Annual Review of Economics

Regression Discontinuity Designs

Matias D. Cattaneo[1] and Rocío Titiunik[2]

[1]Department of Operations Research and Financial Engineering, Princeton University, Princeton, New Jersey, USA; email: cattaneo@princeton.edu

[2]Department of Politics, Princeton University, Princeton, New Jersey, USA; email: titiunik@princeton.edu

Annu. Rev. Econ. 2022. 14:821–51

The *Annual Review of Economics* is online at economics.annualreviews.org

https://doi.org/10.1146/annurev-economics-051520-021409

JEL codes: C10, C18, C21

Keywords

regression discontinuity, causal inference, program evaluation, treatment effects, nonexperimental methods

Abstract

The regression discontinuity (RD) design is one of the most widely used nonexperimental 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 topics: (*a*) designs and parameters, focusing on different types of RD settings and treatment effects of interest; (*b*) estimation and inference, presenting the most popular methods based on local polynomial regression and methods for the analysis of experiments, as well as refinements, extensions, and alternatives; and (*c*) validation and falsification, summarizing an array of mostly empirical approaches to support the validity of RD designs in practice.

1. INTRODUCTION

The regression discontinuity (RD) design was introduced by Thistlethwaite & Campbell (1960, p. 317) as a "method of testing causal hypotheses" 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 the control condition. Under appropriate assumptions, the discontinuous change in the probability of receiving treatment at the cutoff can be used to learn about the causal effects of the treatment on an outcome of interest.

Although overlooked at first, the RD design has become one of the most credible nonexperimental methods for causal inference and program evaluation. In economics, its use is widespread, driven by the wide applicability of the design across multiple subfields. For example, in development economics, cash transfers are often available to households that reach a threshold on a poverty measure. In labor economics, unemployment insurance or reemployment training is offered to individuals who are older than a certain age or who previously worked a minimum number of hours. In political economy, the incumbent party in two-party plurality elections is the candidate that reaches 50% of the vote. In public economics, municipalities with populations above certain thresholds are often subjected to different election or tax-sharing rules. RD designs can be employed to study the effect of such policies on various outcomes of interest, including health status, labor supply, fiscal indicators, and saving decisions, among many others. Given that many such interventions are not amenable to random experimentation, the RD design offers the opportunity to study treatment effects with a rigor and credibility that would otherwise be unavailable.

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 (Cook 2008, Imbens & Lemieux 2008, van der Klaauw 2008, Lee & 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 implementation issues, but refer the reader to Cattaneo et al. (2020a, 2022a) for comprehensive practical guides to empirical analysis. Although we mention examples in various sections, space constraints prevent us from presenting an exhaustive list of empirical applications of RD designs in economics.

We organize our review in three main parts. Section 2 introduces the types of RD designs most commonly encountered in practice and 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 it relies on basic restrictions on the underlying data-generating process—i.e., the statistical model assumed to generate the data. Using the taxonomy proposed by Cattaneo et al. (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 derived from the analysis of experiments. Using both frameworks, we provide an overview of canonical RD settings (sharp and fuzzy), multidimensional RD designs (multi-cutoff, multi-score, geographic, multiple-treatment, 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, and we overview methods for the extrapolation of RD treatment effects.

In Section 3, we focus on estimation and inference methods for RD designs, presenting the two most common approaches: local polynomial regression methods (Fan & Gijbels 1996), which are naturally motivated by the continuity framework, and analysis of experiments methods (Rosenbaum 2010, Imbens & Rubin 2015, Abadie & 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 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 lately received renewed attention among practitioners.

In Section 4, we focus on methods for the 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 analysis.

We conclude in Section 5. General purpose software in `Python`, `R`, and `Stata` for analysis of RD designs, as well as replication files for RD empirical illustrations, further readings, and other resources, is 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 only to units whose score value exceeds a known cutoff (also called threshold). 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, this 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 might be 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 & 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 US 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 threshold 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 et al. (2017) reanalyze this application using the modern methods discussed in Sections 3 and 4. Busse et al. (2006) provide another early application of the RD design in industrial organization.

Unlike other nonexperimental 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 ex-ante and be well defined, with a discontinuous probability of treatment assignment at the cutoff. Importantly, the treatment assignment rule is known and verifiable, and it is 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 nonexperimental strategies, thus endowing the RD design with superior credibility among nonexperimental methods. In particular, the RD treatment assignment rule is the basis for a collection of validation and falsification tests (see 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 (see Section 3) are based on these design-specific features.

The remainder of this section introduces canonical and extended RD designs, focusing on identifying assumptions for treatment effect parameters of interest by employing the potential outcomes framework (Imbens & 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 (i.e., 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 (i.e., some units assigned to treatment remain untreated and/or vice versa) but the researcher is only interested on the intention-to-treat effect of offering treatment. To formalize, suppose that n units, indexed by $i = 1, 2, \ldots, n$, have a score variable X_i. The RD treatment assignment rule is $T_i = \mathbb{1}(X_i \geq c)$, with T_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 $T_i = 1$) and $Y_i(0)$ is the outcome under control (i.e., when $T_i = 0$). The probability of treatment assignment changes abruptly at the cutoff from zero to one, and we have $\mathbb{P}[T_i = 1|X_i = x] = 0$ if $x < c$ and $\mathbb{P}[T_i = 1|X_i = x] = 1$ if $x \geq c$. This is the key distinctive feature of the RD design.

The observed data are $(Y_i, T_i, X_i)'$, $i = 1, 2, \ldots, n$, where $Y_i = T_iY_i(1) + (1 - T_i)Y_i(0)$ captures the fundamental problem of causal inference: $Y_i = Y_i(0)$ for units with $T_i = 0$ (i.e., with $X_i < c$) and $Y_i = Y_i(1)$ for units with $T_i = 1$ (i.e., with $X_i \geq c$). In other words, 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 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 this variable.

Hahn et al. (2001) introduced the continuity framework for canonical 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 two key identifying assumptions are that (*a*) 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, and (*b*) the density of the score near the cutoff is positive. These assumptions capture the idea that units barely below and barely 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 responses of treated and control units at the cutoff can be attributed to the treatment and interpreted as the causal average effect of the treatment at the cutoff, that is, for units with score variable $X_i = c$.

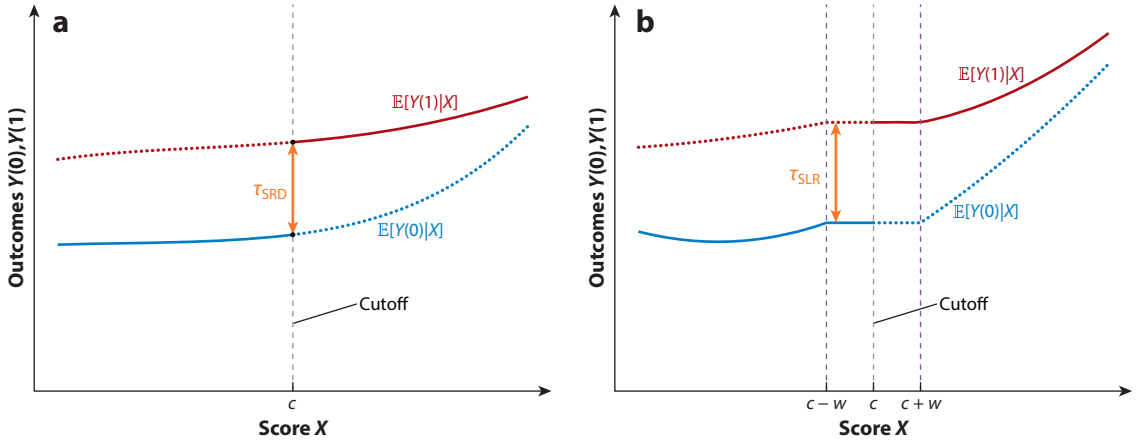

Figure 1

Schematic representation of regression discontinuity (RD) designs, showing (*a*) continuity-based RD design and (*b*) local randomization RD design. X denotes the score; c denotes the cutoff; w denotes the half-length of the local randomization window; $\mathbb{E}[Y(1)|X]$ and $\mathbb{E}[Y(0)|X]$ denote the conditional expectations of the potential outcomes under treatment and control, respectively, given the score; and τ_{SRD} and τ_{SLR} denote the RD treatment effect at the cutoff and in the window, respectively.

Formally, within the continuity-based framework, the canonical sharp RD (SRD) treatment effect is

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

which is represented graphically in **Figure 1*a***. The identity is the definition of the classical sharp RD parameter, while the 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 the data (see Section 3.2).

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

As a complement to the continuity framework, Cattaneo et al. (2015) introduced the local randomization framework for RD analysis. 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 & Campbell (1960), who viewed the RD 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 the latter 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. 675) variation in treatment status near the cutoff, as well as the work of Lee & Lemieux (2010).

The analogy between RD designs and randomized experiments was first formalized without continuity conditions by Cattaneo et al. (2015). Rather than relying on limits as the score tends to the cutoff and on heuristic analogies between units barely above and barely below the cutoff, this framework considers assumptions under 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 assumption behind the local randomization approach to RD analysis is that units near the cutoff are as-if randomly assigned to treatment. The assumption of as-if random assignment requires the existence of a window $\mathcal{W} = [c - w, c + w]$ around the cutoff c, assumed symmetric only for simplicity, such that two conditions hold for all units with $X_i \in \mathcal{W}$: (*a*) The joint probability distribution of scores in $\mathcal{W}$ is known, and (*b*) the potential outcomes are not affected by the score in $\mathcal{W}$. The first condition assumes that the assignment mechanism of the score is known inside the window; 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 (or their distributions, if assumed to be random variables) from being a function of the score inside $\mathcal{W}$.

The first condition is analogous to the requirement of known assignment mechanism in classical randomized experiments, whereas 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 pseudorandom number and then assigning the treatment or control conditions based on this number. If we interpret this pseudorandom 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, might 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, in a local randomization RD approach, the exclusion restriction inside $\mathcal{W}$ must be made explicit. Readers are referred to Cattaneo et al. (2017) for a relaxation of this restriction and to Sekhon & Titiunik (2016, 2017) for more discussion of the differences between randomized experiments and RD designs. De la Cuesta & Imai (2016) and Cattaneo et al. (2020c) further contrast the continuity and local randomization frameworks (see also the related papers in Cattaneo & Escanciano 2017).

The particular formalization of the local randomization framework can take different forms, depending on what approach from the analysis of experiments literature is adopted. The Fisherian approach assumes that the potential outcomes are nonrandom and the population is finite. The Neyman approach assumes that the potential outcomes are nonrandom but are sampled from an underlying infinite population. Finally, the so-called superpopulation approach assumes that the potential outcomes are a random sample. The original formulation of Cattaneo et al. (2015) adopted a Fisherian approach, but Cattaneo et al. (2017) expanded it to include the Neyman and superpopulation approaches.

The local randomization RD design is illustrated in **Figure 1*b***. The exclusion restriction assumption results in the regression functions $\mathbb{E}[Y_i(1)|X_i = x]$ and $\mathbb{E}[Y_i(0)|X_i = x]$ being constant for all values of the score inside the local randomization neighborhood $\mathcal{W}$; the average treatment effect is therefore the constant vertical distance between these functions inside $\mathcal{W}$. For the $N_{\mathcal{W}}$

units with $X_i \in \mathcal{W}$, the sharp local randomization (SLR) RD treatment effect can therefore be defined as

$$\tau_{\text{SLR}} \equiv \frac{1}{N_{\mathcal{W}}} \sum_{i:X_i \in \mathcal{W}} \mathbb{E}_{\mathcal{W}}[Y_i(1) - Y_i(0)]$$

$$= \frac{1}{N_{\mathcal{W}}} \sum_{i:X_i \in \mathcal{W}} \mathbb{E}_{\mathcal{W}}\Big[\frac{T_i Y_i}{\mathbb{P}_{\mathcal{W}}[T_i = 1]}\Big] - \frac{1}{N_{\mathcal{W}}} \sum_{i:X_i \in \mathcal{W}} \mathbb{E}_{\mathcal{W}}\Big[\frac{(1 - T_i) Y_i}{1 - \mathbb{P}_{\mathcal{W}}[T_i = 1]}\Big],$$

where $\mathbb{E}_{\mathcal{W}}$ and $\mathbb{P}_{\mathcal{W}}$ denote, respectively, expectation and probability computed conditionally for all units with $X_i \in \mathcal{W}$. This notation allows for both random and nonrandom potential outcomes as well as for countable or uncountable underlying populations, accommodating Fisherian, Neyman, and superpopulation settings. The estimand τ_{SLR} is identifiable and can be estimated from observable data, as discussed in Section 3.3.

Researchers sometimes incorporate additional (pre-intervention) covariates with the intention of restoring or improving the nonparametric identification of sharp RD parameters such as τ_{SRD} or τ_{SLR}. However, as in randomized experiments, it is not possible to restore the 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 (see Section 3). Similar considerations apply to the other RD designs discussed below. Calonico et al. (2019) provide a more detailed discussion, including how to introduce covariates without changing the RD parameters, and Cattaneo et al. (2022b) offer an overview of covariate-adjusted RD methods.

2.2. Fuzzy Designs

Fuzzy RD designs arise in settings where the treatment assigned and the treatment received do not coincide for at least some units—that is, in 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 an additional variable, D_i, representing the treatment received. Given the binary assignment $T_i = \mathbb{1}(X_i \geq c)$, each unit now has two binary potential treatments, $D_i(0)$ and $D_i(1)$, where $D_i(1)$ is the treatment received when the unit's score is above the cutoff, and $D_i(0)$ is the treatment received when the unit's score is below the cutoff. The observed treatment is therefore $D_i = D_i(0)(1 - T_i) + D_i(1)T_i$. The potential outcomes are generalized to $Y_i(T_i, D_i(T_i))$, leading to four different values depending on the values of T_i and the potential treatments $D_i(T_i)$. For example, $Y_i(0, 1)$ is the outcome that unit i would exhibit when their score is below the cutoff ($T_i = 0$) and they receive the treatment anyway ($D_i(0) = 1$).

Perfect compliance with treatment assignment means that $\mathbb{P}[D_i(0) = 0|X_i = x] = 1$ for $x < c$ (i.e., no units with score below the cutoff receive the treatment) and $\mathbb{P}[D_i(1) = 1|X_i = x] = 1$ for $x \geq c$ (i.e., all units with score above the cutoff receive the treatment); in this case the treatment assignment rule reduces to the sharp RD rule, $D_i = T_i = \mathbb{1}(X_i \geq c)$, and the four potential outcomes reduce to $Y_i(1, 1) := Y_i(1)$ and $Y_i(0, 0) := Y_i(0)$. Readers may consult Imbens & Rubin (2015) for a review of this potential outcomes framework.

As mentioned above, if the goal is to learn about the effect of assigning the treatment rather than of receiving it, researchers can simply follow a sharp RD analysis where X_i is the score, T_i is seen as the treatment of interest, and Y_i and D_i are the outcomes. This requires invoking standard sharp RD continuity assumptions for the regression functions $\mathbb{E}[Y_i(1, D_i(1))|X_i = x]$, $\mathbb{E}[Y_i(0, D_i(0))|X_i = x]$, $\mathbb{E}[D_i(1)|X_i = x]$, and $\mathbb{E}[D_i(0)|X_i = x]$.

However, in many cases researchers are also (or instead) interested in the effect of receiving the treatment on the outcome of interest. In general, identification of such treatment effects in settings

with imperfect compliance cannot be achieved with the same continuity or local randomization assumptions as in the sharp RD case, and it requires additional assumptions. 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 (FRD) estimand is

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

under minimal continuity conditions. The causal interpretation of τ_{FRD} depends on the additional assumptions invoked and the particular way in which treatment compliance is defined. For example, Hahn et al. (2001) showed that if the potential outcomes and the treatment received are independent conditional on the score being near the cutoff, then we have $\tau_{\text{FRD}} = \mathbb{E}[Y_i(1) - Y_i(0) | X = c]$, 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 is not commonly used to interpret the fuzzy RD estimand in a continuity-based approach.

Several authors have derived alternative identification results to interpret τ_{FRD} based on continuity conditions. Many of these approaches seek to characterize the parameter τ_{FRD} as a causal estimand for some subpopulation of compliers at the cutoff, heuristically defined as units who receive the treatment when they are assigned to the treatment condition and remain untreated when they are assigned to the control condition. The result typically requires a type of monotonicity condition, akin to that invoked in instrumental variables (IV) settings, which rules out units whose compliance decisions are the opposite of their assignments and is formalized differently depending on the setup. 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; Arai et al. (2022) instead 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 different definition of types of units is used by Bertanha & 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 by Imbens & Lemieux (2008), Frandsen et al. (2012), Cattaneo et al. (2016), and Choi & Lee (2018).

In addition to monotonicity, the interpretation of τ_{FRD} as the average effect of the treatment requires that the only effect of crossing the cutoff on the outcome occur via the treatment received—a condition analogous to the exclusion restriction in standard IV designs. In other words, conditional on the treatment received, the assignment of the treatment must have no effect on the outcomes. This assumption is violated, for example, when the mere information of having been assigned to the treatment causes participants to change their behaviors. In the continuity framework, the exclusion restriction is implied by the continuity assumptions imposed on the underlying conditional expectations at the cutoff.

Regardless of the particular assumptions adopted to achieve a causal interpretation for τ_{FRD}, most of these methods arrive at the same qualitative identification result: Under continuity of the appropriate functions and some form of monotonicity, 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 because it is analogous to IV results in the analyses of experiments with

noncompliance. The generic fuzzy local randomization (FLR) RD estimand can be written as

$$\tau_{\text{FLR}} = \frac{\vartheta_Y}{\vartheta_D},$$

where

$$\vartheta_Y = \frac{1}{N_{\mathcal{W}}} \sum_{i:X_i \in \mathcal{W}} \mathbb{E}_{\mathcal{W}}\left[\frac{T_i Y_i}{\mathbb{P}_{\mathcal{W}}[T_i = 1]}\right] - \frac{1}{N_{\mathcal{W}}} \sum_{i:X_i \in \mathcal{W}} \mathbb{E}_{\mathcal{W}}\left[\frac{(1 - T_i) Y_i}{1 - \mathbb{P}_{\mathcal{W}}[T_i = 1]}\right]$$

and

$$\vartheta_D = \frac{1}{N_{\mathcal{W}}} \sum_{i:X_i \in \mathcal{W}} \mathbb{E}_{\mathcal{W}}\left[\frac{T_i D_i}{\mathbb{P}_{\mathcal{W}}[T_i = 1]}\right] - \frac{1}{N_{\mathcal{W}}} \sum_{i:X_i \in \mathcal{W}} \mathbb{E}_{\mathcal{W}}\left[\frac{(1 - T_i) D_i}{1 - \mathbb{P}_{\mathcal{W}}[T_i = 1]}\right].$$

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

Analogously to the continuity case, if researchers are interested in the effect of the treatment assignment, they can conduct a local randomization sharp RD analysis where T_i is the treatment of interest and Y_i and D_i are the outcomes, which requires making local randomization assumptions for the potential outcomes as well as for the potential treatments.

2.3. Multidimensional 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 multidimensional RD designs. These generalized RD designs require specific modifications for interpretation and analysis. Conceptually, however, the core of the design remains the same: A discontinuous change in the probability of receiving treatment is the basis for comparing units assigned to treatment and to 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 et al. (2011) introduced the multi-score RD design (see also Reardon & Robinson 2012, Wong et al. 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 multidimensional but the treatment is still binary. For example, students may need to take a language exam and a mathematics exam and to 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. Readers are referred to

Black (1999) for a pioneer use of a geographic RD research design and to 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 & Titiunik (2015) using continuity-based ideas and by Keele et al. (2015) using local randomization ideas. Keele et al. (2017) discuss available options for analysis when the exact geographic location of each unit in the data set is unavailable.

Cattaneo et al. (2016) introduced the multi-cutoff RD design, whereby 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 $\mathcal{C} = \{c_1, c_2, \ldots, c_J\}$, where the values $c_1, c_2, \ldots, c_J$ can be either distinct cutoffs in an RD design or a discretization of a continuous support. 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 et al. 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 the local randomization framework. 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 & Titiunik (2015) and Cattaneo et al. (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. Readers may consult Cattaneo et al. (2022a) for a practical introduction to multi-score and multi-cutoff RD design methods.

In addition to the multi-cutoff and multi-score RD designs, there are several other types of RD designs with multidimensional features. Dong et al. (2021) consider an RD design with a continuous (rather than binary or multi-valued) treatment within the continuity framework, and they show that RD causal effects can be identified based on changes in the probability distribution of the continuous treatment at the cutoff. Caetano et al. (2021) investigate continuity-based identification of RD causal effects where the treatment is multi-valued with finite support. Grembi et al. (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; they call this design a difference-in-discontinuities design, building on ideas from the difference-in-differences design in program evaluation. Cellini et al. (2010) consider a different type of multidimensionality induced by a dynamic context in which the RD design occurs in multiple periods for the same units and the score is redrawn in every period, so that a unit may be assigned to treatment in one period and to control in future periods (see also Hsu & Shen 2021a for an econometric analysis of a dynamic RD design within the continuity-based framework). Lv et al. (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 design 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 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 & Rampichini (2012) and Wing & 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 & Lewbel (2015) and Cerulli et al. (2017) study local extrapolation methods via derivatives of the RD average treatment effect in the continuity-based framework, and Angrist & 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. In the context of continuity-based fuzzy designs, Bertanha & Imbens (2020) exploit variation in treatment assignment generated by imperfect 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 et al. (2016) discuss extrapolation in settings with ignorable cutoff assignment, and Cattaneo et al. (2021) discuss extrapolation allowing for heterogeneity across cutoffs that is invariant as a function of the score. In addition, within the continuity framework, Bertanha (2020) discusses the extrapolation of average RD treatment effects, assuming away cutoff-specific treatment effect heterogeneity. In the specific context of geographic RD designs, Keele & Titiunik (2015) and Keele et al. (2015) discuss extrapolation of average treatment effects along the geographic boundary that determines treatment assignment (see also Galiani et al. 2017 and Keele et al. 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 & Wong (2008) report a study of three empirical RD designs and discuss how they compare to experimental benchmarks created by randomized experiments of the same intervention. Readers are referred to Chaplin et al. (2018), Hyytinen et al. (2018), and De Magalhaes et al. (2020) for more recent results on RD recovery of experimental benchmarks.

2.5. Other Regression Discontinuity–Like Designs

The are several other research designs with features qualitatively similar 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, whereas others require specific modifications.

Card et al. (2015, 2017) introduce the regression kink (RK) design. Canonical and generalized RD designs assume 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 assignment rule that links the policy (or treatment) and the score is assumed to change slope at a known cutoff point

(also known as the kink point). The expectation is that the regression function of the observed outcome will be continuous at all values of the score, but its slope will be discontinuous at the cutoff point where the policy rule has the kink. 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 & Sasaki (2019) extend 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 are 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 et al. 2014, sections 3.1, 3.3).

Another area with interesting connections to the RD design is the literature on bunching and density discontinuities (Kleven 2016, Jales & Yu 2017). In this setting, the objects of interest are related to discontinuities and to other sharp changes in probability density functions, usually motivated by economic or other behavioral models (e.g., Bajari et al. 2011). These models typically exploit discontinuities 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 modeling assumptions that are invoked for extrapolation purposes. Blomquist et al. (2021) provide 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 & Rapson 2018). Although it is generally hard to reconcile those event studies with traditional RD designs, the local randomization RD framework can sometimes be adapted and used to analyze RD designs in time (see Sections 3.3, 3.5 for further discussion). Another RD-like example is provided by Porter & 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 at which behavior changes discontinuously (Card et al. 2008). Hansen (2017) explores a similar problem in the RK context.

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 Regression Discontinuity 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 the 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, and 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, of 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 et al. (2015) propose methods for implementing RD plots in a data-driven and disciplined way, and Cattaneo et al. (2020a, section 3) offer a practical introduction to RD plots. In addition, readers are referred to Pei et al. (2021) for a discussion of how to choose polynomial orders in RD designs to improve their implementation, which could be used in particular to further improve the implementation of RD plots.

Importantly, RD plots should not be used as substitutes for the formal estimation and inference of RD effects. In particular, sometimes the global polynomial fit will exhibit a jump at the cutoff, which might suggest the presence 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 et al. 2015, section 2.1.1). Furthermore, global polynomials for RD estimation and inference lead to several other undesirable features (Gelman & Imbens 2019), and therefore they are not recommended for analysis beyond visualization. The limitations of visual RD analysis are investigated experimentally by Korting et al. (2021), who find that changing the specification of RD plots while keeping the underlying model constant leads participants to draw different conclusions.

3.2. Local Polynomial Methods

The continuity-based identification result of Hahn et al. (2001) naturally justifies estimation of RD effects via local polynomial nonparametric methods (Fan & Gijbels 1996). These methods have become the standard for estimation and inference in the RD design under continuity conditions since the early works of Porter (2003) and Sun (2005) and the more recent works of Imbens & Kalyanaraman (2012) and Calonico et al. (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, 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 remainder 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) is also a reasonable choice in some empirical settings. [Readers may consult Pei et al. (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, which assigns the same weight to all observations. 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 Wald-type 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 with $X_i \in [c - h, c + h]$, 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 to always 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 & 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 et al. (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 et al. (2020) take a different approach and develop optimal bandwidth choices for inference: Their proposed bandwidth choices either (*a*) minimize the coverage error (CE) of confidence intervals or (*b*) optimally trade off coverage error and length of confidence intervals. The CE-optimal bandwidth choice can also be interpreted as optimally reducing the higher-order errors in the distributional approximation used for local polynomial statistical inference. Cattaneo & 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. We obtain

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

and

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

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 the 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: $\widehat{\tau}_{\text{SRD}}(h) = \widehat{\beta}_{+,0} - \widehat{\beta}_{-,0}$. Under standard assumptions, $\widehat{\tau}_{\text{SRD}}(h)$ will be consistent for $\tau_{\text{SRD}} = \mathbb{E}[Y_i(1) - Y_i(0)|X = c]$. It is common practice to use $\widehat{\tau}_{\text{SRD}}(h_{\text{MSE}})$, 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 that the polynomial expansion and the kernel function used capture the correct functional form of the unknown 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

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

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 that 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

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

with $\approx_d$ meaning 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 overrejection of the null hypothesis of no treatment effect ($\mathsf{H}_0 : \tau_{\text{SRD}} = 0$). For example, back-of-the-envelop calculations show that a nominal 95% confidence interval would have an empirical coverage of about 80%.

By choosing h smaller than h_{MSE}, the bias term B will vanish asymptotically, which is called undersmoothing in the nonparametrics literature. Although researchers could choose some

$h < h_{\text{MSE}}$ 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 et al. (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 (RBC) confidence interval is

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

where $\widehat{\mathsf{B}}$ denotes the estimated bias correction and $\widehat{\mathsf{W}}$ 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 more general 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 it 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 et al. 2020a, 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 et al. (2018, 2022) demonstrate the 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 in which the point estimation polynomial order p is increased when conducting inference without requiring any other changes (Calonico et al. 2014, remark 7), which allows to use 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 & Jäger 2018, Hyytinen et al. 2018, De Magalhaes et al. 2020).

The estimation and inference methods described so far extend naturally to fuzzy RD designs, where the local polynomial point estimator is 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 $\widehat{\tau}_{\text{FRD}}(h) = \widehat{\tau}_{\text{SRD}}(h)/\widehat{\varrho}_{\text{SRD}}(h)$, with $\widehat{\varrho}_{\text{SRD}}(h)$ denoting a sharp RD local polynomial estimate using the observed treatment D_i as outcome instead of Y_i—that is, $\widehat{\varrho}_{\text{SRD}}(h)$ is the sharp RD effect on receiving the treatment at the cutoff. Readers are referred to Calonico et al. (2014) for more details on fuzzy and higher-order derivative estimation and inference, and to Cattaneo et al. (2022a, 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 & Ichimura (2016, 2018) propose MSE-optimal bandwidth selection for different control and treatment bandwidths and develop RBC inference; Bartalotti et al. (2017) and He & Bartalotti (2020) investigate resampling methods and their connection with RBC inference; and Bartalotti & Brummet (2017) develop MSE-optimal bandwidth selection and

RBC inference in clustered sampling settings. Calonico et al. (2019) investigate the inclusion of pre-intervention covariates for efficiency purposes in RD designs, and they 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 et al. (2021) have 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 & Titiunik (2015), Cattaneo et al. (2016, 2021), and Bertanha (2020) address estimation and RBC inference in multi-score, geographic, and multi-cutoff RD designs; Dong et al. (2021) offer a complete analysis of estimation and RBC inference in an RD design with continuous treatment; and Chiang et al. (2019), Qu & Yoon (2019), Chen et al. (2020), and Huang & 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 et al. (2016) and He (2018) study methods robust to weak identification, whereas Arai et al. (2022) consider specification testing. In addition, Davezies & Le Barbanchon (2017), Lee (2017), Pei & Shen (2017), and Bartalotti et al. (2021) investigate issues of measurement error in various RD settings.

3.3. Local Randomization Methods

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, or superpopulation). These methods can be used with both continuous and discrete scores (see Section 3.5).

The window selection step is the main implementation challenge for local randomization analysis, and it is analogous to the bandwidth selection step in the continuity framework. Cattaneo et al. (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 assess the validity of randomized experiments. If the RD local randomization assumption holds, pre-treatment covariates 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 preset significance level. The procedure thus searches subsequently smaller windows until it finds the largest possible window such that the null hypothesis cannot be rejected for that window nor for any smaller window contained in it.

Once the window has been chosen, the second step is to deploy methods from the analysis of experiments to estimate RD treatment effects 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 $\binom{N_{\mathcal{W}}}{N_{\mathcal{W}}^+}^{-1}$, where $N_{\mathcal{W}}^+$ denotes the number of treatment units within the window $\mathcal{W}$, and

$N_{\mathcal{W}}^- = N_{\mathcal{W}} - N_{\mathcal{W}}^+$. Another common strategy is to assume that units were assigned independently to treated and control inside the window, with probability equal to $N_{\mathcal{W}}^+/N_{\mathcal{W}}$.

Point estimation of average effects can be based on simple difference-in-means for observations inside $\mathcal{W}$. The sharp and fuzzy parameters, τ_{SLR} and τ_{FLR}, can be respectively estimated with

$$\widehat{\tau}_{\text{SLR}} = \bar{Y}_{\mathcal{W}}^+ - \bar{Y}_{\mathcal{W}}^- \qquad \text{and} \qquad \widehat{\tau}_{\text{FLR}} = \frac{\bar{Y}_{\mathcal{W}}^+ - \bar{Y}_{\mathcal{W}}^-}{\bar{D}_{\mathcal{W}}^+ - \bar{D}_{\mathcal{W}}^-},$$

where

$$\bar{Y}_{\mathcal{W}}^+ = \frac{1}{N_{\mathcal{W}}} \sum_{i:X_i \in \mathcal{W}} \frac{\omega_i}{\mathbb{P}_{\mathcal{W}}[T_i = 1]} T_i Y_i, \qquad \bar{Y}_{\mathcal{W}}^- = \frac{1}{N_{\mathcal{W}}} \sum_{i:X_i \in \mathcal{W}} \frac{\omega_i}{1 - \mathbb{P}_{\mathcal{W}}[T_i = 1]} (1 - T_i) Y_i,$$

$$\bar{D}_{\mathcal{W}}^+ = \frac{1}{N_{\mathcal{W}}} \sum_{i:X_i \in \mathcal{W}} \frac{\omega_i}{\mathbb{P}_{\mathcal{W}}[T_i = 1]} T_i D_i, \qquad \bar{D}_{\mathcal{W}}^- = \frac{1}{N_{\mathcal{W}}} \sum_{i:X_i \in \mathcal{W}} \frac{\omega_i}{1 - \mathbb{P}_{\mathcal{W}}[T_i = 1]} (1 - T_i) D_i,$$

and ω_i denotes appropriate weights that are chosen according to the specific assumptions and framework employed. For example, if the assignment probabilities are not constant, then the observations must be weighted appropriately. In the Neyman framework with complete (i.e., fixed margins) randomization of treatment assignment, we have $\omega_i = 1$ and $\mathbb{P}_{\mathcal{W}}[T_i = 1] = N_{\mathcal{W}}^+/N_{\mathcal{W}}$ for all i, while under Bernoulli (i.e., coin flip) treatment assignment, we have $\omega_i = 1$ and $\mathbb{P}_{\mathcal{W}}[T_i = 1] = p$ (known) for all i. Finally, in the superpopulation framework, the previous estimators are unbiased, but in practice it is common to use the (consistent but biased) estimator that sets $\omega_i = \frac{\mathbb{P}_{\mathcal{W}}[T_i=1]N_{\mathcal{W}}}{N_{\mathcal{W}}^+} T_i + \frac{(1-\mathbb{P}_{\mathcal{W}}[T_i=1])N_{\mathcal{W}}}{N_{\mathcal{W}}^-}(1 - T_i)$. Cattaneo et al. (2017) discuss methods incorporating covariate adjustments.

Methods for inference also depend on the framework (Fisherian, Neyman, or superpopulation) chosen for the analysis. These frameworks, which were developed for the analysis of experiments, differ in the assumptions they make about sampling and in the null hypotheses they consider. In the Fisherian framework, the observations in the study are seen as the population of interest, not as a random sample from a larger population. As a consequence, the potential outcomes are seen as fixed, nonstochastic quantities, and the only randomness in the model stems from the random assignment of the treatment. The null hypothesis of interest is the so-called Fisherian sharp null hypothesis, under which every unit's potential outcome under treatment is equal to that unit's potential outcome under control, and the treatment effect is thus zero for all units. Under this null hypothesis, the randomization distribution of the treatment assignment can be used to impute all potential outcomes for every unit and thus to derive the null distribution of any test statistic of interest, which is then used to calculate exact p-values and can be extended to calculate confidence intervals under additional assumptions. Inferences are therefore finite-sample exact and do not rely on asymptotic approximations.

In the Neyman framework, the potential outcomes are also seen as nonstochastic, but, in contrast to the Fisherian framework, the null hypothesis is that the average treatment effect is zero, not that the individual treatment effect is zero for every unit. Unlike Fisher's sharp null hypothesis, Neyman's null hypothesis does not allow for full imputation of all potential outcomes, so exact finite-sample inferences are unavailable. Instead, assuming that the fixed potential outcomes are drawn from a large population, the law of large numbers and the central limit theorem are used to justify the consistency of the usual difference-in-means estimators and the validity (albeit conservative) of confidence intervals based on the normal distributional approximation.

Finally, in the superpopulation framework, the observations in the study are seen as a random sample taken from a larger population. The potential outcomes are therefore independent and identically distributed random variables, not fixed quantities. As in the Neyman approach, the

canonical null hypothesis of interest is that the average effect is zero, and inferences are also based on the normal distribution via large-sample approximations. Although the Neyman and superpopulation approaches differ conceptually, both approaches are usually implemented similarly in practice. In contrast, the implementation of Fisherian inference relies on considering all possible realizations of the treatment assignment, and it is usually achieved via permutation methods.

Fisherian methods are particularly useful when the number of 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 particularly well suited for analyzing RD designs because it allows researchers to include only the few observations that are closest to the cutoff. As a consequence, Fisherian inference offers a good complement and robustness check to local polynomial methods (see Section 3.2), which typically employ a larger number of 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 the Neyman or superpopulation framework. Readers may consult Cattaneo et al. (2022a, section 2) for a practical discussion of how to use these methods for RD analysis.

The local randomization approach has been applied in other RD settings. Keele et al. (2015) consider multi-score and geographic RD designs, Li et al. (2015) analyze principal stratification in a fuzzy RD setup, Ganong & Jäger (2018) study kink RD designs, and Cattaneo et al. (2016, 2021) investigate multi-cutoff RD designs. In addition, Owen & Varian (2020) study a "tie-breaker" design where the treatment is randomly assigned in a small neighborhood of the cutoff but follows the RD assignment otherwise, and Eckles et al. (2021) consider a setup in which exogenous measurement error is assumed to induce local randomization near the cutoff. Hyytinen et al. (2018) and De Magalhaes et al. (2020) conduct 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 et al. (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. However, a more powerful and robust bias-corrected version of their method can be constructed by increasing the order of the polynomial used (see 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 & Yu (2020) have investigated the higher-order properties of local polynomial empirical likelihood inference in RD designs, building on ideas developed by Calonico et al. (2018, 2022) for RBC inference using local polynomial methods. In addition, Otsu et al. (2013) and Ma et al. (2020) explore empirical likelihood inference methods for 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 generally developed within the local randomization framework. Geneletti et al. (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 & Walker (2015) and Chib & Jacobi (2016) propose alternative Bayesian models for sharp and fuzzy RD designs, paying particular attention to issues of misspecification of unknown regression functions. Branson et al. (2019) employ Gaussian process regression methods that can be interpreted as a Bayesian analog of the local linear RD estimator. Their approach is motivated by Bayesian methods, but they also discuss some frequentist properties of the resulting estimation procedures. 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 $\mathbb{E}[Y_i(0)|X_i = x]$) (see Gelman et al. 2013 for a general introduction to Bayesian methodology).

3.4.3. Uniform-in-bias methods. Imbens & Wager (2019) and Armstrong & Kolesar (2020) apply the ideas of Backus (1989) and Donoho (1994) to develop RD inference methods that are uniformly valid over a smoothness class of functions that determine 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 under treatment and control—$\mathbb{E}[Y_i(1)|X_i = x]$ and $\mathbb{E}[Y_i(0)|X_i = x]$, respectively—near the cutoff. Based on the smoothness constant that is chosen, the 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.

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 the 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. Although there are some proposals for choosing the theoretical smoothness constant using data, their applicability is limited because they are not equivariant to transformations of the data, do not employ a consistent estimator of a well-defined quantity, or are based on visual inspection of the data by the researcher.

3.4.4. Shape-constrained methods. Monotone regression estimation and inference are 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 (see Groeneboom & Jongbloed 2014 and references therein). Motivated by this observation, Babii & 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 positive 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, the shape-constrained RD methodology can be of interest in settings where monotonicity of the regression functions $\mathbb{E}[Y_i(1)|X_i = x]$ and $\mathbb{E}[Y_i(0)|X_i = x]$ near the cutoff is a reasonable assumption.

Hsu & Shen (2021b) also consider the role of monotonicity in RD designs. Within the continuity-based RD framework, they propose a nonparametric monotonicity test for RD treatment effects at the cutoff conditional on observable characteristics. Their testing procedure employs local polynomial methods and therefore requires a choice of bandwidth, and their

implementation relies on undersmoothing techniques that may affect size control of the proposed test. A robust bias-corrected version of their testing procedure can be constructed by increasing the order of the polynomial used for inference (see Section 3.2).

3.4.5. Other methods. Frolich & Huber (2019) and Peng & Ning (2021) consider nonparametric covariate adjustments in RD designs where pre-treatment covariates are discontinuous at the cutoff, which leads to different RD-like parameters in general. Gerard et al. (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 et al. (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 and a continuous score X_i. A continuously distributed score implies that all units have a different score value in finite samples, and it 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 et al. 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 X_i is discrete and 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. On the one hand, within the continuity-based framework, the local polynomial methods discussed in Section 3.2 can still be employed when (*a*) the implicit parametric extrapolation from $X_i = x_{-1}$ to $X_i = c$ is accurate enough, and (*b*) the number of unique values, M, is large. In such settings, conventional bandwidth selection, estimation, and inference methods might still be applicable. When the score has only a 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 & 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 superpopulation methods. If the number of observations per mass point is not large enough and a bigger window is needed, the window selection can be based on predetermined 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 $\tau_{\text{SRD}} = \mathbb{E}[Y_i(1) - Y_i(0)|X_i = c]$

is not identifiable without extrapolation (from the left of the cutoff); but the parameter $\tau_{\text{SDS}} = \mathbb{E}[Y_i(1)|X_i = c] - \mathbb{E}[Y_i(0)|X_i = x_{-1}]$ (where SDS stands for sharp discrete store) is always identifiable, making explicit the inability to identify and estimate $\mathbb{E}[Y_i(0)|X_i = c]$ without additional modeling assumptions. The new parameter τ_{SDS} may be more natural in cases in which 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, and local randomization methods may be a preferable alternative (see Cattaneo et al. 2022a, section 4, for more details and a practical discussion).

There are other specific approaches for RD settings with discrete scores. Dong (2015) and Barreca et al. (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 data set as having the same value. Kolesar & Rothe (2018) employ a uniform-in-bias method, which requires specifying a theoretical bound on the curvature of a class of regression functions that is only identifiable at the mass points of the score (see 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-based or local randomization methods. In recent years, applied researchers and policy makers have been increasingly interested in designing ex-ante program evaluations that leverage RD designs, in which 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 using parametric regression specifications for RD treatment effect estimation. Within the continuity-based framework, Cattaneo et al. (2019) develop power calculations and compute optimal sample sizes for ex-ante experimental design when employing RBC inference (see 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. Bulus (2022) derives MDE formulas for clustered RD designs under parametric assumptions. Importantly, computing MDE effects is preferred for ex-post analysis because ex-post power calculations using observed effect sizes are often unreliable.

4. VALIDATION AND FALSIFICATION

Relative to other nonexperimental 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 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 evaluate 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 we turn to design-based methods that rely on specific RD features. Readers are referred to Cattaneo et al. (2020a, 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 distributions of observed covariates for the control and treatment groups, where the covariates are determined before the treatment is assigned. This so-called covariate balance analysis is usually implemented by testing the null hypothesis that the mean (or other features of the distribution) of predetermined covariates is the same in the group of units assigned to treatment and the group of units assigned to control. Rejecting the null hypothesis implies that the treatment 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 to the RD design (Lee 2008). In addition, the principle of covariate balance can be extended beyond predetermined 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 the effects of Head Start expansion on child mortality, Ludwig & Miller (2007, p. 180, table III) estimated the RD effect on causes of death that should not have been affected by Head Start programs and found such effect to be indistinguishable from zero; they also considered 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 (see Section 3.2) and local randomization inference methods (see Section 3.3) can be used directly upon substituting the outcome variable of interest with the pre-intervention covariates or placebo outcome. In the usual implementation of balance tests, failure to reject the null hypothesis of covariate balance is interpreted as evidence of comparability of control and treatment groups near the cutoff. 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 the probability of concluding that they are different when they are similar, as in the usual approach). Hartman (2021) provides an application in RD settings using RBC inference in the continuity-based framework.

Finally, other methods for testing pre-intervention covariates and placebo outcomes have been proposed within the continuity framework. Shen & 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 (see Section 3.2). Canay & Kamat (2018) discuss an approach based on asymptotic permutation tests, which leads to a randomization inference implementation (see 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 the 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 & Verhoogen (2009) and Bajari et al. (2011) study economic models that predict endogenous sorting around the cutoff, and Crespo (2020) discusses sorting caused by administrative reasons rather than strategic individual behavior.

McCrary's density test has become one of the leading validation and falsification methods for RD designs with a continuously distributed score. The test was originally implemented using the

local polynomial density estimator of Cheng et al. (1997), a two-step nonparametric procedure requiring two distinct bandwidth choices. More recently, Cattaneo et al. (2020b) introduced a local polynomial density estimator that requires only one bandwidth choice and is more robust near boundary points, which leads to 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 et al. (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 it 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 et al. (2013), Jales et al. (2017), and Ma et al. (2020) 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 & Canay (2021) propose a binomial-like test using an asymptotic argument to justify local randomization near the cutoff; both methods employ 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, because 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 a test of the hypothesis that the conditional expectation functions of the potential outcomes are continuous at the artificial cutoffs 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 (see Section 3.2) or local randomization inference methods (see 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. Ganong & Jäger (2018) present 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 reimplementing estimation and inference for the RD treatment effect with different subsets of observations, as determined either by excluding the observations closest to the cutoff or by varying the bandwidth used for estimation and inference. The so-called donut hole approach (Barreca et al. 2011) investigates the sensitivity of the empirical conclusions by removing the few observations closest to the cutoff and then implementing 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 the local extrapolation

properties of the specific method used for estimation and inference near the cutoff. Readers are referred to Dowd (2021) for the latter approach in the context of local polynomial RBC inference methods (see Section 3.2).

The other validation method studies whether the empirical conclusions are sensitive to the choice of bandwidth or local randomization neighborhood. The idea is simply to reestimate 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 if important covariates become imbalanced; in this case, the approach will be uninformative.

For implementation of the donut hole and bandwidth sensitivity approaches, researchers can use either local polynomial RBC inference methods (see Section 3.2) or local randomization inference methods (see Section 3.3).

5. CONCLUSION

Since the seminal work of Thistlethwaite & Campbell (1960), the literature on identification, estimation, inference, and validation of 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 and local randomization neighborhood has become data driven and is now based on principled criteria, and many validation and falsification methods have been developed 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 leading to more credible and replicable empirical conclusions.

Looking ahead, there are still several areas that can benefit from further methodological development. One prominent area is RD extrapolation. Having more robust methods to extrapolate RD treatment effects would bolster the external validity of the conclusions drawn from RD designs, which are otherwise limited by the local nature of the parameters. Such extrapolation methods would also help in the design of optimal policies, a topic that has not yet been explored in the RD literature. Another fruitful area of study is experimental design and data collection in RD settings. It is becoming more common for policy makers and industry researchers to plan RD designs ex-ante as well as to design ex-post data collection exploiting prior RD interventions, often in rich data environments. A set of methodological tools targeted to experimental design in RD settings would improve practice and could also enable better extrapolation of RD treatment effects via principled sampling design and data collection. Finally, another avenue for future research is related to incorporating modern high-dimensional and machine learning methods in RD settings. Such methodological enhancements can aid in principled discovery of heterogeneous treatment effects as well as in developing more robust estimation and inference methods in multidimensional RD designs.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

A preliminary version of this article was titled "Regression discontinuity designs: a review." We are indebted to our collaborators Sebastian Calonico, Max Farrell, Yingjie Feng, Brigham Frandsen, Nicolas Idrobo, Michael Jansson, Luke Keele, Xinwei Ma, Kenichi Nagasawa, Jasjeet Sekhon, and Gonzalo Vazquez-Bare for their continued support and intellectual contribution to our research program on regression discontinuity designs. We also thank Jason Lindo, Filippo Palomba, Zhuan Pei, and a reviewer for valuable comments. Both M.D.C. and R.T. gratefully acknowledge financial support from the National Science Foundation through grants SES-1357561 and SES-2019432, and M.D.C. also gratefully acknowledges financial support from the National Institute of Health (R01 GM072611-16). Parts of this article were presented at the Summer Institute 2021 Methods Lectures of the National Bureau of Economic Research (a video recording is available at **https://www.nber.org/conferences/si-2021-methods-lecture-causal-inference-using-synthetic-controls-and-regression-discontinuity**). General purpose software, replication files, additional articles on regression discontinuity methodology, and other resources are available at **https://rdpackages.github.io/**.

LITERATURE CITED

Abadie A, Cattaneo MD. 2018. Econometric methods for program evaluation. *Annu. Rev. Econ.* 10:465–503

Angrist JD, Rokkanen M. 2015. Wanna get away? Regression discontinuity estimation of exam school effects away from the cutoff. *J. Am. Stat. Assoc.* 110(512):1331–44

Arai Y, Hsu Y, Kitagawa T, Mourifié I, Wan Y. 2022. Testing identifying assumptions in fuzzy regression discontinuity designs. *Quant. Econ.* 13(1):1–28

Arai Y, Ichimura H. 2016. Optimal bandwidth selection for the fuzzy regression discontinuity estimator. *Econ. Lett.* 141(1):103–6

Arai Y, Ichimura H. 2018. Simultaneous selection of optimal bandwidths for the sharp regression discontinuity estimator. *Quant. Econ.* 9(1):441–82

Arai Y, Otsu T, Seo MH. 2021. Regression discontinuity design with potentially many covariates. arXiv:2109.08351 [econ.EM]

Armstrong T, Kolesar M. 2020. Simple and honest confidence intervals in nonparametric regression. *Quant. Econ.* 11(1):1–39

Babii A, Kumar R. 2021. Isotonic regression discontinuity designs. *J. Econom.* In press. **https://doi.org/10.1016/j.jeconom.2021.01.008**

Backus GE. 1989. Confidence set inference with a prior quadratic bound. *Geophys. J. Int.* 97(1):119–50

Bajari P, Hong H, Park M, Town R. 2011. *Regression discontinuity designs with an endogenous forcing variable and an application to contracting in health care*. NBER Work. Pap. 17643

Barreca AI, Guldi M, Lindo JM, Waddell GR. 2011. Saving babies? Revisiting the effect of very low birth weight classification. *Q. J. Econ.* 126(4):2117–23

Barreca AI, Lindo JM, Waddell GR. 2016. Heaping-induced bias in regression-discontinuity designs. *Econ. Inq.* 54(1):268–93

Bartalotti O, Brummet Q. 2017. Regression discontinuity designs with clustered data. See Cattaneo & Escanciano 2017, pp. 383–420

Bartalotti O, Brummet Q, Dieterle S. 2021. A correction for regression discontinuity designs with group-specific mismeasurement of the running variable. *J. Bus. Econ. Stat.* 39(3):833–48

Bartalotti O, Calhoun G, He Y. 2017. Bootstrap confidence intervals for sharp regression discontinuity designs. See Cattaneo & Escanciano 2017, pp. 421–53

Bertanha M. 2020. Regression discontinuity design with many thresholds. *J. Econom.* 218(1):216–41

Bertanha M, Imbens GW. 2020. External validity in fuzzy regression discontinuity designs. *J. Bus. Econ. Stat.* 38(3):593–612

Black SE. 1999. Do better schools matter? Parental valuation of elementary education. *Q. J. Econ.* 114(2):577–99

Blomquist S, Newey WK, Kumar A, Liang CY. 2021. On bunching and identification of the taxable income elasticity. *J. Political Econ.* 129(8):2320–43
Branson Z, Rischard M, Bornn L, Miratrix LW. 2019. A nonparametric Bayesian methodology for regression discontinuity designs. *J. Stat. Plan. Inference* 202:14–30
Bugni FA, Canay IA. 2021. Testing continuity of a density via *g*-order statistics in the regression discontinuity design. *J. Econom.* 221(1):138–59
Bulus M. 2022. Minimum detectable effect size computations for cluster-level regression discontinuity studies: specifications beyond the linear functional form. *J. Res. Educ. Eff.* 15(1):151–77
Busse M, Silva-Risso J, Zettelmeyer F. 2006. $1,000 cash back: the pass-through of auto manufacturer promotions. *Am. Econ. Rev.* 96(4):1253–70
Caetano C, Caetano G, Escanciano JC. 2021. *Regression discontinuity design with multivalued treatments*. Work. Pap., Univ. Carlos III, Madrid, Spain
Calonico S, Cattaneo MD, Farrell MH. 2018. On the effect of bias estimation on coverage accuracy in nonparametric inference. *J. Am. Stat. Assoc.* 113(522):767–79
Calonico S, Cattaneo MD, Farrell MH. 2020. Optimal bandwidth choice for robust bias corrected inference in regression discontinuity designs. *Econom. J.* 23(2):192–210
Calonico S, Cattaneo MD, Farrell MH. 2022. Coverage error optimal confidence intervals for local polynomial regression. *Bernoulli*. In press
Calonico S, Cattaneo MD, Farrell MH, Titiunik R. 2019. Regression discontinuity designs using covariates. *Rev. Econ. Stat.* 101(3):442–51
Calonico S, Cattaneo MD, Titiunik R. 2014. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica* 82(6):2295–326
Calonico S, Cattaneo MD, Titiunik R. 2015. Optimal data-driven regression discontinuity plots. *J. Am. Stat. Assoc.* 110(512):1753–69
Canay IA, Kamat V. 2018. Approximate permutation tests and induced order statistics in the regression discontinuity design. *Rev. Econ. Stud.* 85(3):1577–608
Card D, Lee DS, Pei Z, Weber A. 2015. Inference on causal effects in a generalized regression kink design. *Econometrica* 83(6):2453–83
Card D, Lee DS, Pei Z, Weber A. 2017. Regression kink design: Theory and practice. See Cattaneo & Escanciano 2017, pp. 341–82
Card D, Mas A, Rothstein J. 2008. Tipping and the dynamics of segregation. *Q. J. Econ.* 123(1):177–218
Cattaneo MD, Escanciano JC, eds. 2017. *Regression Discontinuity Designs: Theory and Applications*. Bingley, UK: Emerald
Cattaneo MD, Frandsen B, Titiunik R. 2015. Randomization inference in the regression discontinuity design: an application to party advantages in the U.S. Senate. *J. Causal Inference* 3(1):1–24
Cattaneo MD, Idrobo N, Titiunik R. 2020a. *A Practical Introduction to Regression Discontinuity Designs: Foundations*. Cambridge, UK: Cambridge Univ. Press
Cattaneo MD, Idrobo N, Titiunik R. 2022a. *A Practical Introduction to Regression Discontinuity Designs: Extensions*. Cambridge, UK: Cambridge Univ. Press. In press
Cattaneo MD, Jansson M, Ma X. 2020b. Simple local polynomial density estimators. *J. Am. Stat. Assoc.* 115(531):1449–55
Cattaneo MD, Keele L, Titiunik R. 2022b. Covariate adjustment in regression discontinuity designs. In *Handbook of Matching and Weighting in Causal Inference*, ed. JR Zubizarreta, EA Stuart, DS Small, PR Rosenbaum. London: Chapman & Hall. In press
Cattaneo MD, Keele L, Titiunik R, Vazquez-Bare G. 2016. Interpreting regression discontinuity designs with multiple cutoffs. *J. Politics* 78(4):1229–48
Cattaneo MD, Keele L, Titiunik R, Vazquez-Bare G. 2021. Extrapolating treatment effects in multi-cutoff regression discontinuity designs. *J. Am. Stat. Assoc.* 116(536):1941–52
Cattaneo MD, Titiunik R, Vazquez-Bare G. 2017. Comparing inference approaches for RD designs: a reexamination of the effect of head start on child mortality. *J. Policy Anal. Manag.* 36(3):643–81
Cattaneo MD, Titiunik R, Vazquez-Bare G. 2019. Power calculations for regression discontinuity designs. *Stata J.* 19(1):210–45

Cattaneo MD, Titiunik R, Vazquez-Bare G. 2020c. The regression discontinuity design. In *Handbook of Research Methods in Political Science and International Relations*, ed. L Curini, RJ Franzese, pp. 835–57. London: Sage

Cattaneo MD, Vazquez-Bare G. 2016. The choice of neighborhood in regression discontinuity designs. *Obs. Stud.* 2:134–46

Cellini SR, Ferreira F, Rothstein J. 2010. The value of school facility investments: evidence from a dynamic regression discontinuity design. *Q. J. Econ.* 125(1):215–61

Cerulli G, Dong Y, Lewbel A, Poulsen A. 2017. Testing stability of regression discontinuity models. See Cattaneo & Escanciano 2017, pp. 317–39

Chaplin DD, Cook TD, Zurovac J, Coopersmith JS, Finucane MM, et al. 2018. The internal and external validity of the regression discontinuity design: a meta-analysis of 15 within-study comparisons. *J. Policy Anal. Manag.* 37(2):403–29

Chen H, Chiang HD, Sasaki Y. 2020. Quantile treatment effects in regression kink designs. *Econom. Theory* 36(6):1167–91

Cheng MY, Fan J, Marron JS. 1997. On automatic boundary corrections. *Ann. Stat.* 25(4):1691–708

Chiang HD, Hsu YC, Sasaki Y. 2019. Robust uniform inference for quantile treatment effects in regression discontinuity designs. *J. Econom.* 211(2):589–618

Chiang HD, Sasaki Y. 2019. Causal inference by quantile regression kink designs. *J. Econom.* 210(2):405–33

Chib S, Jacobi L. 2016. Bayesian fuzzy regression discontinuity analysis and returns to compulsory schooling. *J. Appl. Econom.* 31(6):1026–47

Choi JY, Lee MJ. 2018. Relaxing conditions for local average treatment effect in fuzzy regression discontinuity. *Econ. Lett.* 173:47–50

Cook TD. 2008. "Waiting for life to arrive": a history of the regression-discontinuity design in psychology, statistics and economics. *J. Econom.* 142(2):636–54

Cook TD, Wong VC. 2008. Empirical tests of the validity of the regression discontinuity design. *Ann. Econ. Stat.* 91–92:127–50

Crespo C. 2020. Beyond manipulation: administrative sorting in regression discontinuity designs. *J. Causal Inference* 8(1):164–81

Davezies L, Le Barbanchon T. 2017. Regression discontinuity design with continuous measurement error in the running variable. *J. Econom.* 200(2):260–81

De la Cuesta B, Imai K. 2016. Misunderstandings about the regression discontinuity design in the study of close elections. *Annu. Rev. Political Sci.* 19:375–96

De Magalhaes L, Hangartner D, Hirvonen S, Meriläinen J, Ruiz NA, Tukiainen J. 2020. *How much should we trust regression discontinuity design estimates? Evidence from experimental benchmarks of the incumbency advantage*. Work. Pap., Univ. Bristol, Bristol, UK

Dell M. 2010. The persistent effects of Peru's mining *mita*. *Econometrica* 78(6):1863–903

Dong Y. 2015. Regression discontinuity applications with rounding errors in the running variable. *J. Appl. Econom.* 30(3):422–46

Dong Y. 2018a. Alternative assumptions to identify late in fuzzy regression discontinuity designs. *Oxf. Bull. Econ. Stat.* 80(5):1020–27

Dong Y. 2018b. *Jump or kink? Regression probability jump and kink design for treatment effect evaluation*. Work. Pap., Univ. Calif., Irvine

Dong Y. 2019. Regression discontinuity designs with sample selection. *J. Bus. Econ. Stat.* 37(1):171–86

Dong Y, Lee YY, Gou M. 2021. Regression discontinuity designs with a continuous treatment. *J. Am. Stat. Assoc.* **https://doi.org/10.1080/01621459.2021.1923509**

Dong Y, Lewbel A. 2015. Identifying the effect of changing the policy threshold in regression discontinuity models. *Rev. Econ. Stat.* 97(5):1081–92

Donoho DL. 1994. Statistical estimation and optimal recovery. *Ann. Stat.* 22(1):238–70

Dowd C. 2021. *Donuts and distant CATEs: Derivative bounds for RD extrapolation*. Work. Pap., Booth Sch. Bus., Univ. Chicago, Chicago, IL

Eckles D, Ignatiadis N, Wager S, Wu H. 2021. Noise-induced randomization in regression discontinuity designs. arXiv:2004.09458 [stat.ME]

Fan J, Gijbels I. 1996. *Local Polynomial Modelling and Its Applications*. London: Chapman & Hall
Feir D, Lemieux T, Marmer V. 2016. Weak identification in fuzzy regression discontinuity designs. *J. Bus. Econ. Stat.* 34(2):185–96
Frandsen B. 2017. Party bias in union representation elections: testing for manipulation in the regression discontinuity design when the running variable is discrete. See Cattaneo & Escanciano 2017, pp. 281–315
Frandsen B, Frolich M, Melly B. 2012. Quantile treatments effects in the regression discontinuity design. *J. Econom.* 168(2):382–95
Frolich M, Huber M. 2019. Including covariates in the regression discontinuity design. *J. Bus. Econ. Stat.* 37(4):736–48
Galiani S, McEwan PJ, Quistorff B. 2017. External and internal validity of a geographic quasi-experiment embedded in a cluster-randomized experiment. See Cattaneo & Escanciano 2017, pp. 195–236
Ganong P, Jäger S. 2018. A permutation test for the regression kink design. *J. Am. Stat. Assoc.* 113(522):494–504
Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. 2013. *Bayesian Data Analysis*. Boca Raton, FL: CRC Press
Gelman A, Imbens GW. 2019. Why high-order polynomials should not be used in regression discontinuity designs. *J. Bus. Econ. Stat.* 37(3):447–56
Geneletti S, O'Keeffe AG, Sharples LD, Richardson S, Baio G. 2015. Bayesian regression discontinuity designs: incorporating clinical knowledge in the causal analysis of primary care data. *Stat. Med.* 34(15):2334–52
Gerard F, Rokkanen M, Rothe C. 2020. Bounds on treatment effects in regression discontinuity designs with a manipulated running variable. *Quant. Econ.* 11(3):839–70
Grembi V, Nannicini T, Troiano U. 2016. Do fiscal rules matter? A difference-in-discontinuities design. *Am. Econ. J. Appl. Econ.* 8(3):1–30
Groeneboom P, Jongbloed G. 2014. *Nonparametric Estimation Under Shape Constraints*. Cambridge, UK: Cambridge Univ. Press
Hahn J, Todd P, van der Klaauw W. 2001. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica* 69(1):201–9
Hansen BE. 2017. Regression kink with an unknown threshold. *J. Bus. Econ. Stat.* 35(2):228–40
Hartman E. 2021. Equivalence testing for regression discontinuity designs. *Political Anal.* 29(4):505–21
Hausman C, Rapson DS. 2018. Regression discontinuity in time: considerations for empirical applications. *Annu. Rev. Resour. Econ.* 10:533–52
He Y. 2018. *Three essays on regression discontinuity design and partial identification*. PhD Thesis, Iowa State Univ., Ames
He Y, Bartalotti O. 2020. Wild bootstrap for fuzzy regression discontinuity designs: obtaining robust bias-corrected confidence intervals. *Econom. J.* 23(2):211–31
Hsu YC, Shen S. 2019. Testing treatment effect heterogeneity in regression discontinuity designs. *J. Econom.* 208(2):468–86
Hsu YC, Shen S. 2021a. *Dynamic regression discontinuity under treatment effect heterogeneity*. Work. Pap., Univ. Calif., Davis
Hsu YC, Shen S. 2021b. Testing monotonicity of conditional treatment effects under regression discontinuity designs. *J. Appl. Econom.* 36(3):346–66
Huang X, Zhan Z. 2021. Local composite quantile regression for regression discontinuity. *J. Bus. Econ. Stat.* **https://doi.org/10.1080/07350015.2021.1990072**
Hyytinen A, Meriläinen J, Saarimaa T, Toivanen O, Tukiainen J. 2018. When does regression discontinuity design work? Evidence from random election outcomes. *Quant. Econ.* 9(2):1019–51
Imbens GW, Lemieux T. 2008. Regression discontinuity designs: a guide to practice. *J. Econom.* 142(2):615–35
Imbens GW, Kalyanaraman K. 2012. Optimal bandwidth choice for the regression discontinuity estimator. *Rev. Econ. Stud.* 79(3):933–59
Imbens GW, Rubin DB. 2015. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge, UK: Cambridge Univ. Press

Imbens GW, Wager S. 2019. Optimized regression discontinuity designs. *Rev. Econ. Stat.* 101(2):264–78
Jales H, Ma J, Yu Z. 2017. Optimal bandwidth selection for local linear estimation of discontinuity in density. *Econ. Lett.* 153:23–27
Jales H, Yu Z. 2017. Identification and estimation using a density discontinuity approach. See Cattaneo & Escanciano 2017, pp. 29–72
Kamat V. 2018. On nonparametric inference in the regression discontinuity design. *Econom. Theory* 34(3):694–703
Karabatsos G, Walker SG. 2015. A Bayesian nonparametric causal model for regression discontinuity designs. In *Nonparametric Bayesian Inference in Biostatistics*, ed. R Mitra, P Müller, pp. 403–21. Cham, Switz.: Springer
Keele LJ, Lorch S, Passarella M, Small D, Titiunik R. 2017. An overview of geographically discontinuous treatment assignments with an application to children's health insurance. See Cattaneo & Escanciano 2017, pp. 147–94
Keele LJ, Titiunik R. 2015. Geographic boundaries as regression discontinuities. *Political Anal.* 23(1):127–55
Keele LJ, Titiunik R, Zubizarreta J. 2015. Enhancing a geographic regression discontinuity design through matching to estimate the effect of ballot initiatives on voter turnout. *J. R. Stat. Soc. A* 178(1):223–39
Kleven HJ. 2016. Bunching. *Annu. Rev. Econ.* 8:435–64
Kolesar M, Rothe C. 2018. Inference in regression discontinuity designs with a discrete running variable. *Am. Econ. Rev.* 108(8):2277–304
Korting C, Lieberman C, Matsudaira J, Pei Z, Shen Y. 2021. Visual inference and graphical representation in regression discontinuity designs. arXiv:2112.03096 [econ.EM]
Lee DS. 2008. Randomized experiments from nonrandom selection in U.S. House elections. *J. Econom.* 142(2):675–97
Lee DS, Card D. 2008. Regression discontinuity inference with specification error. *J. Econom.* 142(2):655–74
Lee DS, Lemieux T. 2010. Regression discontinuity designs in economics. *J. Econ. Lit.* 48(2):281–355
Lee MJ. 2017. Regression discontinuity with errors in the running variable: effect on truthful margin. *J. Econom. Methods* 6(1):281–355
Li F, Mattei A, Mealli F. 2015. Evaluating the causal effect of university grants on student dropout: evidence from a regression discontinuity design using principal stratification. *Ann. Appl. Stat.* 9(4):1906–31
Ludwig J, Miller DL. 2007. Does head start improve children's life chances? Evidence from a regression discontinuity design. *Q. J. Econ.* 122(1):159–208
Lv X, Sun XR, Lu Y, Li R. 2019. Nonparametric identification and estimation of dynamic treatment effects for survival data in a regression discontinuity design. *Econ. Lett.* 184:108665
Ma J, Jales H, Yu Z. 2020. Minimum contrast empirical likelihood inference of discontinuity in density. *J. Bus. Econ. Stat.* 38(4):934–50
Ma J, Yu Z. 2020. Coverage optimal empirical likelihood inference for regression discontinuity design. arXiv:2008.09263 [econ.EM]
Matzkin RL. 2013. Nonparametric identification in structural economic models. *Annu. Rev. Econ.* 5:457–86
McCrary J. 2008. Manipulation of the running variable in the regression discontinuity design: a density test. *J. Econom.* 142(2):698–714
Mealli F, Rampichini C. 2012. Evaluating the effects of university grants by using regression discontinuity designs. *J. R. Stat. Soc. A* 175(3):775–98
Mukherjee D, Banerjee M, Ritov Y. 2021. Estimation of a score-explained non-randomized treatment effect in fixed and high dimensions. arXiv:2102.11229 [stat.ME]
Otsu T, Xu KL, Matsushita Y. 2013. Estimation and inference of discontinuity in density. *J. Bus. Econ. Stat.* 31(4):507–24
Otsu T, Xu KL, Matsushita Y. 2015. Empirical likelihood for regression discontinuity design. *J. Econom.* 186(1):94–112
Owen AB. 2001. *Empirical Likelihood*. London: Chapman & Hall
Owen AB, Varian H. 2020. Optimizing the tie-breaker regression discontinuity design. *Electron. J. Stat.* 14(2):4004–27
Papay JP, Willett JB, Murnane RJ. 2011. Extending the regression-discontinuity approach to multiple assignment variables. *J. Econom.* 161(2):203–7

Pei Z, Lee DS, Card D, Weber A. 2021. Local polynomial order in regression discontinuity designs. *J. Bus. Econ. Stat.* 31(4):507–24
Pei Z, Shen Y. 2017. The devil is in the tails: regression discontinuity design with measurement error in the assignment variable. See Cattaneo & Escanciano 2017, pp. 455–502
Peng S, Ning Y. 2021. Regression discontinuity design under self-selection. *Proc. Mach. Learn. Res.* 130:118–26
Porter J. 2003. *Estimation in the regression discontinuity model.* Work. Pap., Univ. Wis., Madison
Porter J, Yu P. 2015. Regression discontinuity designs with unknown discontinuity points: testing and estimation. *J. Econom.* 189(1):132–47
Qu Z, Yoon J. 2019. Uniform inference on quantile effects under sharp regression discontinuity designs. *J. Bus. Econ. Stat.* 37(4):625–47
Reardon SF, Robinson JP. 2012. Regression discontinuity designs with multiple rating-score variables. *J. Res. Educ. Eff.* 5(1):83–104
Rokkanen M. 2015. *Exam schools, ability, and the effects of affirmative action: latent factor extrapolation in the regression discontinuity design*. Discuss. Pap. 1415-03, Dep. Econ., Columbia Univ., New York
Rosenbaum PR. 2010. *Design of Observational Studies*. New York: Springer
Schochet PZ. 2009. Statistical power for regression discontinuity designs in education evaluations. *J. Educ. Behav. Stat.* 34(2):238–66
Sekhon JS, Titiunik R. 2016. Understanding regression discontinuity designs as observational studies. *Obs. Stud.* 2:174–82
Sekhon JS, Titiunik R. 2017. On interpreting the regression discontinuity design as a local experiment. See Cattaneo & Escanciano 2017, pp. 1–28
Shen S, Zhang X. 2016. Distributional tests for regression discontinuity: theory and empirical examples. *Rev. Econ. Stat.* 98(4):685–700
Sun Y. 2005. *Adaptive estimation of the regression discontinuity model.* Work. Pap., Univ. Calif., San Diego
Thistlethwaite DL, Campbell DT. 1960. Regression-discontinuity analysis: an alternative to the ex post facto experiment. *J. Educ. Psychol.* 51(6):309–17
Titiunik R. 2021. Natural experiments. In *Advances in Experimental Political Science*, ed. JN Druckman, DP Green, pp. 103–29. Cambridge, UK: Cambridge Univ. Press
Tuvaandorj P. 2020. Regression discontinuity designs, white noise models, and minimax. *J. Econom.* 218(2):587–608
Urquiola M, Verhoogen E. 2009. Class-size caps, sorting, and the regression-discontinuity design. *Am. Econ. Rev.* 99(1):179–215
van der Klaauw W. 2008. Regression–discontinuity analysis: a survey of recent developments in economics. *Labour* 22(2):219–45
Wing C, Cook TD. 2013. Strengthening the regression discontinuity design using additional design elements: a within-study comparison. *J. Policy Anal. Manag.* 32(4):853–77
Wong VC, Steiner PM, Cook TD. 2013. Analyzing regression-discontinuity designs with multiple assignment variables: a comparative study of four estimation methods. *J. Educ. Behav. Stat.* 38(2):107–41
Xu KL. 2017. Regression discontinuity with categorical outcomes. *J. Econom.* 201(1):1–18
Xu KL. 2018. A semi-nonparametric estimator of regression discontinuity design with discrete duration outcomes. *J. Econom.* 206(1):258–78

Annual Review of Economics

Volume 14, 2022

Contents

The Great Divide: Education, Despair, and Death
Anne Case and Angus Deaton 1

The Impact of Health Information and Communication Technology on Clinical Quality, Productivity, and Workers
Ari Bronsoler, Joseph Doyle, and John Van Reenen 23

Household Financial Transaction Data
Scott R. Baker and Lorenz Kueng 47

Media and Social Capital
Filipe Campante, Ruben Durante, and Andrea Tesei 69

The Elusive Explanation for the Declining Labor Share
Gene M. Grossman and Ezra Oberfield 93

The Past and Future of Economic Growth: A Semi-Endogenous Perspective
Charles I. Jones 125

Risks and Global Supply Chains: What We Know and What We Need to Know
Richard Baldwin and Rebecca Freeman 153

Managing Retirement Incomes
James Banks and Rowena Crawford 181

The Economic Impacts of the US–China Trade War
Pablo D. Fajgelbaum and Amit K. Khandelwal 205

How Economic Development Influences the Environment
Seema Jayachandran 229

The Economics of the COVID-19 Pandemic in Poor Countries
Edward Miguel and Ahmed Mushfiq Mobarak 253

The Affordable Care Act After a Decade: Industrial Organization of the Insurance Exchanges
Benjamin Handel and Jonathan Kolstad 287

Helicopter Money: What Is It and What Does It Do?
Ricardo Reis and Silvana Tenreyro 313

Relational Contracts and Development
Rocco Macchiavello 337

Trade Policy Uncertainty
Kyle Handley and Nuno Limão 363

Bureaucracy and Development
Timothy Besley, Robin Burgess, Adnan Khan, and Guo Xu 397

Misperceptions About Others
Leonardo Bursztyn and David Y. Yang 425

The Affordable Care Act After a Decade: Its Impact on the Labor Market and the Macro Economy
Hanming Fang and Dirk Krueger 453

Expecting Brexit
Swati Dhingra and Thomas Sampson 495

Salience
Pedro Bordalo, Nicola Gennaioli, and Andrei Shleifer 521

Enough Potential Repudiation: Economic and Legal Aspects of Sovereign Debt in the Pandemic Era
Anna Gelpern and Ugo Panizza 545

The Great Gatsby Curve
Steven N. Durlauf, Andros Kourtellos, and Chih Ming Tan 571

Inequality and the COVID-19 Crisis in the United Kingdom
Richard Blundell, Monica Costa Dias, Jonathan Cribb, Robert Joyce, Tom Waters, Thomas Wernham, and Xiaowei Xu 607

The Aftermath of Debt Surges
M. Ayhan Kose, Franziska L. Ohnsorge, Carmen M. Reinhart, and Kenneth S. Rogoff 637

Networks and Economic Fragility
Matthew Elliott and Benjamin Golub 665

Central Bank Digital Currencies: Motives, Economic Implications, and the Research Frontier
Raphael Auer, Jon Frost, Leonardo Gambacorta, Cyril Monnet, Tara Rice, and Hyun Song Shin 697

The Use of Scanner Data for Economics Research
Pierre Dubois, Rachel Griffith, and Martin O'Connell 723

The Marginal Propensity to Consume in Heterogeneous Agent Models
Greg Kaplan and Giovanni L. Violante 747

Experimental Economics: Past and Future
Guillaume R. Fréchette, Kim Sarnoff, and Leeat Yariv 777

Spatial Sorting and Inequality
Rebecca Diamond and Cecile Gaubert 795

Regression Discontinuity Designs
Matias D. Cattaneo and Rocío Titiunik 821

Early Childhood Development, Human Capital, and Poverty
Orazio Attanasio, Sarah Cattan, and Costas Meghir 853

The Econometric Model for Causal Policy Analysis
James J. Heckman and Rodrigo Pinto 893

Indexes

Cumulative Index of Contributing Authors, Volumes 10–14 925

Errata

An online log of corrections to *Annual Review of Economics* articles may be found at http://www.annualreviews.org/errata/economics