Regression Discontinuity Designs[*]

Matias D. Cattaneo[†]   Rocío Titiunik[‡]

August 20, 2021

Abstract

The Regression Discontinuity (RD) design is one of the most widely used non-experimental methods for causal inference and program evaluation. Over the last two decades, statistical and econometric methods for RD analysis have expanded and matured, and there is now a large number of methodological results for RD identification, estimation, inference, and validation. We offer a curated review of this methodological literature organized around the two most popular frameworks for the analysis and interpretation of RD designs: the continuity framework and the local randomization framework. For each framework, we discuss three main areas: (i) designs and parameters, which focuses on different types of RD settings and treatment effects of interest; (ii) estimation and inference, which presents the most popular methods based on local polynomial regression and analysis of experiments, as well as refinements, extensions and other methods; and (iii) validation and falsification, which summarizes an array of mostly empirical approaches to support the validity of RD designs in practice.

Keywords: regression discontinuity, causal inference, program evaluation, treatment effects, non-experimental methods.

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

[‡]Department of Politics, Princeton University.

Contents

1 Introduction

The regression discontinuity (RD) design was introduced by Thistlethwaite and Campbell (1960) as a "method of testing causal hypotheses" (p. 317) in settings where the random assignment of treatment is unavailable. The design is applicable in situations where units receive a score, and a treatment is assigned on the basis of a specific rule that links the score with a known cutoff: all units whose score is above the cutoff are assigned to the treatment condition, while all units whose score is below the cutoff are assigned to control condition. Under appropriate assumptions, the discontinuous change in the probability of receiving treatment at the cutoff can be used to learn about causal effects of the treatment on an outcome of interest.

While overlooked at first, the RD design has become one of the most credible non-experimental methods for causal inference and program evaluation. We offer a curated review of the methodological literature on the analysis and interpretation of RD designs, providing an update over prior review pieces in economics (van der Klaauw, 2008; Cook, 2008; Imbens and Lemieux, 2008; Lee and Lemieux, 2010). Our discussion aims to be comprehensive in terms of conceptual and methodological aspects but omits technical details, which are available in the specific references we give for each topic. We also discuss general ideas about implementation, but for in-depth companion methodological guides to practice see Cattaneo, Idrobo, and Titiunik (2020, 2021).

We organize our review in three main parts. Section 2 introduces the types of RD designs most commonly encountered in practice, and it also briefly discusses other RD-like research designs. This section focuses on identifiability (Matzkin, 2013) of the main treatment effects of interest in each setting, and therefore relies on basic restrictions on the underlying data generating process— the statistical model assumed to generate the data. Using the taxonomy proposed by Cattaneo, Titiunik, and Vazquez-Bare (2017), we consider two distinct conceptual frameworks for RD designs: the *continuity* framework, which relies on limiting arguments via local regression ideas, and the *local randomization* framework, which relies on ideas from the analysis of experiments. Using both frameworks, we provide an overview of canonical RD settings (sharp and fuzzy), multi-dimensional RD designs (multi-cutoff, multi-score, geographic, multiple treatments, and time-varying designs), and other related designs such as kink, bunching, before-and-after, and threshold regression designs. We also discuss issues of internal vs. external validity in the context of RD designs, and overview

methods for extrapolation of RD treatment effects.

In the second part, Section 3, we focus on estimation and inference methods for RD designs, presenting the two most common approaches: *local polynomial regression* methods (Fan and Gijbels, 1996), which are naturally motivated by the continuity framework, and *analysis of experiments* methods (Rosenbaum, 2010; Imbens and Rubin, 2015; Abadie and Cattaneo, 2018), which are naturally motivated by the local randomization framework. For completeness, we also summarize refinements, extensions, and other methods that have been proposed in the literature over the years. Building on the two canonical estimation and inference methods for RD designs, we overview the applicability of those methods when the score has discrete support, with either few or many distinct values. To close Section 3, we discuss power calculations for experimental design in RD settings, a topic that has received renewed attention among practitioners.

In the third and final part, Section 4, we focus on methods for validation and falsification of RD designs. These empirical methods are used to provide indirect evidence in favor of the underlying assumptions commonly invoked for RD identification, estimation, and inference. Our discussion includes generic program evaluation methods such as tests based on pre-intervention covariates and placebo outcomes, as well as RD-specific methods such as binomial and density continuity tests, placebo cutoff analysis, donut hole strategies, and bandwidth sensitivity.

We conclude in Section 5. General purpose software in `Python`, `R`, and `Stata` for analysis of RD designs, replication files for RD empirical illustrations, further readings, and other resources are available at https://rdpackages.github.io/.

2   Designs and Parameters

The RD design is defined by three key components—a score, a cutoff, and a discontinuous treatment assignment rule—related to each other as follows: all the units in the study are assigned a value of the score (also called a running variable or index), and the treatment is assigned to units whose score value exceeds a known cutoff (also called threshold), while not assigned to those units whose score value is below the cutoff. The distinctive feature of the RD design is that the probability of treatment assignment changes from zero to one at the cutoff, and this abrupt change in the probability of being assigned to treatment induces a (possibly smaller) abrupt change in the probability

of receiving treatment at the cutoff. Under formal assumptions that guarantee the "comparability" of treatment and control units at or near the cutoff, the discontinuous change in the probability of receiving treatment can be used to learn about causal effects of the treatment on outcomes of interest for units with scores at or near the cutoff, because units with scores barely below the cutoff can be used as a comparison (or control) group for units with scores barely above it. The most important threat to any RD design is the possibility that units are able to strategically and precisely change their score to be assigned to their preferred treatment condition (Lee, 2008; McCrary, 2008), which might induce a discontinuous change in their observable and/or unobservable characteristics at or near the cutoff and thus confound causal conclusions.

Ludwig and Miller (2007) offer a prototypical empirical application of early RD methodology with a study of the effect of expanding the social services program Head Start on child mortality. In 1965, all U.S. counties whose poverty was above the poverty rate of the 300th poorest county received grant-writing assistance to solicit Head Start funding, while counties below this poverty level did not. The increase in assistance led to an increase in funding for Head Start social services. The authors used an RD design to compare counties barely above and barely below the cutoff to estimate the effect of the increased assistance and funding on child mortality from causes that might be affected by the bundle of Head Start services. In this design, the unit of analysis is the county, the score is the county's poverty rate in 1960, and the cutoff is the poverty rate of the 300th poorest county (which was 59.1984). Cattaneo, Titiunik, and Vazquez-Bare (2017) re-analyze this application using the modern methods discussed in Sections 3 and 4. See Busse, Silva-Risso, and Zettelmeyer (2006) for another early application of the RD design in industrial organization.

Unlike other non-experimental methods, RD methodology is only applicable in situations where the treatment is assigned based on whether a score exceeds a cutoff (or where some generalization of the RD rule is employed, as we discuss below). These design-specific features make the RD strategy stand out among observational methods: to study causal RD treatment effects, the score, cutoff and treatment assignment rule must exist and be well-defined, with a discontinuous probability of treatment assignment at the cutoff. Importantly, the treatment assignment rule is known and verifiable, and not just postulated or hypothesized by the researcher. These empirically verifiable characteristics give the RD design an objective basis for implementation and validation that is usually lacking in other non-experimental strategies, and thus endow it with superior credibility

among non-experimental methods. In particular, the RD treatment assignment rule is the basis for a collection of validation and falsification tests (Section 4) that can offer empirical support for the RD assumptions and increase the credibility of RD applications. Furthermore, best practices for estimation and inference in RD designs (Section 3) also exploit these design-specific features.

The remaining of this section introduces canonical and extended RD designs, focusing on identifying assumptions for treatment effect parameters of interest employing the potential outcomes framework (Imbens and Rubin, 2015).

2.1 Sharp Designs

RD designs where the treatment assigned and the treatment received coincide for all units are referred to as *sharp*. Sharp RD designs can be used in settings where compliance with treatment is perfect (every unit assigned to treatment receives the treatment and no unit assigned to control receives the treatment), or in cases where compliance with treatment is imperfect (some units assigned to treatment remain untreated and/or viceversa) but the researcher is only interested on the *intention-to-treat* effect of offering treatment. To formalize, suppose that n units, indexed by $i = 1, 2, \cdots, n$, have a score variable X_i. The RD treatment assignment rule is $D_i = \mathbb{1}(X_i \geq c)$ with D_i denoting the observed treatment assignment, c denoting the RD cutoff, and $\mathbb{1}(\cdot)$ denoting the indicator function. Each unit is endowed with two potential outcomes $Y_i(0)$ and $Y_i(1)$, where $Y_i(1)$ is the outcome under the treatment (i.e., when $D_i = 1$) and $Y_i(0)$ is the outcome under control (i.e., when $D_i = 0$). The probability of treatment assignment changes abruptly at the cutoff from zero to one: $\mathbb{P}[D_i = 1 | X_i = x] = 0$ if $x < c$ and $\mathbb{P}[D_i = 1 | X_i = x] = 1$ if $x \geq c$; this is the key distinctive feature of the RD design.

The observed data is $(Y_i, D_i, X_i)'$, $i = 1, 2, \ldots, n$, where $Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$ captures the fundamental problem of causal inference: $Y_i = Y_i(0)$ for units with $D_i = 0$ (control units, with $X_i < c$) and $Y_i = Y_i(1)$ for units with $D_i = 1$ (treated units, with $X_i \geq c$), that is, only one potential outcome is observed for each unit. Crucially, units in the control and treatment groups will not be comparable in general, as the score variable X_i often captures important (observable and unobservable) differences between them, and the RD treatment assignment rule induces a fundamental lack of common support in the score.

Hahn, Todd, and van der Klaauw (2001) introduced the continuity framework for canonical
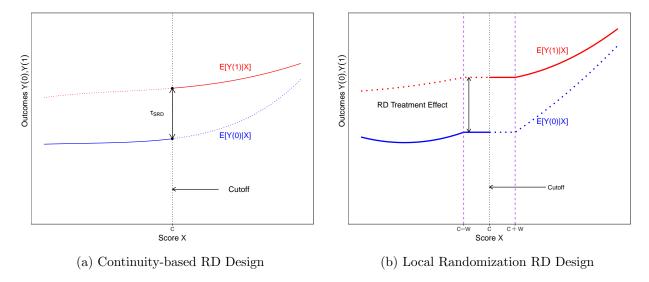
4

(a) Continuity-based RD Design   (b) Local Randomization RD Design

Figure 1: Schematic Representation of RD designs.

RD designs. In this framework, potential outcomes are taken to be random variables, with the n units of analysis forming a (random) sample from an underlying population, and the score X_i is assumed to be continuously distributed. Focusing on average treatment effects, the key identifying assumption is that the regression functions $\mathbb{E}[Y_i(0)|X_i = x]$ and $\mathbb{E}[Y_i(1)|X_i = x]$ are continuous in x at c. This assumption captures the idea that units barely "below" and "above" the cutoff c would exhibit the same average response if their treatment status did not change. Then, by implication, any difference between the average response of treated and control units at the cutoff can be attributed to the causal average effect of the treatment at the cutoff, that is, for units with score variable $X_i = c$.

Formally, within the continuity-based framework, the canonical RD treatment effect is

$$\tau_{\mathsf{SRD}} = \mathbb{E}[Y_i(1) - Y_i(0)|X = c] = \lim_{x \downarrow c} \mathbb{E}[Y_i|X_i = x] - \lim_{x \uparrow c} \mathbb{E}[Y_i|X_i = x], \qquad (1)$$

which is represented graphically in Figure (1a). The first equality in (1) is the definition of the classical Sharp RD parameter, while the second equality is the key continuity-based identification result because it links features of the distribution of potential outcomes to features of the distribution of observed outcomes, via a continuity argument. The result asserts that the sharp RD treatment effect is identifiable as the vertical distance at the cutoff between the two conditional expectations, $\lim_{x \uparrow c} \mathbb{E}[Y_i|X_i = x]$ and $\lim_{x \downarrow c} \mathbb{E}[Y_i|X_i = x]$, which can be estimated from data (Section 3.2).

5

The continuity-based identification idea for average sharp RD treatment effects has been extended in several directions. For example, Frandsen, Frolich, and Melly (2012) study quantile treatment effects, Xu (2017) investigates treatment effects using generalized nonlinear models for limited dependent outcome variables, and Hsu and Shen (2019) focus on treatment effects conditional on pre-intervention covariates. All extensions and generalizations retain the two basic sharp RD features: (i) treatment assignment is binary and applies to all units in a cross-sectional random sample, and (ii) there is a maximum discontinuity in the treatment assignment rule at the cutoff, that is, $\lim_{x \downarrow c} \mathbb{E}[D_i|X_i = x] = 1 \neq 0 = \lim_{x \uparrow c} \mathbb{E}[D_i|X_i = x]$. (Recall that $\mathbb{E}[D_i|X_i = x] = \mathbb{P}[D_i = 1|X_i = x]$.) In addition, the score variable is assumed to be continuously distributed (Section 3.5).

As a complement to the continuity framework, Cattaneo, Frandsen, and Titiunik (2015) introduced the local randomization framework for canonical RD designs, which was further developed in Cattaneo, Titiunik, and Vazquez-Bare (2017). This framework builds on the idea that "near" the cutoff the RD design can be interpreted as a randomized experiment or, more precisely, as a natural experiment (Titiunik, 2021). The heuristic idea of local randomization near the cutoff can be traced back to the original paper of Thistlethwaite and Campbell (1960), and was more recently advocated by Lee (2008) and Lee and Lemieux (2010). The local randomization framework can be formalized in various ways, building on the literature on analysis of experiments for causal inference and program evaluation (Rosenbaum, 2010; Imbens and Rubin, 2015; Abadie and Cattaneo, 2018). The original formulation of Cattaneo, Frandsen, and Titiunik (2015) adopted a Fisherian design-based approach where potential outcomes are non-random and the population is finite, implying that the only source of randomness comes from the treatment assignment mechanism near the cutoff. Cattaneo, Titiunik, and Vazquez-Bare (2017) expanded the local randomization framework to also include both the Neyman design-based approach, where the potential outcomes are non-random but sampled from an underlying infinite population, and the so-called super-population approach, where potential outcomes are random variables and their sampling uncertainty is also considered. In all three approaches, the score variable can be discrete or continuously distributed (Section 3.5).

The local randomization RD design is illustrated in Figure (1b). The key assumption in the local randomization framework is that there exists a neighborhood $\mathcal{W}$ around the cutoff such that, for all units with $X_i \in \mathcal{W}$, (i) the treatment assignment mechanism is known and independent of

potential outcomes, and (ii) the potential outcomes are independent of the score. The first condition formalizes the idea of an as-if random assignment near the cutoff, while the second condition aims to handle issues related to lack of common support via an exclusion restriction. Sekhon and Titiunik (2016, 2017) offer further conceptual discussions on the local randomization framework for RD designs.

Under the local randomization conditions in the preceding paragraph, Fisherian randomization inference is valid for testing the sharp null hypothesis of no treatment effect for units with score $X_i \in \mathcal{W}$, as well as for constructing confidence intervals and point estimators under additional assumptions. Furthermore, unbiased and/or consistent estimators of treatment effects in the neighborhood $\mathcal{W}$ can be constructed, depending on whether a Fisherian, Neyman or super-population approach is adopted. For example, for the $N_{\mathcal{W}}$ units with $X_i \in \mathcal{W}$, the average RD treatment effect in the local randomization neighborhood can be defined as

$$\tau_{\mathrm{SLR}} = \frac{1}{N_{\mathcal{W}}} \sum_{i:X_i \in \mathcal{W}} \mathbb{E}[Y_i(1) - Y_i(0)|X_i \in \mathcal{W}].$$

This notation allows for both random and non-random potential outcomes, as well as for countable or uncountable underlying populations, accommodating Fisherian, Neyman, and super-population settings. The estimand τ_{SLR} is identifiable, and can be estimated as a difference in means between control and treatment units with scores $X_i \in \mathcal{W}$, or using local regression adjustments. See Section 3.3 for details.

Researchers sometimes incorporate additional (pre-intervention) covariates to "restore" or "improve" the nonparametric identification of sharp RD parameters such as τ_{SRD} or τ_{SLR}, depending on the conceptual framework considered. However, just as in standard randomized experiments, it is not possible to restore nonparametric identification of canonical parameters of interest using covariate-adjustment without additional strong assumptions. On the other hand, covariate-adjustment can be used to improve precision in estimation and inference (Section 3). Similar considerations apply to the other RD designs discussed below. See Calonico, Cattaneo, Farrell, and Titiunik (2019) for a detailed discussion, including how to introduce covariates without changing the RD parameters, and Cattaneo, Keele, and Titiunik (2021) for an overview on covariate-adjusted RD methods.

2.2 Fuzzy Designs

Fuzzy RD designs arise in settings where treatment assignment and treatment status do not coincide for at least some units, that is, settings where some units assigned to treatment fail to receive the treatment and/or some units assigned to the control condition receive the treatment anyway. To formalize this imperfect compliance setup, we expand the statistical model for sharp RD designs and introduce two binary potential treatments for each unit, $D_i(0)$ and $D_i(1)$, where $D_i(1)$ is the actual treatment received when assigned to treatment (i.e., when $X_i \geq c$) and $D_i(0)$ is the actual treatment received when assigned to control (i.e., when $X_i < c$). The observed treatment for each unit is then redefined to $D_i = D_i(0)\mathbb{1}(X_i < c) + D_i(1)\mathbb{1}(X_i \geq c)$. Perfect compliance with treatment assignment means $\mathbb{P}[D_i(0) = 0|X_i = x] = 1$ for $x < c$ (i.e., all units with score below the cutoff fail to receive the treatment) and $\mathbb{P}[D_i(1) = 1|X_i = x] = 1$ for $x \geq c$ (i.e., all units with score below the cutoff receive the treatment), in which case the treatment assignment rule reduces to the sharp RD rule, $D_i = \mathbb{1}(X_i \geq c)$.

As mentioned above, if the goal is to learn about the effect of assigning the treatment rather than receiving it, researchers can simply set $D_i = \mathbb{1}(X_i \geq c)$ and follow a Sharp RD analysis of the effect of treatment assignment (i.e., compare units below and above the cutoff), regardless of their actual treatment received. However, in many cases researchers are also interested in the effect of actually receiving the treatment on the outcome of interest. In general, identification of such treatment effects in settings with imperfect compliance cannot be achieved with either continuity or local randomization assumptions alone, and thus additional identifying assumptions are needed. Therefore, learning about interpretable causal treatment effects in Fuzzy RD designs is more difficult than in Sharp RD designs.

Within the continuity-based framework, the generic Fuzzy RD estimand is

$$\tau_{\mathsf{FRD}} = \frac{\lim_{x \downarrow c} \mathbb{E}[Y_i|X_i = x] - \lim_{x \uparrow c} \mathbb{E}[Y_i|X_i = x]}{\lim_{x \downarrow c} \mathbb{E}[D_i|X_i = x] - \lim_{x \uparrow c} \mathbb{E}[D_i|X_i = x]}.$$

The causal interpretation of τ_{FRD} depends on the additional assumptions invoked, and the particular way in which treatment compliance is defined. Hahn, Todd, and van der Klaauw (2001) showed that if the potential outcomes and the treatment received are independent conditional on the score being near the cutoff, then $\tau_{\mathsf{FRD}} = \mathbb{E}[Y_i(1) - Y_i(0)|X = c]$, removing the problem of imperfect

compliance and identifying the same parameter as in a Sharp RD design. However, this local conditional independence assumption is strong and thus not commonly used to interpret the Fuzzy RD estimand in a continuity-based approach.

Several authors have derived identification results to interpret τ_{FRD} based on continuity conditions. In all cases, the results seek to characterize the parameter τ_{FRD} as a causal estimand for some subpopulation of "compliers" at the cutoff, heuristically defined as units who actually receive the treatment when they are assigned to the treatment condition and remain untreated when they are assigned to the control condition. For instance, Dong (2018a) defines compliance types of units for a given arbitrary value of their score and imposes a monotonicity condition in a neighborhood of the cutoff to achieve identification, while Arai, Hsu, Kitagawa, Mourifié, and Wan (2021) define units' compliance types in a neighborhood near the cutoff, assume that each unit's type is constant for all values of x in this neighborhood, and impose a monotonicity condition at the cutoff. A somewhat more restrictive definition of types of units is used by Bertanha and Imbens (2020), who assume that, given the value of the cutoff c, the decision to take the treatment is only a function of the score via the indicator $\mathbb{1}(X_i \geq c)$, implying that a unit's compliance type is constant for all values of the score with equal treatment assignment. Other related approaches are discussed in Imbens and Lemieux (2008), Frandsen, Frolich, and Melly (2012), Cattaneo, Keele, Titiunik, and Vazquez-Bare (2016), and Choi and Lee (2018).

Regardless of the particular assumptions adopted to achieve a causal interpretation for τ_{FRD}, all methods arrive at the same qualitative identification result: under continuity of the appropriate functions, and some form of monotonicity and exclusion restrictions, the fuzzy RD estimand τ_{FRD} can be interpreted as the average treatment effect at the cutoff for compliers, where the specific definition of complier varies by approach.

In the local randomization framework, causal interpretability of Fuzzy RD treatment effects is better understood but specific to the approach employed (Fisherian, Neyman or super-population). The generic Fuzzy RD estimand can be written as

$$\tau_{\mathsf{FLR}} = \frac{\frac{1}{N_{\mathcal{W}}^+}\sum_{i:X_i\in\mathcal{W},X_i\geq c}\mathbb{E}[Y_i|X_i\in\mathcal{W},X_i\geq c] - \frac{1}{N_{\mathcal{W}}^-}\sum_{i:X_i\in\mathcal{W},X_i<c}\mathbb{E}[Y_i|X_i\in\mathcal{W},X_i<c]}{\frac{1}{N_{\mathcal{W}}^+}\sum_{i:X_i\in\mathcal{W},X_i\geq c}\mathbb{E}[D_i|X_i\in\mathcal{W},X_i\geq c] - \frac{1}{N_{\mathcal{W}}^-}\sum_{i:X_i\in\mathcal{W},X_i<c}\mathbb{E}[D_i|X_i\in\mathcal{W},X_i<c]},$$

where $N_{\mathcal{W}}^-$ denotes the number of control units within the window $\mathcal{W}$, and $N_{\mathcal{W}}^+ = N_{\mathcal{W}} - N_{\mathcal{W}}^-$.

Regardless of the specific local randomization framework, standard ideas from the literature on analysis of experiments can be used to attach a causal interpretation to the estimand $\tau_{\mathtt{FLR}}$, for example, as an average treatment effect for compliers with scores within the local randomization neighborhood $\mathcal{W}$.

2.3   Multi-Dimensional Designs

In the canonical RD designs discussed so far, both the score and the cutoff are one-dimensional, and the treatment is binary. We now consider generalizations of these conditions, which we generically label *multi-dimensional* RD designs. These generalized RD designs require specific modifications for interpretation and analysis. Conceptually, however, the core of the design remain the same: a discontinuous change in the probability of receiving treatment is the basis for comparing units assigned to treatment and control near the region where such a change occurs, under assumptions that guarantee the comparability of those units. Most of the designs discussed in this subsection admit sharp and fuzzy versions, although we avoid distinguishing each case due to space limitations.

Papay, Willett, and Murnane (2011) introduced the Multi-Score RD design; see also Reardon and Robinson (2012) and Wong, Steiner, and Cook (2013). In its more general version, the Multi-Score RD design includes two or more scores assigning units to a range of different treatment conditions, but a more common setup is one where the score is multi-dimensional but the treatment is still binary. For example, students may need to take a language exam and a mathematics exam, and obtain a high enough grade in both exams to be admitted to a school. In such two-dimensional example, the score is the vector $\mathbf{X}_i = (X_{1i}, X_{2i})'$, the cutoff is the vector $\mathbf{c} = (c_1, c_2)$, and the treatment assignment rule is $D_i = \mathbb{1}(X_{1i} \geq c_1)\mathbb{1}(X_{2i} \geq c_2)$. This rule shows that there are potentially infinite ways for a unit to be barely treated or barely control, that is, there are infinitely many points at which the treatment assignment jumps discontinuously from zero to one, leading to the definition of a treatment effect curve. This idea generalizes to higher-dimensions.

An important particular case of the Multi-Score RD design is the Geographic RD design, where the treatment is assigned to units that reside in a particular geographic area (the *treated area*), and not assigned to units who are in an adjacent area (the *control area*). The RD score is two-dimensional to reflect each unit's position in space, usually latitude and longitude. The geographic boundary that separates the treated and control areas thus forms an infinite collection of points

at which the treatment assignment jumps discontinuously from zero to one. See Black (1999) for a pioneer use of a Geographic RD research design, and Dell (2010) for the first paper to explicitly exploit latitude and longitude to achieve RD-specific geographic identification of an average treatment effect. Modern analysis of Geographic RD designs is discussed by Keele and Titiunik (2015) using continuity-based ideas, and by Keele, Titiunik, and Zubizarreta (2015) using local randomization ideas. Keele, Lorch, Passarella, Small, and Titiunik (2017) discuss available options for analysis when the exact geographic location of each unit in the dataset is unavailable.

Cattaneo, Keele, Titiunik, and Vazquez-Bare (2016) introduced the Multi-Cutoff RD design, where different units in the study receive the treatment according to different cutoff values along a univariate score. In this setting, the variable X_i continues to denote the scalar score for unit i, but instead of a fixed common cutoff the setting now allows for an additional random variable, C_i, that denotes the cutoff faced by each unit. The cutoff random variable C_i has support $\mathcal{C}$, which can be countable or uncountable. For example, a natural assumption is $\mathcal{C} = \{c_1, c_2, ..., c_J\}$, where the values $c_1, c_2, ..., c_J$ can either be distinct cutoffs in an RD design or a discretization of the continuous support of the cutoff variable. If $\mathcal{C} = \{c\}$, then the canonical (Sharp and Fuzzy) RD designs are recovered.

The Multi-Score RD design can be recast and analyzed as a Multi-Cutoff RD design by appropriate choice of the score and cutoff variables, X_i and C_i. For example, it is common practice to study Multi-Score RD treatment effects for a finite set of values (i.e., a discretization) of the treatment assignment curve, which can be naturally interpreted as a Multi-Cutoff RD design with each cutoff being one point on the discretization of the curve. In addition, as discussed in Section 2.4, Multi-Cutoff RD designs may be useful for extrapolation of RD treatment effects (Cattaneo, Keele, Titiunik, and Vazquez-Bare, 2021).

Both Multi-Score and Multi-Cutoff RD designs offer a rich class of RD treatment effects. For a given treatment assignment boundary curve (in the Multi-Score RD design) or each cutoff value (in the Multi-Cutoff RD design), a Sharp or Fuzzy RD treatment effect can be defined, using either the continuity or local randomization frameworks. Furthermore, it is common practice to also normalize and pool the data along the treatment assignment boundary curve or the multiple cutoff values in order to consider a single, pooled RD treatment effect. Keele and Titiunik (2015) and Cattaneo, Keele, Titiunik, and Vazquez-Bare (2016) discuss a causal link between the normalized-

and-pooled RD treatment effect and the RD treatment effects along the treatment assignment curve or at multiple cutoffs. See Cattaneo, Idrobo, and Titiunik (2021) for a practical introduction to Multi-Score and Multi-Cutoff RD designs methods.

In addition to the Multi-Cutoff and Multi-Score RD designs, there are several other types of RD designs with multi-dimensional features. Dong, Lee, and Gou (2021) consider an RD design with a continuous (rather than binary or multi-valued) treatment, within the continuity framework, and show that RD causal effects can be identified based on changes in the probability distribution of the continuous treatment at the cutoff. Caetano, Caetano, and Escanciano (2021) investigate continuity-based identification of RD causal effects where the treatment is multi-valued with finite support. Grembi, Nannicini, and Troiano (2016) introduce a second dimension to the RD design in a setup where there are two time periods, and discuss continuity-based identification of RD treatment effects across the two periods, in a design they call *difference-in-discontinuities* design, building on ideas from the difference-in-differences design in program evaluation (Abadie and Cattaneo, 2018). Cellini, Ferreira, and Rothstein (2010) consider a different type of multi-dimensionality induced by a dynamic context in which the RD design occurs in multiple periods for the same units and the score is re-drawn in every period so that a unit may be assigned to treatment in one period but control in future periods; see also Hsu and Shen (2021a) for an econometric analysis of a dynamic RD design within the continuity-based framework. Lv, Sun, Lu, and Li (2019) consider the generalization of the RD design to survival data settings, where the treatment is assigned at most once per unit and the outcome of interest is the units' survival time in a particular state, which may be censored. Xu (2018) also studies an RD designs with duration outcomes, but assumes the outcomes have discrete support.

2.4 Extrapolation

RD designs offer credible identification of treatment effects at or near the cutoff, via either continuity or local randomization frameworks, often with superior internal validity relative to most other observational methods. However, an important limitation of RD designs is that they have low external validity. In the absence of additional assumptions, it is not possible to learn about treatment effects away from the cutoff—that is, to extrapolate the effect of the treatment. In the continuity-based Sharp RD design, for example, the average treatment effect at the cutoff is

$\tau_{\text{SRD}} = \tau_{\text{SRD}}(c)$ with $\tau_{\text{SRD}}(x) = \mathbb{E}[Y_i(1) - Y_i(0)|X_i = x]$. However, researchers and policy-makers may also be interested in average treatment effects at other values of the score variable such as $\tau_{\text{SRD}}(x)$ for x different (and possibly far) from c, which are not immediately available.

In recent years, several approaches have been proposed to enable valid extrapolation of RD treatment effects. In the context of sharp designs, Mealli and Rampichini (2012) and Wing and Cook (2013) rely on an external pre-intervention measure of the outcome variable and employ parametric methods to impute the treated-control differences of the post-intervention outcome above the cutoff. Dong and Lewbel (2015) and Cerulli, Dong, Lewbel, and Poulsen (2017) study local extrapolation methods via derivatives of the RD average treatment effect in the continuity-based framework, while Angrist and Rokkanen (2015) employ pre-intervention covariates under a conditional ignorability condition within a local randomization framework. Rokkanen (2015) relies on multiple measures of the score, which are assumed to capture a common latent factor, and employs ideas from a local randomization framework to extrapolate RD treatment effects away from the cutoff assuming that the potential outcomes are conditionally independent of the available measurements given the latent factor. In the context of continuity-based fuzzy designs, Bertanha and Imbens (2020) exploit variation in treatment assignment generated by imperfect treatment compliance imposing independence between potential outcomes and compliance types to extrapolate average RD treatment effects.

In the context of (Sharp and Fuzzy) Multi-Cutoff RD designs, and from both continuity and local randomization perspectives, Cattaneo, Keele, Titiunik, and Vazquez-Bare (2016) discuss extrapolation in settings with ignorable cutoff assignment, while Cattaneo, Keele, Titiunik, and Vazquez-Bare (2021) discuss extrapolation allowing for heterogeneity across cutoffs that is invariant as function of the score. In addition, within the continuity framework, Bertanha (2020) discusses extrapolation of average RD treatment effects assuming away cutoff-specific treatment effect heterogeneity. In the specific context of geographic RD designs, Keele and Titiunik (2015) and Keele, Titiunik, and Zubizarreta (2015) discuss extrapolation of average treatment effects along the geographic boundary determining treatment assignment. See also Galiani, McEwan, and Quistorff (2017) and Keele, Lorch, Passarella, Small, and Titiunik (2017) for related discussion.

Finally, there is a literature that investigates the relationship between RD designs and randomized experiments and discusses the relationship between local RD effects and more general

parameters. For example, Cook and Wong (2008) report a study of three empirical RD designs and how they compare to experimental benchmarks created by randomized experiments of the same intervention. See Chaplin, Cook, Zurovac, Coopersmith, Finucane, Vollmer, and Morris (2018), Hyytinen, Meriläinen, Saarimaa, Toivanen, and Tukiainen (2018), and De Magalhaes, Hangartner, Hirvonen, Meriläinen, Ruiz, and Tukiainen (2020) for more recent results on RD recovery of experimental benchmarks.

2.5 Other RD-like Designs

The are several other research designs with qualitatively similar features to RD designs but also important differences. Some of these research designs can be analyzed using the econometric tools discussed in Sections 3 and 4, while for others specific modifications would be needed.

Card, Lee, Pei, and Weber (2015, 2017) introduce the regression kink (RK) design. Canonical and generalized RD designs assume a discontinuity in the treatment assignment rule, leading to a discontinuous jump in the probability of receiving treatment at the cutoff. In contrast, in an RK design there is no such discontinuous jump; instead, the policy rule is assumed to have a kink in the relationship between the policy (or treatment) and the score at a known cutoff point (also known as the kink point). Such kinks arise naturally in social and behavioral sciences, such as when a tax or benefit schedule is a piecewise linear function of baseline income. The core idea in the RK design is to examine the slope of the conditional relationship between the outcome and the score at the kink point in the policy formula. Chiang and Sasaki (2019) extends this idea to quantile treatment effects. See also Dong (2018b) for related discussion. From the point of view of implementation, estimation and inference of kink treatment effects is equivalent to estimating RD treatment effects for derivatives of regression functions. In the RD estimation and inference literature, generic estimands defined as differences of first derivatives of regression functions at the cutoff, or ratios thereof, are referred to as kink RD designs (e.g., Calonico, Cattaneo, and Titiunik, 2014, Sections 3.1 and 3.3).

Another area with interesting connections to the RD design is the literature on bunching and density discontinuities (Kleven, 2016; Jales and Yu, 2017). In this setting, the objects of interest are related to discontinuities and other sharp changes in probability density functions, usually motivated by economic or other behavioral models. These models typically exploit discontinuities

14

on treatment assignment or policy rules, like RD or RK designs, but they focus on estimands that are outside the support of the observed data. As a consequence, identification (as well as estimation and inference) requires additional parametric modelling assumptions that are invoked for extrapolation purposes. See Blomquist, Newey, Kumar, and Liang (2021) for a discussion of nonparametric identification in bunching and density discontinuities settings.

Before-and-after (or event) studies are also sometimes portrayed as "RD designs in time" where the time index is taken as the score variable and the analysis is conducted using RD methods (Hausman and Rapson, 2018). While it is hard to reconcile those event studies as traditional RD designs in general, the local randomization RD framework can sometimes be adapted and used to analyze RD designs in time (see Sections 3.3 and 3.5 for more discussion). Another RD-like example is Porter and Yu (2015), who consider settings where the cutoff in an assignment rule is unknown and must be estimated from the data, a problem that can occur, for example, when unknown preferences generate tipping points where behavior changes discontinuously (Card, Mas, and Rothstein, 2008).

3 Estimation and Inference

We now discuss methods for RD analysis, focusing on the continuity-based and local randomization approaches separately. Before presenting formal methods for estimation and inference, we briefly discuss how to present the RD design visually using a combination of global and local techniques.

3.1 Visualization of the RD design

One of the advantages of the RD design is its transparency. The running variable gives the observations a natural ordering, and the treatment assignment rule links this ordering to treatment assignment, which in turn is hypothesized to affect an outcome variable of interest. These ingredients can all be displayed graphically to provide a visual illustration of the design, usually referred to as an RD Plot. This plot has become a central component of RD empirical analysis, regardless of the formal methodology employed for estimation and inference. Although a graphical illustration cannot replace formal econometric analysis, it is often helpful to display the general features of the study, to gauge whether the discontinuity observed at the cutoff is unusual with respect to the

overall shape of the regression function, as well as to have an overall understanding of the data to be used in the subsequent formal analysis.

The RD plot graphs different estimates of the conditional expectation of the outcome function given the score, $\mathbb{E}[Y_i|X_i = x]$, as a function of x. The typical plot has two ingredients: a global polynomial fit of the outcome on the score, and local sample means of the outcome computed in small bins of the score variable. Both ingredients are calculated separately for treated and control observations because $\mathbb{E}[Y_i|X_i = x]$ may exhibit a discontinuity at the cutoff, thereby capturing two different regression functions: $\mathbb{E}[Y_i(0)|X_i = x]$ for observations below the cutoff, and $\mathbb{E}[Y_i(1)|X_i = x]$ for observations above the cutoff (Figure 1).

Each ingredient plays a different role. The two global polynomial fits are designed to show a smooth global approximation of the regression functions that gives a sense of their overall shape—for example, to see whether they are increasing or decreasing in x, or whether there are nonlinearities. The typical strategy is to fit a polynomial of order four, but this should be modified as needed. The local means are constructed by partitioning the support of the score into disjoint bins, and computing sample averages of the outcome for each bin, using only observations whose score value falls within that bin; each local mean is plotted against the bin mid-point. In contrast to the global fit, the local means are intended to give a sense of the local behavior of the regression function. In this sense, by simultaneously depicting two representations of the regression function, one smooth (the global polynomial fit) and the other discontinuous (the local means), the RD plot provides rich information about the overall data underlying the RD design.

Calonico, Cattaneo, and Titiunik (2015) propose methods for implementing RD plots in a data-driven and disciplined way, and Cattaneo, Idrobo, and Titiunik (2020, Section 3) offer a practical introduction to RD Plots. In addition, see Pei, Lee, Card, and Weber (2021) for a discussion on how to choose polynomial orders in RD designs, which could be used to further improve the implementation of RD Plots.

Importantly, RD plots should not be used as substitutes for formal estimation and inference of RD effects. In particular, sometimes the global polynomial fit will exhibit a "jump" at the cutoff, which may be indicative of a treatment effect. Although useful as graphical and suggestive evidence, such jump should not be interpreted as the definitive RD effect, as higher-order polynomial approximations are poorly behaved at or near boundary points—a problem known as the Runge's

phenomenon in approximation theory (Calonico, Cattaneo, and Titiunik, 2015, Section 2.1.1). Furthermore, global polynomials for RD estimation and inference lead to several other undesirable features (Gelman and Imbens, 2019), and are therefore not recommended for analysis beyond visualization.

3.2 Local Polynomial Methods

The continuity-based identification result of Hahn, Todd, and van der Klaauw (2001) naturally justifies estimation of RD effects via local polynomial nonparametric methods (Fan and Gijbels, 1996). These methods have become the standard for estimation and inference in the RD design under continuity conditions since the early work of Porter (2003) and Sun (2005), and the more recent work of Imbens and Kalyanaraman (2012) and Calonico, Cattaneo, and Titiunik (2014). Local polynomial methods are preferable to global polynomial methods because, as mentioned above, they avoid several of the methodological problems created by the use of global polynomials such as erratic behavior near boundary points (Runge's phenomenon), counterintuitive weighting, overfitting, and general lack of robustness.

Local polynomial analysis for RD designs is implemented by fitting Y_i on a low-order polynomial expansion of X_i, separately for treated and control observations, and in each case using only observations near the cutoff rather than all available observations, as determined by the choice of a kernel function and a bandwidth parameter. The implementation requires four steps. First, choose the polynomial order p and the weighting scheme or kernel function $K(\cdot)$. Second, given the choices of p and $K(\cdot)$, choose a bandwidth h that determines a neighborhood around the cutoff so that only observations with scores within that neighborhood are included in the estimation. Third, given the choices of p, $K(\cdot)$, and h, construct point estimators using standard least-squares methods. Fourth, given the above steps, conduct valid statistical inference for the RD parameter of interest. We discuss these steps in more detail in the remaining of this subsection, focusing on methodological aspects.

The order of the polynomial should always be low, to avoid overfitting and erratic behavior near the cutoff point. The default recommendation is $p = 1$ (local linear regression), but $p = 2$ (local quadratic regression) or $p = 3$ (local cubic regression) are also reasonable choices in some empirical settings. See Pei, Lee, Card, and Weber (2021) for a data-driven choice of $p \in \{0, 1, 2, 3\}$

based on minimizing the asymptotic mean squared error (MSE) of the RD point estimator. The choice of kernel, $K(\cdot)$, assigns different weights for different observations according to the proximity of each observation's score to the cutoff. Common choices are a triangular kernel which assigns weights that are highest at the cutoff and decrease linearly for values of the score away from the cutoff, and a uniform kernel that assigns all observations the same weight. The triangular kernel has a point estimation MSE optimality property when coupled with an MSE-optimal bandwidth choice (discussed below), while the uniform kernel has an inference optimality property (i.e., it minimizes the asymptotic variance of the local polynomial estimator, and hence the interval length of commonly used confidence intervals). Both choices of kernel function are reasonable, although in practice the triangular kernel is often the default.

Once p and $K(\cdot)$ have been selected, the researcher must choose the bandwidth. When a particular value of h is chosen, this means that the polynomial fitting will only use observations such that $X_i \in [c-h, c+h]$, since the kernel functions used in RD designs are always compact supported. The notation assumes that the same bandwidth h is used both above and below the cutoff for simplicity, but several approaches have been proposed to accommodate distinct bandwidths for control and treatment groups. The choice of bandwidth is critical because empirical results are generally sensitive to changes in this tuning parameter, and therefore choosing the bandwidth in an arbitrary manner is discouraged. Instead, the recommendation is always to use a data-driven procedure that is optimal for a given criterion, leading to a transparent, principled, and objective choice for implementation that ultimately removes the researcher's discretion. Such an approach to tuning parameter selection enhances replicability and reduces the potential for p-hacking in empirical work.

Imbens and Kalyanaraman (2012) first proposed to choose h to minimize a first-order approximation to the MSE of the RD point estimator, leading to an MSE-optimal bandwidth choice. For implementation, they use a first-generation plug-in rule, where unknown quantities in the MSE-optimal bandwidth formula are replaced by inconsistent estimators. Calonico, Cattaneo, and Titiunik (2014) later obtained more general MSE expansions and MSE-optimal bandwidth choices, which are implemented using a second-generation plug-in rule (i.e., unknown quantities in the MSE-optimal bandwidth formula are replaced by consistent estimators thereof). Calonico, Cattaneo, and Farrell (2020) take a different approach and develop optimal bandwidth choices for

18

inference: their proposed bandwidth choices either (i) minimize the coverage error (CE) of confidence intervals or (ii) optimally trade off coverage error and length of confidence intervals. The CE-optimal bandwidth choice can also be interpreted as optimal to reduce the higher-order errors in the distributional approximation used for local polynomial statistical inference. Cattaneo and Vazquez-Bare (2016) provide an overview of bandwidth selection methods for RD designs.

Given a polynomial order p, a kernel function $K(\cdot)$, and a bandwidth h, the RD local polynomial point estimator is obtained by fitting two weighted least squares regressions of Y_i on a polynomial expansion of X_i as follows:

$$\widehat{\boldsymbol{\beta}}_- = \underset{b_0,\ldots,b_p}{\arg\min} \sum_{i=1}^n \mathbb{1}(X_i < c) \left(Y_i - b_0 - b_1(X_i - c) - b_2(X_i - c)^2 - \ldots - b_p(X_i - c)^p\right)^2 K\left(\frac{X_i - c}{h}\right)$$

$$\widehat{\boldsymbol{\beta}}_+ = \underset{b_0,\ldots,b_p}{\arg\min} \sum_{i=1}^n \mathbb{1}(X_i \geq c) \left(Y_i - b_0 - b_1(X_i - c) - b_2(X_i - c)^2 - \ldots - b_p(X_i - c)^p\right)^2 K\left(\frac{X_i - c}{h}\right)$$

where $\widehat{\boldsymbol{\beta}}_- = (\widehat{\beta}_{-,0}, \widehat{\beta}_{-,1}, \ldots, \widehat{\beta}_{-,p})'$ and $\widehat{\boldsymbol{\beta}}_+ = (\widehat{\beta}_{+,0}, \widehat{\beta}_{+,1}, \ldots, \widehat{\beta}_{+,p})'$ denote the resulting least squares estimates for control and treatment groups, respectively. The RD point estimator of the sharp treatment effect τ_{SRD} is the estimated vertical distance at the cutoff, that is, the difference in intercepts: $\widehat{\tau}_{\mathsf{SRD}}(h) = \widehat{\beta}_{+,0} - \widehat{\beta}_{-,0}$. Under standard assumptions, $\widehat{\tau}_{\mathsf{SRD}}(h)$ will be consistent for $\tau_{\mathsf{SRD}} = \mathbb{E}[Y_i(1) - Y_i(0)|X = c]$. It is common practice to use $\widehat{\tau}_{\mathsf{SRD}}(h_{\mathsf{MSE}})$, where h_{MSE} denotes an MSE-optimal bandwidth choice, thereby delivering not only a consistent but also an MSE-optimal point estimator.

A crucial question is how to perform valid statistical inference for the RD parameter based on local polynomial methods with an MSE-optimal bandwidth. At first sight, it seems that standard least squares results could be applied to construct confidence intervals because, given a bandwidth choice, estimation is implemented via weighted least-squares linear regression. However, conventional least squares inference methods are parametric, assuming the polynomial expansion and kernel function used capture the correct functional form of the unknwon conditional expectations. Conventional least squares methods thus assume away any misspecification error to conclude that the usual t-statistic has an asymptotic standard normal distribution, which (if true) would lead to

the standard approximately 95% least squares confidence interval:

$$I_{\mathsf{LS}} = \left[\ \widehat{\tau}_{\mathsf{SRD}}(h_{\mathsf{MSE}}) \ \pm \ 1.96 \cdot \sqrt{\widehat{\mathsf{V}}} \ \right],$$

where $\sqrt{\widehat{\mathsf{V}}}$ denotes any of the usual standard error estimators in least-squares regression settings.

However, RD local polynomial methods are nonparametric in nature and thus view the polynomial fit near the cutoff as an approximation to unknown regression functions, which generally carries a misspecification error (or smoothing bias). In fact, if there was no misspecification bias, then standard MSE-optimal bandwidth choices would not be well-defined to begin with. It is therefore not possible to simultaneously rely on MSE-optimal bandwidth choices and conduct inference without taking into account misspecification errors in the approximation of the conditional expectations near the cutoff. Crucially, the approximation errors will translate into a bias in the distributional approximation of the RD local polynomial estimator, where in general we have

$$\frac{\widehat{\tau}_{\mathsf{SRD}}(h_{\mathsf{MSE}}) - \tau_{\mathsf{SRD}}}{\sqrt{\widehat{\mathsf{V}}}} \approx_d \mathcal{N}(\mathsf{B}, 1),$$

with $\approx_d$ denoting approximately in distribution, $\mathcal{N}(\cdot)$ denoting the Gaussian distribution, and B capturing the (standardized) misspecification bias or approximation error. Ignoring the misspecification bias can lead to substantial over-rejection of the null hypothesis of no treatment effect ($\mathsf{H}_0 : \tau_{\mathsf{SRD}} = 0$)—for example, back-of-the-envelop calculations show that a nominal 95% confidence interval would have an empirical coverage of about 80%.

For some choices of h smaller than h_{MSE}, the bias term B will vanish asymptotically, which is called "undersmoothing" in the nonparametric literature. Although researchers could choose some $h < h_{\mathsf{MSE}}$ and appeal to asymptotics, such approach is undisciplined and arbitrary (the bias may still be large in finite samples), leads to reduced power (fewer observations are used for inference), and requires employing different observations for MSE-optimal point estimation and for valid inference. As a principled alternative, Calonico, Cattaneo, and Titiunik (2014) propose to use a different confidence interval that estimates and removes the misspecification bias in the asymptotic approximation and, at the same time, adjusts the standard error formula to account for the additional variability introduced in the bias-estimation step. Specifically, the approximately

95% *robust bias-corrected* confidence interval is:

$$I_{\text{RBC}} = \left[\left(\widehat{\tau}_{\text{SRD}}(h_{\text{MSE}}) - \widehat{\text{B}} \right) \; \pm \; 1.96 \cdot \sqrt{\widehat{\text{V}} + \widehat{\text{W}}} \; \right],$$

where $\widehat{\text{B}}$ denotes the estimated bias correction and $\widehat{\text{W}}$ denotes the adjustment in the standard errors.

The robust bias-corrected confidence interval I_{RBC} is recentered and rescaled with respect to the conventional interval I_{LS}. The new standardization is theoretically justified by a nonstandard large-sample distributional approximation of the bias-corrected estimator that explicitly accounts for the potential contribution of bias correction to the finite-sample variability of the usual t-statistic. This confidence interval is valid whenever the conventional confidence interval is valid, and continues to offer correct coverage in large samples even when the conventional confidence interval does not— hence the name *robust*. In particular, I_{RBC} is valid even when the MSE-optimal bandwidth h_{MSE} is employed, thereby allowing researchers to use the same bandwidth for both (optimal) point estimation and (valid, possibly optimal) inference. See Cattaneo, Idrobo, and Titiunik (2020, Section 4) for a practical guide.

Robust bias correction (RBC) inference methods have been further developed in recent years. From the theoretical side, Kamat (2018) gives conditions under which RBC inference has uniform validity, Calonico, Cattaneo, and Farrell (2018, 2021) demonstrate superior higher-order properties of RBC inference both pointwise and uniformly over standard data generating classes, and Tuvaandorj (2020) establishes minimax optimality. Methodologically, these theoretical results justify a simple implementation of RBC inference where the point estimation polynomial order p is increased when conducting inference without requiring any other changes (Calonico, Cattaneo, and Titiunik, 2014, Remark 7), which allows for using the same data for MSE-optimal point estimation and valid inference. Finally, from a practical perspective, RBC inference has been found to exhibit excellent performance in both simulations and replication studies (e.g., Ganong and Jäger, 2018; Hyytinen, Meriläinen, Saarimaa, Toivanen, and Tukiainen, 2018; De Magalhaes, Hangartner, Hirvonen, Meriläinen, Ruiz, and Tukiainen, 2020).

The estimation and inference methods described so far extend naturally to Fuzzy RD designs, where the local polynomial point estimator is now a ratio of sharp RD estimators, and to sharp and fuzzy kink RD designs where the point estimator is based on estimated higher-order derivatives at

the kink point. For example, the Fuzzy RD estimand τ_{FRD} is estimated by $\widehat{\tau}_{\mathsf{FRD}}(h) = \widehat{\tau}_{\mathsf{SRD}}(h)/\widehat{\varrho}_{\mathsf{SRD}}(h)$ with $\widehat{\varrho}_{\mathsf{SRD}}(h)$ denoting a sharp RD local polynomial estimate using the observed treatment D_i as outcome instead of Y_i—that is, $\widehat{\varrho}_{\mathsf{SRD}}(h)$ is the sharp RD effect on receiving the treatment at the cutoff. For more details on fuzzy and higher-order derivative estimation and inference see Calonico, Cattaneo, and Titiunik (2014), and Cattaneo, Idrobo, and Titiunik (2021, Section 3) for a practical discussion.

The estimation and inference results for canonical Sharp and Fuzzy RD designs have been expanded in multiple directions. Arai and Ichimura (2016, 2018) propose optimal MSE-optimal bandwidth selection for different control and treatment bandwidths and develop RBC inference, Bartalotti, Calhoun, and He (2017) and He and Bartalotti (2020) investigate resampling methods and their connection with RBC inference, and Bartalotti and Brummet (2017) develop MSE-optimal bandwidth selection and RBC inference in clustered sampling settings. Calonico, Cattaneo, Farrell, and Titiunik (2019) investigate inclusion of pre-intervention covariates for efficiency purposes in RD designs, and develop MSE-optimal bandwidth selection and RBC inference with possibly clustered data and/or different bandwidth choices for control and treatment groups. More recently, Arai, Otsu, and Seo (2021) proposed novel high-dimensional methods for pre-intervention covariate selection, which lead to increased efficiency of RBC inference.

Continuity-based estimation and RBC inference methods have also been studied in other (sharp, fuzzy and kink) RD settings. For example, Xu (2017) considers generalized linear models and develops RBC inference methods, Keele and Titiunik (2015), Cattaneo, Keele, Titiunik, and Vazquez-Bare (2016, 2021), Bertanha (2020) address estimation and RBC inference in Multi-Score, Geographic and Multi-Cutoff RD designs, Dong, Lee, and Gou (2021) offer a complete analysis of estimation and RBC inference in an RD design with continuous treatment, and Chiang, Hsu, and Sasaki (2019), Qu and Yoon (2019), Chen, Chiang, and Sasaki (2020), and Huang and Zhan (2021) investigate quantile RD designs and also propose RBC inference methods.

Finally, within the continuity-based framework, scholars have addressed different failures of standard assumptions in canonical and other RD designs. Dong (2019) explores issues of sample selection in sharp and fuzzy RD designs. For Fuzzy RD designs, Feir, Lemieux, and Marmer (2016) and He (2018) study methods robust to weak identification, while Arai, Hsu, Kitagawa, Mourifié, and Wan (2021) consider specification testing. In addition, Pei and Shen (2017), Lee (2017),

Davezies and Le Barbanchon (2017) and Bartalotti, Brummet, and Dieterle (2021) investigate issues of measurement error in various RD settings.

3.3   Local Randomization Methods

In their original article, Thistlethwaite and Campbell (1960) conceived the RD treatment assignment rule as inducing a sort of random assignment near the cutoff, and drew a strong analogy between the RD design and a randomized experiment, describing a true experiment as a design "for which the regression-discontinuity analysis may be regarded as a substitute" (p. 310). The analogy between RD designs and randomized experiments took greater force after the work of Lee (2008), who advocated for interpreting the RD design as inducing "as good as randomized" (p. 1) variation in treatment status near the cutoff. See also Lee and Lemieux (2010).

The analogy between RD designs and randomized experiments was first formalized without continuity conditions by Cattaneo, Frandsen, and Titiunik (2015) using a Fisherian framework, and later expanded by Cattaneo, Titiunik, and Vazquez-Bare (2017) to also allow for Neyman or super-population frameworks, after possible score-adjustments of potential outcomes. Rather than relying on limits as the score tends to the cutoff and heuristic analogies between units barely above and barely below the cutoff, these works consider assumptions under the which the RD design would produce conditions equivalent to the conditions that would have occurred if a randomized experiment had been conducted in a neighborhood around the cutoff.

The key idea behind the local randomization approach to RD analysis is to assume that units near the cutoff are as-if randomly assigned to treatment. The formalization of the as-if random assignment requires that there exist a window $\mathcal{W} = [c - w, c + w]$ around the cutoff c, assumed symmetric only for simplicity, such that the following two conditions hold:

1. The joint probability distribution of scores in $\mathcal{W}$ is known: $\mathbb{P}[\mathbf{X}_{\mathcal{W}} \leq \mathbf{x}] = F(\mathbf{x})$, for some known joint c.d.f. $F(\mathbf{x})$ where $\mathbf{X}_{\mathcal{W}}$ denotes the vector of scores for all i such that $X_i \in \mathcal{W}$.

2. The potential outcomes are not affected by the score in $\mathcal{W}$: $Y_i(0) = Y_i(0, x)$ and $Y_i(1) = Y_i(1, x)$, for all units such that $X_i \in \mathcal{W}$.

The first condition assumes that, inside the window, the assignment mechanism of the score is known, as it would occur in an experiment. For example, this condition holds when all units

have the same probability of receiving all score values in $\mathcal{W}$. The second condition is an exclusion restriction that prevents the potential outcomes from being a function of the score inside $\mathcal{W}$.

The first condition is analogous to the requirement of classical randomized experiments, while the second condition is not typically stated in definitions of randomized experiments because it holds (implicitly) by construction. For example, when a treatment is randomly assigned, this usually involves generating a pseudo-random number and then assigning the treatment or control conditions based on this number. If we interpret this pseudo-random number as a score, it will be unrelated to the potential outcomes by virtue of having been generated independently of any information about the characteristics of the units. In contrast, in an RD design, the score is often correlated with the potential outcomes, and there are multiple channels by which a change in the score, even if small, can induce a change in the potential outcomes. Such a relationship between the score and the potential outcomes would hinder the comparability of units above and below the cutoff within $\mathcal{W}$, because of the lack of common support in the score. For this reason, the exclusion restriction inside $\mathcal{W}$ must be made explicit. See Cattaneo, Titiunik, and Vazquez-Bare (2017) for a relaxation of that exclusion condition, and Sekhon and Titiunik (2017, 2016) for more discussion on the differences between randomized experiments and RD designs. De la Cuesta and Imai (2016) and Cattaneo, Titiunik, and Vazquez-Bare (2020) further contrast the continuity and local randomization frameworks; see also the related papers in Cattaneo and Escanciano (2017).

Once a local randomization assumption has been invoked, the analysis of the RD design can proceed by deploying various methods from the analysis of experiments. This requires two steps. First, a window $\mathcal{W}$ where the local randomization is assumed to hold must be chosen. Second, given $\mathcal{W}$, enough assumptions about the assignment mechanism must be imposed to perform estimation and inference, as appropriate for the specific approach chosen (Fisherian, Neyman, super-population). These methods can be used with both continuous and discrete scores (Section 3.5).

The window selection step is the main implementation challenge for local randomization analysis, and is analogous to the bandwidth selection step in the continuity framework. Cattaneo, Frandsen, and Titiunik (2015) recommend a data-driven procedure to select the window based on pre-treatment covariates or placebo outcomes known to be unaffected by the treatment. The procedure is based on the common approach of testing for covariate balance to asses the validity of randomized experiments. If the RD local randomization assumption holds, pre-treatment covari-

ates should be unrelated to the score inside $\mathcal{W}$. Assuming that the covariates are related to the score outside of $\mathcal{W}$ leads to the following procedure for selecting $\mathcal{W}$: consider a sequence of nested windows around the cutoff, and for each candidate window perform "balance" tests of the null hypothesis of no treatment effect, beginning with the largest window and sequentially shrinking it until the null hypothesis fails to be rejected at some pre-set significance level. The procedure thus searches subsequently smaller windows until it finds the largest possible window such that the covariate balance null hypothesis cannot be rejected for that window and any smaller windows contained in it.

Once the window has been chosen, the second step is to deploy methods from the analysis of experiments to estimate the RD effect and perform statistical inference inside the selected window. These methods require to specify a randomization mechanism inside the window. A common strategy is to assume that every unit with score within $\mathcal{W}$ was assigned to treatment according to a fixed-margins randomization, where the probability of each treatment assignment vector is $\binom{N_{\mathcal{W}}}{N_{\mathcal{W}}^+}$. Another common strategy is to assume that units were assigned independently to treated and control inside the window, with constant probability equal to $N_{\mathcal{W}}^+/N_{\mathcal{W}}$.

The particular implementation of estimation and inference depends on the framework chosen for the analysis (Fisherian, Neyman or super-population). Fisherian methods are particularly useful when the number observations with score within $\mathcal{W}$ is small, because they provide finite-sample valid inference for the sharp null hypothesis of no treatment effect under reasonable assumptions. This inference method is called randomization inference in the causal inference and program evaluation literature (Rosenbaum, 2010; Imbens and Rubin, 2015; Abadie and Cattaneo, 2018), and is particularly well-suited for analyzing RD designs because it allows researchers to include only the few observations that are closest to the cutoff. As consequence, randomization inference offers a good complement and robustness check to local polynomial methods (Section 3.2), which typically employ a larger number observations further away from the cutoff for estimation and inference. On the other hand, when the number of observations inside $\mathcal{W}$ is large enough, researchers can also use methods based on large-sample approximations within either the Neyman framework, where potential outcomes are non-random and uncertainty is design-based, or the super-population framework, where potential outcomes are random and uncertainty is both design- and sampling-based. See Cattaneo, Idrobo, and Titiunik (2021, Section 2) for a practical introduction.

The local randomization approach has been applied in other RD settings. Keele, Titiunik, and Zubizarreta (2015) consider Multi-Score and Geographic RD designs, Li, Mattei, and Mealli (2015) analyze principal stratification in a fuzzy RD setup, Ganong and Jäger (2018) study kink RD designs, and Cattaneo, Keele, Titiunik, and Vazquez-Bare (2016, 2021) investigate Multi-Cutoff RD designs. In addition, Owen and Varian (2020) study a "tie-breaker" design where the treatment is actually randomly assigned in a small neighborhood of the cutoff but follows the RD assignment otherwise, while Eckles, Ignatiadis, Wager, and Wu (2021) consider a setup where exogenous measurement error is assumed to induce local randomization near the cutoff. See also Hyytinen, Meriläinen, Saarimaa, Toivanen, and Tukiainen (2018) and De Magalhaes, Hangartner, Hirvonen, Meriläinen, Ruiz, and Tukiainen (2020) for related work.

3.4 Refinements, Extensions and Other Methods

We review alternative approaches that have been proposed to refine, extend or generalize the standard RD estimation and inference methods discussed in the last two sections.

3.4.1 Empirical Likelihood Methods

Otsu, Xu, and Matsushita (2015) propose an inference approach for sharp and fuzzy RD designs combining local polynomial and empirical likelihood methods, which leads to asymptotically valid confidence intervals for average RD treatment effects under an undersmoothed bandwidth choice, but a more powerful robust bias-corrected version of their method can be constructed by increasing the order of the polynomial used (Section 3.2). Empirical likelihood inference is based on a data-driven likelihood ratio function, permits incorporating known constraints on parameters, and does not require explicit estimation of asymptotic variance formulas. See Owen (2001) for an introduction to empirical likelihood methodology. More recently, Ma and Yu (2020) investigated the higher-order properties of local polynomial empirical likelihood inference in RD designs, building on ideas in Calonico, Cattaneo, and Farrell (2018, 2021) for RBC inference using local polynomial methods. In addition, Otsu, Xu, and Matsushita (2013) and Ma, Jales, and Yu (2020) explored empirical likelihood methods for inference on a discontinuity in a density function, which has direct application both to the RD-like designs discussed in Section 2 and to the validation and falsification methodology discussed in Section 4. These empirical likelihood methods are developed within the

continuity framework.

3.4.2 Bayesian Methods

There are also several Bayesian methods for RD designs, which are developed almost exclusively within the local randomization framework. Geneletti, O'Keeffe, Sharples, Richardson, and Baio (2015) discuss a Bayesian methodology for RD designs in the biomedical sciences, focusing on how to incorporate external information in estimation and inference for RD treatment effects. Similarly, Karabatsos and Walker (2015) and Chib and Jacobi (2016) propose alternative Bayesian models for sharp and fuzzy RD designs, paying particular attention to issues of misspecifcation of unknown regression functions. Branson, Rischard, Bornn, and Miratrix (2019) employ Gaussian process regression methodology that can be interpreted as a Bayesian analog of the local linear RD estimator, thereby offering a flexible fit for regression functions of treatment and control units via a general prior on the mean response functions of the underlying Gaussian processes. Their approach is motivated by Bayesian methods, but they also discuss some frequentist properties of the resulting estimation procedures. All these Bayesian methods for RD analysis rely on specifying prior distributions for functional parameters (e.g., $\mathbb{E}[Y_i(1)|X_i = x]$ and $\mathbb{E}[Y_i(0)|X_i = x]$, the conditional expectation functions for treatment and control units). See Gelman, Carlin, Stern, Dunson, Vehtari, and Rubin (2013) for a general introduction to Bayesian methodology.

3.4.3 Uniform-in-Bias Methods

Imbens and Wager (2019) and Armstrong and Kolesar (2020) apply ideas from Backus (1989) and Donoho (1994) to develop RD inference methods that are uniformly valid over a smoothness class of functions determining the largest possible misspecification bias of the RD treatment effect estimator. Their procedures require pre-specifying a theoretical smoothness constant that determines the largest possible "curvature" of the unknown conditional expectation functions for treatment and control groups, $\mathbb{E}[Y_i(1)|X_i = x]$ and $\mathbb{E}[Y_i(0)|X_i = x]$, near the cutoff. Given that choice of smoothness constant, the inference approach computes an optimal uniform-in-bias bandwidth choice, which is then used to construct local polynomial confidence intervals for RD treatment effects that account for the worst possible bias of the RD point estimator (by inflating the quantile used). These methods are developed within the continuity framework for RD designs.

The uniform-in-bias methods depend on the theoretical smoothness constant, which is a tuning parameter that must be chosen. If this constant is estimated from data, the resulting inference procedures cease to be uniformly valid. As a consequence, assuming the goal is to achieve uniformity, the choice of the theoretical smoothness constant cannot be data-driven and must be set manually by the researcher. However, specifying this constant manually is equivalent to manually choosing the local polynomial bandwidth because there is a one-to-one correspondence between the two parameters. Finally, although there are some proposals for choosing the theoretical smoothness constant using data, their applicability is limited because they are often not equivariant to transformations of the data, or because they do not employ a consistent estimator of a well-defined quantity.

3.4.4 Shape-Constrained Methods

Monotone regression estimation and inference is a useful methodology whenever a regression function is known to be monotonic because it does not require a choice of bandwidth or kernel function at interior points (Groeneboom and Jongbloed, 2014, and reference therein). Motivated by this observation, Babii and Kumar (2021) study RD treatment effect estimation and inference under the assumption that the conditional expectation functions of potential outcomes are monotonic near the cutoff. Due to the inherent boundary bias in RD settings, however, some of the nice features of monotone regression estimators are lost: a choice of tuning parameter is still required for implementation, since otherwise the monotone regression procedures would not be valid at the cutoff. Nevertheless, within the continuity-based RD framework, their shape-constrained RD methodology can be of interest in settings where monotonicity of conditional expectations $\mathbb{E}[Y_i(1)|X_i = x]$ and $\mathbb{E}[Y_i(0)|X_i = x]$ near the cutoff is a reasonable assumption.

Hsu and Shen (2021b) also consider the role of monotonicity in RD designs. Within the continuity-based RD framework, they propose a nonparametric monotonicity test for conditional on observable characteristics RD treatment effects at the cutoff. Their testing procedure employs local polynomial methods and therefore requires a choice of bandwidth, and their implementation relies on undersmoothing techniques, which may affect size control of the proposed test. A robust bias-correction version of their testing procedure can be constructed by increasing the order of the polynomial used for inference (Section 3.2).

3.4.5 Other Methods

Frolich and Huber (2019) and Peng and Ning (2021) develop nonparametric covariate adjustments in RD designs where pre-treatment covariates are discontinuous at the cutoff, which leads to different RD-like parameters in general, while Gerard, Rokkanen, and Rothe (2020) propose partial identification methods in settings where the RD design is invalidated by endogenous one-sided sorting of units near the cutoff. Finally, Mukherjee, Banerjee, and Ritov (2021) augment the usual RD setup by including a second equation that assumes the score is a linear function of observed covariates to construct alternative semiparametric estimators of average RD treatment effects.

3.5 Discrete Score Variable

The defining feature of the RD design is the treatment assignment rule $\mathbb{1}(X_i \geq c)$, which in principle allows for both a discrete or continuous score X_i. A continuously distributed score implies that all units have a different score value in finite samples, and ensures that some observations will have scores arbitrarily close to the cutoff in large samples (provided the density of the score is positive around the cutoff). In contrast, a score with discrete support implies that multiple units will share the same value of X_i, leading to repeated values or "mass points" in the data. In the absence of a continuously distributed score, standard continuity-based RD treatment effects are not nonparametrically identifiable at the cutoff—this problem is equivalent to the lack of nonparametric identifiability in bunching designs (Blomquist, Newey, Kumar, and Liang, 2021). The main issue is that, with discrete scores, identification and estimation of continuity-based RD treatment effects would necessarily require extrapolation outside the support of the score.

There are different approaches for RD designs with discrete scores, depending on the assumptions imposed and the conceptual framework under consideration. To describe these methods, suppose the discrete score X_i can take $M = 2K+1$ distinct values $\{x_{-K}, \ldots, x_{-2}, x_{-1}, c, x_1, x_2, \ldots x_K\}$, where we assume the cutoff c is the median value only for simplicity. Within the continuity-based framework, the local polynomial methods discussed in Section 3.2 can still be employed when (i) the implicit parametric extrapolation from $X_i = x_{-1}$ to $X_i = c$ is accurate enough, and (ii) the number of unique values M is large. In such settings, conventional bandwidth selection, estimation, and inference methods would still be applicable. On the other hand, when the score has only a

29

few distinct values, continuity-based methods are not advisable for analysis of RD designs unless the underlying local polynomial regression model for the conditional expectations is believed to be correctly specified. See Lee and Card (2008) for early methods under this type of assumption.

Local randomization methods, on the other hand, remain valid under the same assumptions discussed in Section 3.3 when the score is discrete, and the estimation and inference methods in that section can be applied regardless of whether the number of distinct values of X_i is large or small. If the number of observations per mass point is large enough, the window selection step is not needed, as it suffices to use observations with $X_i = x_{-1}$ and $X_i = c$ to conduct estimation and inference using Fisherian, Neyman, or super-population methods. If the number of observations per mass point is not large enough so that a bigger window is needed, the window selection can be based on pre-determined covariates, but the implementation is easier because the nested windows considered can be increased in length one mass point at the time on either side of the cutoff.

In some cases, it may be desirable to redefine the parameter of interest. For example, within the continuity-based framework, the average RD treatment effect $\tau_{\mathsf{SRD}} = \mathbb{E}[Y_i(1) - Y_i(0)|X_i = c]$ is not identifiable without extrapolation (from the left of the cutoff), but the parameter $\tau_{\mathsf{SDS}} = \mathbb{E}[Y_i(1)|X_i = c] - \mathbb{E}[Y_i(0)|X_i = x_{-1}]$ is identifiable, making explicit the inability to identify and estimate $\mathbb{E}[Y_i(0)|X_i = c]$ without additional modelling assumptions. The new parameter τ_{SDS} may be more natural in cases where the score is inherently discrete, but its usefulness is ultimately application-specific.

In sum, from a practical perspective, a key consideration for RD analysis with discrete scores is the number of distinct values M in the support of the running variable. When M is large, continuity-based methods may still be useful, though the effective number of observations for the purposes of local polynomial estimation and inference will be the number of mass points (not the total number of units). When M is small, continuity-based methods are only useful under additional assumptions for extrapolation, while local randomization methods may be a preferable alternative. See Cattaneo, Idrobo, and Titiunik (2021, Section 4) for more details and a practical discussion.

Other specific issues in RD settings with discrete scores have been studied recently. Dong (2015) and Barreca, Lindo, and Waddell (2016) investigate the phenomenon of heaping, which occurs when the score variable is rounded so that units that initially had different score values appear in the dataset as having the same value. Kolesar and Rothe (2018) employ a uniform-in-bias method,

which requires specifying a theoretical bound on the "curvature" of a class of regression functions that are only identifiable at the mass points of the score; see also Section 3.4.3.

3.6 Power Calculations for Experimental Design

The discussion so far has focused on ex-post analysis of RD designs using either continuity or local randomization methods. In recent years, applied researchers and policy-makers have been increasingly interested in designing ex-ante program evaluations leveraging RD designs, where researchers have access to the scores of all units but must choose a subset of units for whom to collect outcomes due to budget considerations or other restrictions. This experimental RD setting requires computing power functions for RD hypothesis tests, which depend on the specific estimation and inference methods adopted. Schochet (2009) discusses power calculations and optimal sample size selection for experimental design when using parametric regression specifications for RD treatment effect estimation. Within the continuity-based framework, Cattaneo, Titiunik, and Vazquez-Bare (2019) develop power calculations and compute optimal sample sizes for ex-ante experimental design when employing RBC inference (Section 3.2). These power calculation results can also be used to compute Minimum Detectable Effects (MDE), which are useful for both ex-ante and ex-post analysis of RD designs. Importantly, for ex-post analysis, computing MDE effects is preferred because ex-post power calculations using observed effect sizes are often unreliable.

4 Validation and Falsification

Relative to other non-experimental designs, the RD design offers a wealth of design-specific validation and falsification methods to assess the plausibility of the core RD assumptions. Regardless of whether the continuity or local randomization framework is used, a fundamental step in any RD analysis is to provide such supporting empirical evidence. This evidence can only be indirect, because the core RD assumptions are inherently untestable. Nonetheless, there are important steps that can be taken to assess their plausibility in most settings, and hence the potential validity of the RD design, as well as the overall robustness of the main empirical findings. We first discuss generic falsification methods that are used in most program evaluation settings, and then turn to design-based methods that rely on specific RD features. See Cattaneo, Idrobo, and Titiunik (2020,

Section 5) for a recent practical overview and more details.

4.1 Pre-intervention Covariates and Placebo Outcomes

In randomized experiments, it is common to check whether the randomization produced similar control and treatment group distributions of observed covariates determined before treatment assignment. This so-called *covariate balance* analysis is usually implemented by testing the null hypothesis that the mean (or other feature of the distribution) of pre-determined covariates is the same in the treated and control groups. Rejecting the null hypothesis implies that the treated and control groups are not comparable in terms of pre-treatment characteristics, thereby casting doubts on the research design. The same idea can be applied in the RD design (Lee, 2008). In addition, the principle of covariate balance can be extended beyond pre-determined covariates to variables that are determined after the treatment is assigned but are known to be unaffected by the treatment, commonly referred to as "unaffected" (Rosenbaum, 2010, p. 121) or "placebo" outcomes. This kind of validation analysis can be powerful, as it relies on the hypothesized mechanism by which the treatment affects the outcome. For example, in their analysis of Head Start expansion on child mortality, Ludwig and Miller (2007, Table III, p. 180) estimated the RD effect on causes of death that should not be affected by Head Start programs and found such effect to be indistinguishable from zero, in addition to also considering other pre-intervention covariates.

Analysis of pre-intervention covariates and placebo outcomes is a fundamental component of the validation and falsification toolkit for RD designs. Its application is straightforward: local polynomial RBC inference methods (Section 3.2) and local randomization inference methods (Section 3.3) can directly be used upon substituting the outcome variable of interest by the pre-intervention covariates and placebo outcomes. In the usual implementation of balance tests, failure to reject the null hypothesis of covariate balance is interpreted as evidence of comparability. However, it may be more appropriate to test the null hypothesis that the covariate distributions are different and thus control the probability of concluding that the groups are similar when they are not (rather than controlling the probability of concluding that they are different when they are similar, as in the usual approach). See Hartman (2020) for an application in RD settings using RBC inference in the continuity-based framework.

Finally, other methods for pre-intervention covariates and placebo outcomes testing have been

proposed within the continuity framework. First, Shen and Zhang (2016) propose local polynomial distributional tests using an undersmoothed bandwidth for implementation, but a more robust version using RBC inference can be constructed by increasing the polynomial order when implemented (Section 3.2). Second, Canay and Kamat (2018) discuss an approach based on asymptotic permutation tests, which leads to a randomization inference implementation (Section 3.3).

4.2 Density Continuity and Binomial Testing

McCrary (2008) proposed a density test for score "manipulation" near the cutoff, where the null hypothesis is that the density function of the score is continuous at the cutoff. While continuity of the density function of the score at the cutoff is neither necessary nor sufficient for validity of an RD design, a discontinuous density can be heuristically associated with the idea of "endogenous sorting" of units around the cutoff. Furthermore, under additional assumptions, such discontinuity can be formally linked to failure of the RD design. For example, Urquiola and Verhoogen (2009) and Bajari, Hong, Park, and Town (2011) give economic models that predict endogenous sorting around the cutoff, while Crespo (2020) discusses sorting caused by administrative reasons rather than strategic individual behavior.

As a consequence, McCrary's density test has become one of the leading validation and falsification methods for RD design with a continuously distributed score. His density test was originally implemented using the local polynomial density estimator of Cheng, Fan, and Marron (1997), a two-step nonparametric procedure requiring two distinct bandwidth choices. More recently, Cattaneo, Jansson, and Ma (2020) introduced a local polynomial density estimator that requires only one bandwidth choice and is more robust near boundary points, which they use to provide a hypothesis test based on RBC inference, thereby offering an improved density test for RD applications.

Building on McCrary's idea, but originally motivated by the local randomization framework, Cattaneo, Titiunik, and Vazquez-Bare (2017) propose a binomial test for counts near the cutoff as an additional manipulation test. Unlike the continuity-based density test, the binomial test can be used when the score is continuous or discrete, and does not rely on asymptotic approximations, as the associated p-value is exact for a given assignment probability in a fixed window around the cutoff. This validation method offers a useful complement and robustness check for the continuity-based density test.

Within the continuity-based framework, there are other extensions and refinements for density testing. Otsu, Xu, and Matsushita (2013), Ma, Jales, and Yu (2020) and Jales, Ma, and Yu (2017) discuss density testing combining local polynomials and empirical likelihood methods; see Section 3.4.1. Frandsen (2017) proposes a manipulation test for discrete scores and Bugni and Canay (2021) propose a binomial-like test using an asymptotic argument to justify local randomization near the cutoff, both employing a uniform-in-bias approach that requires the manual choice of a theoretical smoothness constant; see Section 3.4.3.

4.3 Placebo Cutoffs and Continuity of Regression Functions

Another falsification method specific to the RD design looks at artificial (or placebo) cutoffs away from the real cutoff determining treatment assignment. The intuition is that, since the true cutoff is the only score value at which the probability of receiving the treatment changes discontinuously, it should also be the only score value at which the outcome changes discontinuously. This test assumes that there are no other policies (or changes) introduced at different score values that may affect the outcome of interest. More formally, this method can be viewed as hypothesis test for continuity of the conditional expectation functions of potential outcomes at values of the score equal to the placebo cutoff considered. For implementation, researchers usually choose a grid of artificial cutoff values, and repeat estimation and inference of the RD effect on the outcome of interest at each artificial cutoff value, using either local polynomial RBC inference methods (Section 3.2) or local randomization inference methods (Section 3.3). Importantly, in order to avoid treatment effect contamination, this validation method should be implemented for units below and above the cutoff separately. See Ganong and Jäger (2018) for a related approach in the context of kink RD designs.

4.4 Donut hole and Bandwidth Sensitivity

Another class of RD-specific falsification methods involves re-implementing estimation and inference for the RD treatment effect with different subsets of observations, as determined by either excluding those observations closest to the cutoff or varying the bandwidth used for estimation and inference. The so-called "donut hole" approach (Bajari, Hong, Park, and Town, 2011) investigates the sensitivity of the empirical conclusions by removing the few observations closest to the cutoff,

and then implements estimation and inference. Heuristically, the goal is to assess how much influence the few observations closest to the cutoff have on the overall empirical results. The intuition is that if there is endogenous sorting of units across the cutoff, such sorting might occur only among units whose scores are very close to the cutoff, and thus when those observations are excluded the RD treatment effect may change. More formally, this approach can be related to issues of local extrapolation properties of the specific method used for estimation and inference near the cutoff. See Dowd (2021) for the latter approach in the context of local polynomial RBC inference methods (Section 3.2).

The other validation method whether the empirical conclusions are sensitive to the choice of bandwidth or local randomization neighborhood. The idea is simply to re-estimate the RD treatment effect for bandwidths (or neighborhood lengths) that are smaller or larger than the one originally chosen. In the continuity-based framework, if the original bandwidth is MSE-optimal, considering much larger bandwidths is not advisable due to the implied misspecification bias. Similarly, in the local randomization framework, considering larger neighborhoods may not be justifiable and hence uninformative.

Implementation of the donut hole and bandwidth sensitivity approaches, researchers can rely on either local polynomial RBC inference methods (Section 3.2) or local randomization inference methods (Section 3.3). See Cattaneo, Idrobo, and Titiunik (2020, Section 5) for more details.

5   Conclusion

Since the seminal work of Thistlethwaite and Campbell (1960), the literature on identification, estimation, inference, and validation for RD designs has blossomed and matured. Thanks to the collective work of many scholars, current RD methodology provides a wide range of both practical and theoretical tools. The original analogy between RD designs and randomized experiments has been clarified and in many ways strengthened, the choice of local polynomial bandwidth or local randomization neighborhood has become data-driven and based on principled criteria, and there has been a plethora of validation and falsification methods exploiting the particular features of RD designs. It is now understood that best practices for estimation and inference in RD designs should be based on tuning parameter choices that are objective, principled, and data-driven, thereby

removing researchers' ability to engage in arbitrary specification searching, and thus leading to credible and replicable empirical conclusions.

Looking ahead, there are still several areas that require further methodological development. For example, having more robust methods for extrapolation of RD treatment effects can contribute to expand the external validity of the conclusions drawn from RD designs, which are otherwise limited by the local nature of the design. Such extrapolation methods can also help in the design of optimal policies, a topic that has not been explored in the RD literature yet. As another example, it has become more common for policy makers and industry researchers to partner with experimental scholars to plan RD designs ex-ante, often in data-complex settings. A more complete methodology for ex-ante experimental design exploiting RD-specific features can help not only in improving field work in practice, but also enabling extrapolation of RD treatment effects by careful sampling design and data collection. Another important avenue for future research in RD settings is related to successfully incorporating modern high-dimensional and machine learning methods. Such methodological enhancements can aid in principled discovery of heterogeneous treatment effects as well as in developing more robust estimation and inference methods in multi-dimensional RD designs and related program evaluation settings.

References

ABADIE, A., AND M. D. CATTANEO (2018): "Econometric Methods for Program Evaluation," *Annual Review of Economics*, 10, 465–503.

ANGRIST, J. D., AND M. ROKKANEN (2015): "Wanna Get Away? Regression Discontinuity Estimation of Exam School Effects Away from the Cutoff," *Journal of the American Statistical Association*, 110(512), 1331–1344.

ARAI, Y., Y. HSU, T. KITAGAWA, I. MOURIFIÉ, AND Y. WAN (2021): "Testing Identifying Assumptions in Fuzzy Regression Discontinuity Designs," *Quantitative Economics*.

ARAI, Y., AND H. ICHIMURA (2016): "Optimal bandwidth selection for the fuzzy regression discontinuity estimator," *Economic Letters*, 141(1), 103–106.

——— (2018): "Simultaneous Selection of Optimal Bandwidths for the Sharp Regression Discontinuity Estimator," *Quantitative Economics*, 9(1), 441–482.

ARAI, Y., T. OTSU, AND M. H. SEO (2021): "Regression Discontinuity Design with Potentially Many Covariates," working paper.

ARMSTRONG, T., AND M. KOLESAR (2020): "Simple and Honest Confidence Intervals in Nonparametric Regression," *Quantitative Economics*, 11(1), 1–39.

BABII, A., AND R. KUMAR (2021): "Isotonic Regression Discontinuity Designs," *Journal of Econometrics*.

BACKUS, G. E. (1989): "Confidence Set Inference with a Prior Quadratic Bound," *Geophysical Journal International*, 97(1), 119–150.

BAJARI, P., H. HONG, M. PARK, AND R. TOWN (2011): "Regression Discontinuity Designs with an Endogenous Forcing Variable and an Application to Contracting in Health Care," NBER Working Paper No. 17643.

BARRECA, A. I., J. M. LINDO, AND G. R. WADDELL (2016): "Heaping-Induced Bias in Regression-Discontinuity Designs," *Economic Inquiry*, 54(1), 268–293.

BARTALOTTI, O., AND Q. BRUMMET (2017): "Regression Discontinuity Designs with Clustered Data," in *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, volume 38)*, ed. by M. D. Cattaneo, and J. C. Escanciano, pp. 383–420. Emerald Group Publishing.

BARTALOTTI, O., Q. BRUMMET, AND S. DIETERLE (2021): "A Correction for Regression Discontinuity Designs with Group-specific Mismeasurement of the Running Variable," *Journal of Business & Economic Statistics*, 39(3), 833–848.

BARTALOTTI, O., G. CALHOUN, AND Y. HE (2017): "Bootstrap Confidence Intervals for Sharp Regression Discontinuity Designs," in *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, volume 38)*, ed. by M. D. Cattaneo, and J. C. Escanciano, pp. 421–453. Emerald Group Publishing.

BERTANHA, M. (2020): "Regression Discontinuity Design with Many Thresholds," *Journal of Econometrics*, 218(1), 216–241.

BERTANHA, M., AND G. W. IMBENS (2020): "External Validity in Fuzzy Regression Discontinuity Designs," *Journal of Business & Economic Statistics*, 38(3), 593–612.

BLACK, S. E. (1999): "Do better schools matter? Parental valuation of elementary education," *The quarterly journal of economics*, 114(2), 577–599.

BLOMQUIST, S., W. K. NEWEY, A. KUMAR, AND C.-Y. LIANG (2021): "On Bunching and Identification of the Taxable Income Elasticity," *Journal of Political Economy*, 129(8), 2320–2343.

BRANSON, Z., M. RISCHARD, L. BORNN, AND L. W. MIRATRIX (2019): "A Nonparametric Bayesian Methodology for Regression Discontinuity Designs," *Journal of Statistical Planning and Inference*, 202, 14–30.

BUGNI, F. A., AND I. A. CANAY (2021): "Testing Continuity of a Density via g-order Statistics in the Regression Discontinuity Design," *Journal of Econometrics*, 221(1), 138–159.

BUSSE, M., J. SILVA-RISSO, AND F. ZETTELMEYER (2006): "$1,000 Cash Back: The Pass-through of Auto Manufacturer Promotions," *American Economic Review*, 96(4), 1253–1270.

CAETANO, C., G. CAETANO, AND J. C. ESCANCIANO (2021): "Regression Discontinuity Design with Multivalued Treatments," Working paper, Universidad Carlos III.

CALONICO, S., M. D. CATTANEO, AND M. H. FARRELL (2018): "On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference," *Journal of the American Statistical Association*, 113(522), 767–779.

——— (2020): "Optimal Bandwidth Choice for Robust Bias Corrected Inference in Regression Discontinuity Designs," *Econometrics Journal*, 23(2), 192–210.

——— (2021): "Coverage Error Optimal Confidence Intervals for Local Polynomial Regression," arXiv:1808.01398.

CALONICO, S., M. D. CATTANEO, M. H. FARRELL, AND R. TITIUNIK (2019): "Regression Discontinuity Designs using Covariates," *Review of Economics and Statistics*, 101(3), 442–451.

CALONICO, S., M. D. CATTANEO, AND R. TITIUNIK (2014): "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 82(6), 2295–2326.

——— (2015): "Optimal Data-Driven Regression Discontinuity Plots," *Journal of the American Statistical Association*, 110(512), 1753–1769.

CANAY, I. A., AND V. KAMAT (2018): "Approximate Permutation Tests and Induced Order Statistics in the Regression Discontinuity Design," *Review of Economic Studies*, 85(3), 1577–1608.

CARD, D., D. S. LEE, Z. PEI, AND A. WEBER (2015): "Inference on Causal Effects in a Generalized Regression Kink Design," *Econometrica*, 83(6), 2453–2483.

——— (2017): "Regression Kink Design: Theory and Practice," in *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, volume 38)*, ed. by M. D. Cattaneo, and J. C. Escanciano, pp. 341–382. Emerald Group Publishing.

CARD, D., A. MAS, AND J. ROTHSTEIN (2008): "Tipping and the Dynamics of Segregation," *Quarterly Journal of Economics*, 123(1), 177–218.

CATTANEO, M. D., AND J. C. ESCANCIANO (2017): *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, volume 38)*. Emerald Group Publishing.

CATTANEO, M. D., B. FRANDSEN, AND R. TITIUNIK (2015): "Randomization Inference in the Regression Discontinuity Design: An Application to Party Advantages in the U.S. Senate," *Journal of Causal Inference*, 3(1), 1–24.

CATTANEO, M. D., N. IDROBO, AND R. TITIUNIK (2020): *A Practical Introduction to Regression Discontinuity Designs: Foundations*. Cambridge University Press.

——— (2021): *A Practical Introduction to Regression Discontinuity Designs: Extensions*. Cambridge University Press (to appear).

CATTANEO, M. D., M. JANSSON, AND X. MA (2020): "Simple Local Polynomial Density Estimators," *Journal of the American Statistical Association*, 115(531), 1449–1455.

CATTANEO, M. D., L. KEELE, AND R. TITIUNIK (2021): "Covariate Adjustment in Regression Discontinuity Designs," in *Handbook of Matching and Weighting in Causal Inference*, ed. by D. S. S. J. R. Zubizarreta, E. A. Stuart, and P. R. Rosenbaum. Chapman & Hall.

CATTANEO, M. D., L. KEELE, R. TITIUNIK, AND G. VAZQUEZ-BARE (2016): "Interpreting Regression Discontinuity Designs with Multiple Cutoffs," *Journal of Politics*, 78(4), 1229–1248.

——— (2021): "Extrapolating Treatment Effects in Multi-Cutoff Regression Discontinuity Designs," *Journal of the American Statistical Association*, forthcoming.

CATTANEO, M. D., R. TITIUNIK, AND G. VAZQUEZ-BARE (2017): "Comparing Inference Approaches for RD Designs: A Reexamination of the Effect of Head Start on Child Mortality," *Journal of Policy Analysis and Management*, 36(3), 643–681.

——— (2019): "Power Calculations for Regression Discontinuity Designs," *Stata Journal*, 19(1), 210–245.

——— (2020): "The Regression Discontinuity Design," in *Handbook of Research Methods in Political Science and International Relations*, ed. by L. Curini, and R. J. Franzese, chap. 44, pp. 835–857. Sage Publications.

CATTANEO, M. D., AND G. VAZQUEZ-BARE (2016): "The Choice of Neighborhood in Regression Discontinuity Designs," *Observational Studies*, 2, 134–146.

CELLINI, S. R., F. FERREIRA, AND J. ROTHSTEIN (2010): "The Value of School Facility Investments: Evidence from a Dynamic Regression Discontinuity Design," *Quarterly Journal of Economics*, 125(1), 215–261.

CERULLI, G., Y. DONG, A. LEWBEL, AND A. POULSEN (2017): "Testing Stability of Regression Discontinuity Models," in *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, volume 38)*, ed. by M. D. Cattaneo, and J. C. Escanciano, pp. 317–339. Emerald Group Publishing.

CHAPLIN, D. D., T. D. COOK, J. ZUROVAC, J. S. COOPERSMITH, M. M. FINUCANE, L. N. VOLLMER, AND R. E. MORRIS (2018): "The Internal and External Validity of the Regression Discontinuity Design: A Meta-analysis of 15 Within-study Comparisons," *Journal of Policy Analysis and Management*, 37(2), 403–429.

CHEN, H., H. D. CHIANG, AND Y. SASAKI (2020): "Quantile Treatment Effects in Regression Kink Designs," *Econometric Theory*, 36(6), 1167–1191.

CHENG, M.-Y., J. FAN, AND J. S. MARRON (1997): "On Automatic Boundary Corrections," *Annals of Statistics*, 25(4), 1691–1708.

CHIANG, H. D., Y.-C. HSU, AND Y. SASAKI (2019): "Robust Uniform Inference for Quantile Treatment Effects in Regression Discontinuity Designs," *Journal of Econometrics*, 211(2), 589–618.

CHIANG, H. D., AND Y. SASAKI (2019): "Causal Inference by Quantile Regression Kink Designs," *Journal of Econometrics*, 210(2), 405–433.

CHIB, S., AND L. JACOBI (2016): "Bayesian Fuzzy Regression Discontinuity Analysis and Returns to Compulsory Schooling," *Journal of Applied Econometrics*, 31(6), 1026–1047.

CHOI, J.-Y., AND M.-J. LEE (2018): "Relaxing Conditions for Local Average Treatment Effect in Fuzzy Regression Discontinuity," *Economics Letters*, 173, 47–50.

COOK, T. D. (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.

COOK, T. D., AND V. C. WONG (2008): "Empirical Tests of the Validity of the Regression Discontinuity Design," *Annales d'Economie et de Statistique*, 91/92, 127–150.

CRESPO, C. (2020): "Beyond Manipulation: Administrative Sorting in Regression Discontinuity Designs," *Journal of Causal Inference*, 8(1), 164–181.

DAVEZIES, L., AND T. LE BARBANCHON (2017): "Regression Discontinuity Design with Continuous Measurement Error in the Running Variable," *Journal of Econometrics*, 200(2), 260–281.

DE LA CUESTA, B., AND K. IMAI (2016): "Misunderstandings About the Regression Discontinuity Design in the Study of Close Elections," *Annual Review of Political Science*, 19, 375–396.

DE MAGALHAES, L., D. HANGARTNER, S. HIRVONEN, J. MERILÄINEN, N. A. RUIZ, AND J. TUKIAINEN (2020): "How Much Should We Trust Regression Discontinuity Design Estimates? Evidence from Experimental Benchmarks of the Incumbency Advantage," SSRN working paper 3664069.

DELL, M. (2010): "The Persistent Effects of Peru's Mining Mita," *Econometrica*, 78(6), 1863–1903.

DONG, Y. (2015): "Regression Discontinuity Applications with Rounding Errors in the Running Variable," *Journal of Applied Econometrics*, 30(3), 422–446.

———— (2018a): "Alternative Assumptions to Identify LATE in Fuzzy Regression Discontinuity Designs," *Oxford Bulletin of Economics and Statistics*, 80(5), 1020–1027.

———— (2018b): "Jump or Kink? Regression Probability Jump and Kink Design for Treatment Effect Evaluation," Working paper, UC-Irvine.

———— (2019): "Regression Discontinuity Designs with Sample Selection," *Journal of Business & Economic Statistics*, 37(1), 171–186.

DONG, Y., Y.-Y. LEE, AND M. GOU (2021): "Regression Discontinuity Designs with a Continuous Treatment," *Journal of the American Statistical Association*, forthcoming.

DONG, Y., AND A. LEWBEL (2015): "Identifying the Effect of Changing the Policy Threshold in Regression Discontinuity Models," *Review of Economics and Statistics*, 97(5), 1081–1092.

DONOHO, D. L. (1994): "Statistical Estimation and Optimal Recovery," *Annals of Statistics*, 22(1), 238–270.

DOWD, C. (2021): "Donuts and Distant CATEs: Derivative Bounds for RD Extrapolation," *Working paper*.

ECKLES, D., N. IGNATIADIS, S. WAGER, AND H. WU (2021): "Noise-Induced Randomization in Regression Discontinuity Designs," arXiv:2004.09458.

FAN, J., AND I. GIJBELS (1996): *Local Polynomial Modelling and Its Applications*. Chapman & Hall/CRC, New York.

FEIR, D., T. LEMIEUX, AND V. MARMER (2016): "Weak Identification in Fuzzy Regression Discontinuity Designs," *Journal of Business & Economic Statistics*, 34(2), 185–196.

FRANDSEN, B. (2017): "Party Bias in Union Representation Elections: Testing for Manipulation in the Regression Discontinuity Design When the Running Variable is Discrete," in *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, volume 38)*, ed. by M. D. Cattaneo, and J. C. Escanciano, pp. 281–315. Emerald Group Publishing.

FRANDSEN, B., M. FROLICH, AND B. MELLY (2012): "Quantile Treatments Effects in the Regression Discontinuity Design," *Journal of Econometrics*, 168(2), 382–395.

FROLICH, M., AND M. HUBER (2019): "Including Covariates in the Regression Discontinuity Design," *Journal of Business & Economic Statistics*, 37(4), 736–748.

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

GANONG, P., AND S. JÄGER (2018): "A Permutation Test for the Regression Kink Design," *Journal of the American Statistical Association*, 113(522), 494–504.

GELMAN, A., J. B. CARLIN, H. S. STERN, D. B. DUNSON, A. VEHTARI, AND D. B. RUBIN (2013): *Bayesian Data Analysis*. CRC press.

GELMAN, A., AND G. W. IMBENS (2019): "Why High-Order Polynomials Should Not be Used in Regression Discontinuity Designs," *Journal of Business & Economic Statistics*, 37(3), 447–456.

GENELETTI, S., A. G. O'KEEFFE, L. D. SHARPLES, S. RICHARDSON, AND G. BAIO (2015): "Bayesian Regression Discontinuity Designs: Incorporating Clinical Knowledge in the Causal Analysis of Primary Care Data," *Statistics in Medicine*, 34(15), 2334–2352.

GERARD, F., M. ROKKANEN, AND C. ROTHE (2020): "Bounds on Treatment Effects in Regression Discontinuity Designs with a Manipulated Running Variable," *Quantitative Economics*, 11(3), 839–870.

GREMBI, V., T. NANNICINI, AND U. TROIANO (2016): "Do Fiscal Rules Matter? A Difference-in-Discontinuities Design," *American Economic Journal: Applied Economics*, 8(3), 1–30.

GROENEBOOM, P., AND G. JONGBLOED (2014): *Nonparametric Estimation under Shape Constraints*. Cambridge University Press.

HAHN, J., P. TODD, AND W. VAN DER KLAAUW (2001): "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 69(1), 201–209.

HARTMAN, E. (2020): "Equivalence Testing for Regression Discontinuity Designs," *Political Analysis*.

HAUSMAN, C., AND D. S. RAPSON (2018): "Regression Discontinuity in Time: Considerations for Empirical Applications," *Annual Review of Resource Economics*, 10, 533–552.

HE, Y. (2018): "Three Essays on Regression Discontinuity Design and Partial Identification," *Ph.D. thesis, Iowa State University*.

HE, Y., AND O. BARTALOTTI (2020): "Wild Bootstrap for Fuzzy Regression Discontinuity Designs: Obtaining Robust Bias-corrected Confidence Intervals," *Econometrics Journal*, 23(2), 211–231.

HSU, Y.-C., AND S. SHEN (2019): "Testing Treatment Effect Heterogeneity in Regression Discontinuity Designs," *Journal of Econometrics*, 208(2), 468–486.

——— (2021a): "Dynamic Regression Discontinuity under Treatment Effect Heterogeneity," Working paper, UC-Davis.

——— (2021b): "Testing Monotonicity of Conditional Treatment Effects under Regression Discontinuity Designs," *Journal of Applied Econometrics*, 36(3), 346–366.

HUANG, X., AND Z. ZHAN (2021): "Local Composite Quantile Regression for Regression Discontinuity," *arXiv preprint arXiv:2009.03716*.

HYYTINEN, A., J. MERILÄINEN, T. SAARIMAA, O. TOIVANEN, AND J. TUKIAINEN (2018): "When Does Regression Discontinuity Design Work? Evidence from Random Election Outcomes," *Quantitative Economics*, 9(2), 1019–1051.

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

IMBENS, G., AND S. WAGER (2019): "Optimized Regression Discontinuity Designs," *Review of Economics and Statistics*, 101(2), 264–278.

IMBENS, G. W., AND K. KALYANARAMAN (2012): "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *Review of Economic Studies*, 79(3), 933–959.

IMBENS, G. W., AND D. B. RUBIN (2015): *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

JALES, H., J. MA, AND Z. YU (2017): "Optimal Bandwidth Selection for Local Linear Estimation of Discontinuity in Density," *Economics Letters*, 153, 23–27.

JALES, H., AND Z. YU (2017): "Identification and Estimation using a Density Discontinuity Approach," in *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, volume 38)*, ed. by M. D. Cattaneo, and J. C. Escanciano, pp. 29–72. Emerald Group Publishing.

KAMAT, V. (2018): "On Nonparametric Inference in the Regression Discontinuity Design," *Econometric Theory*, 34(3), 694–703.

KARABATSOS, G., AND S. G. WALKER (2015): "A Bayesian Nonparametric Causal Model for Regression Discontinuity Designs," in *Nonparametric Bayesian inference in biostatistics*, pp. 403–421. Springer.

KEELE, L. J., S. LORCH, M. PASSARELLA, D. SMALL, AND R. TITIUNIK (2017): "An Overview of Geographically Discontinuous Treatment Assignments with an Application to Children's Health Insurance," in *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, volume 38)*, ed. by M. D. Cattaneo, and J. C. Escanciano, pp. 147–194. Emerald Group Publishing.

KEELE, L. J., AND R. TITIUNIK (2015): "Geographic Boundaries as Regression Discontinuities," *Political Analysis*, 23(1), 127–155.

KEELE, L. J., R. TITIUNIK, AND J. ZUBIZARRETA (2015): "Enhancing a Geographic Regression Discontinuity Design Through Matching to Estimate the Effect of Ballot Initiatives on Voter Turnout," *Journal of the Royal Statistical Society: Series A*, 178(1), 223–239.

KLEVEN, H. J. (2016): "Bunching," *Annual Review of Economics*, 8, 435–464.

KOLESAR, M., AND C. ROTHE (2018): "Inference in Regression Discontinuity Designs with a Discrete Running Variable," *American Economic Review*, 108(8), 2277–2304.

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

LEE, D. S., AND D. CARD (2008): "Regression Discontinuity Inference with Specification Error," *Journal of Econometrics*, 142(2), 655–674.

LEE, D. S., AND T. LEMIEUX (2010): "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 48(2), 281–355.

LEE, M.-J. (2017): "Regression Discontinuity with Errors in the Running Variable: Effect on Truthful Margin," *Journal of Econometric Methods*, 6(1).

LI, F., A. MATTEI, AND F. MEALLI (2015): "Evaluating the Causal Effect of University Grants on Student Dropout: Evidence from a Regression Discontinuity Design using Principal Stratification," *Annals of Applied Statistics*, 9(4), 1906–1931.

LUDWIG, J., AND D. L. MILLER (2007): "Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design," *Quarterly Journal of Economics*, 122(1), 159–208.

LV, X., X.-R. SUN, Y. LU, AND R. LI (2019): "Nonparametric Identification and Estimation of Dynamic Treatment Effects for Survival Data in a Regression Discontinuity Design," *Economics Letters*, 184, 108665.

MA, J., H. JALES, AND Z. YU (2020): "Minimum Contrast Empirical Likelihood Inference of Discontinuity in Density," *Journal of Business & Economic Statistics*, 38(4), 934–950.

MA, J., AND Z. YU (2020): "Coverage Optimal Empirical Likelihood Inference for Regression Discontinuity Design," *arXiv preprint arXiv:2008.09263*.

MATZKIN, R. L. (2013): "Nonparametric Identification in Structural Economic Models," *Annual Review of Economics*, 5, 457–486.

MCCRARY, J. (2008): "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test," *Journal of Econometrics*, 142(2), 698–714.

MEALLI, F., AND C. RAMPICHINI (2012): "Evaluating the Effects of University Grants by Using Regression Discontinuity Designs," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 175(3), 775–798.

MUKHERJEE, D., M. BANERJEE, AND Y. RITOV (2021): "Regression Discontinuity Design: Estimating the Treatment Effect with Standard Parametric Rate," *arXiv preprint arXiv:2102.11229*.

OTSU, T., K.-L. XU, AND Y. MATSUSHITA (2013): "Estimation and inference of discontinuity in density," *Journal of Business & Economic Statistics*, 31(4), 507–524.

———— (2015): "Empirical Likelihood for Regression Discontinuity Design," *Journal of Econometrics*, 186(1), 94–112.

OWEN, A. B. (2001): *Empirical likelihood*. Chapman and Hall/CRC.

OWEN, A. B., AND H. VARIAN (2020): "Optimizing the Tie-breaker Regression Discontinuity Design," *Electronic Journal of Statistics*, 14(2), 4004–4027.

PAPAY, J. P., J. B. WILLETT, AND R. J. MURNANE (2011): "Extending the regression-discontinuity approach to multiple assignment variables," *Journal of Econometrics*, 161(2), 203–207.

PEI, Z., D. S. LEE, D. CARD, AND A. WEBER (2021): "Local Polynomial Order in Regression Discontinuity Designs," *Journal of Business & Economic Statistics*, 31(4), 507–524.

PEI, Z., AND Y. SHEN (2017): "The Devil is in the Tails: Regression Discontinuity Design with Measurement Error in the Assignment Variable," in *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, volume 38)*, ed. by M. D. Cattaneo, and J. C. Escanciano, pp. 455–502. Emerald Group Publishing.

PENG, S., AND Y. NING (2021): "Regression Discontinuity Design under Self-selection," in *International Conference on Artificial Intelligence and Statistics*, pp. 118–126. PMLR.

PORTER, J. (2003): "Estimation in the Regression Discontinuity Model," working paper, University of Wisconsin.

PORTER, J., AND P. YU (2015): "Regression Discontinuity Designs with Unknown Discontinuity Points: Testing and Estimation," *Journal of Econometrics*, 189(1), 132–147.

QU, Z., AND J. YOON (2019): "Uniform Inference on Quantile Effects under Sharp Regression Discontinuity Designs," *Journal of Business & Economic Statistics*, 37(4), 625–647.

REARDON, S. F., AND J. P. ROBINSON (2012): "Regression Discontinuity Designs with Multiple Rating-Score Variables," *Journal of Research on Educational Effectiveness*, 5(1), 83–104.

ROKKANEN, M. (2015): "Exam Schools, Ability, and the Effects of Affirmative Action: Latent Factor Extrapolation in the Regression Discontinuity Design," *Unpublished manuscript*.

ROSENBAUM, P. R. (2010): *Design of Observational Studies*. Springer, New York.

SCHOCHET, P. Z. (2009): "Statistical Power for Regression Discontinuity Designs in Education Evaluations," *Journal of Educational and Behavioral Statistics*, 34(2), 238–266.

SEKHON, J. S., AND R. TITIUNIK (2016): "Understanding Regression Discontinuity Designs as Observational Studies," *Observational Studies*, 2, 174–182.

———— (2017): "On Interpreting the Regression Discontinuity Design as a Local Experiment," in *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, volume 38)*, ed. by M. D. Cattaneo, and J. C. Escanciano, pp. 1–28. Emerald Group Publishing.

SHEN, S., AND X. ZHANG (2016): "Distributional Tests for Regression Discontinuity: Theory and Empirical Examples," *Review of Economics and Statistics*, 98(4), 685–700.

Sun, Y. (2005): "Adaptive Estimation of the Regression Discontinuity Model," woking paper, UCSD.

Thistlethwaite, D. L., and D. T. Campbell (1960): "Regression-Discontinuity Analysis: An Alternative to the Ex-Post Facto Experiment," *Journal of Educational Psychology*, 51(6), 309–317.

Titiunik, R. (2021): "Natural Experiments," in *Advances in Experimental Political Science*, ed. by J. N. Druckman, and D. P. Gree, chap. 6, pp. 103–129. Cambridge University Press.

Tuvaandorj, P. (2020): "Regression Discontinuity Designs, White Noise Models, and Minimax," *Journal of Econometrics*, 218(2), 587–608.

Urquiola, M., and E. Verhoogen (2009): "Class-size caps, sorting, and the regression-discontinuity design," *American Economic Review*, 99(1), 179–215.

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

Wing, C., and T. D. Cook (2013): "Strengthening the Regression Discontinuity Design Using Additional Design Elements: A Within-Study Comparison," *Journal of Policy Analysis and Management*, 32(4), 853–877.

Wong, V. C., P. M. Steiner, and T. D. Cook (2013): "Analyzing Regression-Discontinuity Designs With Multiple Assignment Variables: A Comparative Study of Four Estimation Methods," *Journal of Educational and Behavioral Statistics*, 38(2), 107–141.

Xu, K.-L. (2017): "Regression Discontinuity with Categorical Outcomes," *Journal of Econometrics*, 201(1), 1–18.

——— (2018): "A Semi-Nonparametric Estimator of Regression Discontinuity Design with Discrete Duration Outcomes," *Journal of Econometrics*, 206(1), 258–278.