hat{h}_{CCT,n,p}) = 7.429$, $\hat{V}^{bc}_{n,q,p}(\hat{h}_{CCT,n,p}, \hat{b}_{CCT,n,q}) = (1.573)^2$, and $\hat{CI}^{rbc}_{1-\alpha}(\hat{h}_{CCT,n,p}, \hat{b}_{CCT,n,q}) = [\, 4.011 \,,\, 11.281 \,]$ ($\alpha = 0.05$).

The command `rdrobust` also offers a more detailed output, which includes all the point estimators, standard-errors estimators and confidence intervals discussed in Section 2. These results are retrieved by including the option `all`. The corresponding output is as follows:

```
> rdrobust(y=vote, x=margin, all=TRUE)
```

```
$Details
           Value
Cutoff   0.0000000
N_l    332.0000000
N_r    298.0000000
p        1.0000000
q        2.0000000
h       16.0073008
b       26.0502314
rho      0.6144783


$AddDetails
          Value
BW Type      "CCT"
Kernel Type "Triangular"


$RDresults
                  Coef. Std. Err.        z       P>|z| 95% CI Left
Conventional    7.429218  1.573146 4.722523 2.329362e-06    4.345909
Bias Correction 7.646383  1.573146 4.860569 1.170491e-06    4.563074
Robust          7.646383  1.854524 4.123097 3.738128e-05    4.011582
                95% CI Right
Conventional       10.51253
Bias Correction    10.72969
Robust             11.28118
```

Finally, we explore all the bandwidth selection procedures contained in our package. Specifically, we may employ our companion package `rdbwselect` to compare the CCT bandwidth selectors with the IK and CV approaches. We have:

```
> rdbwselect(y=vote, x=margin, all=TRUE)


          h        b
CCT 16.00730 26.05023
IK  19.11819 18.13554
CV  35.42113       NA
```

In this case we employed the option `all`, which computes the three bandwidth selectors briefly discussed above. Notice that the option `CV` is currently not available for derivative estimation. To further understand the performance of the CV approach, we include a graph of the CV objective function over the grid being considered. This is done using the option `cvplot` as shown next (in this example we also changed the grid features to obtain a better plot, and to show this additional functionality in action as well).

```
> rdbwselect(y=vote, x=margin, bwselect="CV",
+            cvgrid_min=10, cvgrid_max=80, cvplot=TRUE)
```

```
         h  b
CCT    NA NA
IK     NA NA
CV   34.5 NA
```

Cross−Validation Objective Function
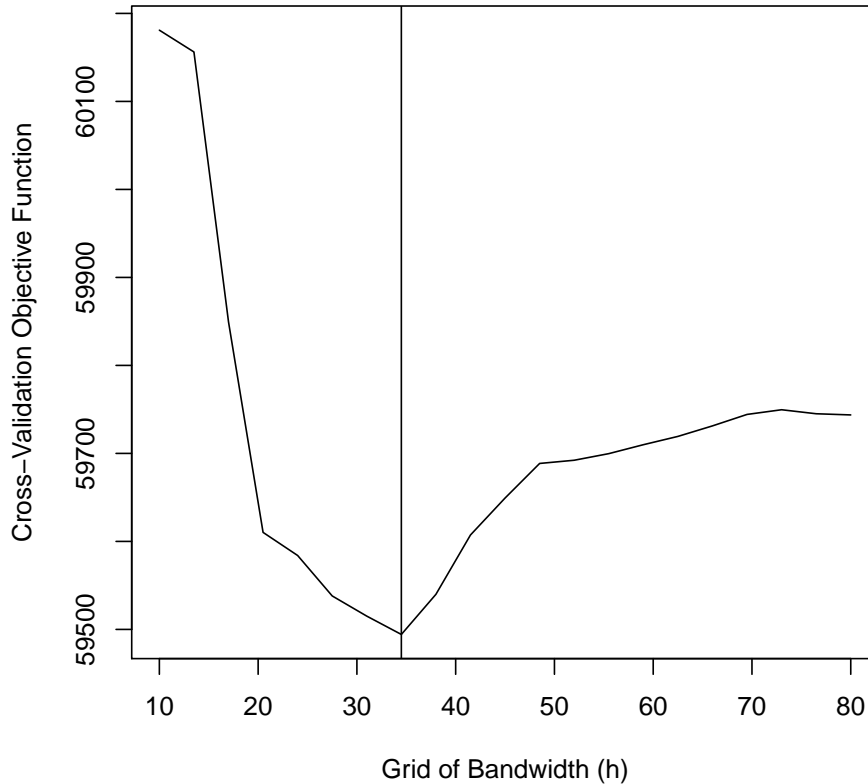
As discussed above, our commands have many other options. For example, for the main command rdrobust we have the following additional examples (output is not provided to conserve space):

1. `rdrobust`(y= *vote*, x= *margin*, `kernel`= *"uniform"*)
Estimation using uniform kernel.

2. `rdrobust`(y= *vote*, x= *margin*, `bwselect`= *"IK"*)
Estimation using the IK bandwidth selectors.

3. `rdrobust`(y= *vote*, x= *margin*, `bwselect`= *"CV"*)
Estimation using the CV bandwidth selector

4. `rdrobust`(y= *vote*, x= *margin*, `h`= 15, `rho`= 0.8)

Estimation using

5. `rdrobust`(y= *vote*, x= *margin*, `p`= 2, `q`= 4)

Estimation using $p = 2$ and $q = 4$.

6. `rdrobust`(y= *vote*, x= *margin*, `vce`= *"resid"*)

Estimation using plug-in residuals estimates in the VCE.

Finally, our commands may also be used to conduct inference in other RD design settings. For example, assuming y is the output variable, t is the treatment status variable, and x is the running variable:

1. `rdrobust`(y, x, `deriv`= 1)

Estimation for sharp kink RD.

2. `rdrobust`(y, x, `fuzzy`= t)

Estimation for fuzzy RD.

3. `rdrobust`(y, x, `fuzzy`= t, `deriv`= 1)

Estimation for fuzzy kink RD.

4. Conclusions

We introduced and discussed the main features of the R package **rdrobust**, which provides several functions to conduct data-driven, local-polynomial based (robust) inference in RD designs. In particular, this package includes three main functions (`rdrobus`, `rdbwselect` and `rdbinselect`), which together offer an array of data-driven nonparametric inference methods useful to perform empirical work in the context of RD applications. This implementation covers average RD treatment effects at the cutoff in the sharp RD, sharp kink RD, fuzzy RD and fuzzy kink RD designs, among other possibilities.

A full description of the R package **rdrobust** capabilities may be found in the help files. A companion Stata package offering the same structure and capabilities is described in Calonico, Cattaneo, and Titiunik (2013b). In future versions, we plan to include other estimands (e.g., quantile RD treatment effects at the cutoff) as well as other estimation procedures (e.g., RD estimators based on generalized linear models).

5. Acknowledgments

References

Abadie A, Imbens GW (2006). "Large Sample Properties of Matching Estimators for Average Treatment Effects." *Econometrica*, **74**(1), 235–267.

Calonico S, Cattaneo MD, Farrell MH (2013a). "On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Estimation." Working paper, University of Michigan.

Calonico S, Cattaneo MD, Titiunik R (2013b). "Robust Data-Driven Inference in the Regression-Discontinuity Design." Revision requested by *Stata Journal*.

Calonico S, Cattaneo MD, Titiunik R (2013c). "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." Revision requested by *Econometrica*.

Calonico S, Cattaneo MD, Titiunik R (2013d). "Supplement to 'Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs'." Supplemental Material, University of Michigan.

Card D, Lee DS, Pei Z, Weber A (2012). "Nonlinear Policy Rules and the Identification and Estimation of Causal Effects in a Generalized Regression Kink Design." *Working paper, University of California at Berkeley*.

Cattaneo MD, Farrell MH (2013). "Optimal Convergence Rates, Bahadur Representation, and Asymptotic Normality of Partitioning Estimators." *Journal of Econometrics*, **174**(2), 127–143.

Cattaneo MD, Frandsen B, Titiunik R (2013). "Randomization Inference in the Regression Discontinuity Design: An Application to Party Advantages in the U.S. Senate." Working paper, University of Michigan.

Cheng MY, Fan J, Marron JS (1997). "On Automatic Boundary Corrections." *Annals of Statistics*, **25**(4), 1691–1708.

Cook TD (2008). ""Waiting for Life to Arrive": A History of the Regression-Discontinuity Design in Psychology, Statistics and Economics." *Journal of Econometrics*, **142**(2), 636–654.

Dinardo J, Lee DS (2011). "Program Evaluation and Research Designs." In O Ashenfelter, D Card (eds.), *Handbook of Labor Economics*, volume 4A, pp. 463–536. Elsevier Science B.V.

Dong Y (2012). "Jumpy or Kinky? Regression Discontinuity Without The Discontinuity." *Working paper, University of California at Irvine*.

Dong Y, Lewbel A (2012). "Regression Discontinuity Marginal Threshold Treatment Effects." *Working paper, Boston College*.

Fan J, Gijbels I (1996). *Local Polynomial Modelling and Its Applications*. Chapman & Hall/CRC, New York.

Hahn J, Todd P, van der Klaauw W (2001). "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica*, **69**(1), 201–209.

Heckman JJ, Vytlacil EJ (2007). "Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation." In J Heckman, E Leamer (eds.), *Handbook of Econometrics, vol. VI*, pp. 4780–4874. Elsevier Science B.V.

Imbens G, Lemieux T (2008). "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics*, **142**(2), 615–635.

Imbens GW, Kalyanaraman K (2012). "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *Review of Economic Studies*, **79**(3), 933–959.

Imbens GW, Wooldridge JM (2009). "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature*, **47**(1), 5–86.

Lee DS (2008). "Randomized Experiments from Non-random Selection in U.S. House Elections." *Journal of Econometrics*, **142**(2), 675–697.

Lee DS, Lemieux T (2010). "Regression Discontinuity Designs in Economics." *Journal of Economic Literature*, **48**(2), 281–355.

Porter J (2003). "Estimation in the Regression Discontinuity Model." *Working paper, University of Wisconsin.*

van der Klaauw W (2008). "Regression-Discontinuity Analysis: A Survey of Recent Developments in Economics." *Labour*, **22**(2), 219–245.

Wand M, Jones M (1995). *Kernel Smoothing.* Chapman & Hall/CRC, Florida.

Affiliation:

Sebastian Calonico
Department of Economics
University of Michigan
Ann Arbor, Michigan, USA
E-mail: calonico@umich.edu
URL: http://www.umich.edu/~calonico

Matias D. Cattaneo
Department of Economics
University of Michigan
Ann Arbor, Michigan, USA
E-mail: cattaneo@umich.edu
URL: http://www.umich.edu/~cattaneo

Rocio Titiunik
Department of Political Science and Center for Political Studies
University of Michigan
Ann Arbor, Michigan, USA
E-mail: titiunik@umich.edu
URL: http://www.umich.edu/~titiunik