UNIVERSITÄT BERN

Faculty of Business, Economics and Social Sciences

Department of Economics

Estimation, Inference, and Interpretation in the Regression Discontinuity Design

Blaise Melly
Rafael Lalive

20-16
November 2020

DISCUSSION PAPERS

Schanzeneckstrasse 1
CH-3012 Bern, Switzerland
http://www.vwi.unibe.ch

Estimation, Inference, and Interpretation in the Regression Discontinuity Design

Blaise Melly*and Rafael Lalive†

November 2, 2020

Abstract

The Regression Discontinuity Design (RDD) has proven to be a compelling and transparent research design to estimate treatment effects. We provide a review of the main assumptions and key challenges faced when adopting an RDD. We cover the most recent developments and advanced methods, and provide the key intuitions that underlie the statistical arguments. Among others, we summarize new insights that we consider to be highly relevant about the choice of bandwidth, optimal inference, discrete running variables, distributional effects, estimation in the presence of covariates, and the regression kink design. We also show how structural parameters can be estimated by combining an RDD identification strategy with theoretical models. We illustrate the procedures by applying them to data and we provide codes to replicate the results.

1 Introduction

The Regression Discontinuity Design (RDD) was first introduced by Thistlethwaite and Campbell (1960) as a quasi-experimental design for evaluating social programs and interventions. The predictions made by Campbell and Stanley (1963) that the RDD is "very limited in range of possible applications" and that "those limited settings are mainly educational" have been proven wrong by the recent literature. The RDD has received tremendous attention in many fields, e.g. labor economics, political economy, health economics, criminology, environmental economics, and development economics. Among the non-experimental identification strategies, it is often viewed as one of the most credible one.

There are two main requirements for a valid RDD. First, the treatment must be a discontinuous function of a continuous, observed variable, which is called running, assignment or forcing variable. Often, administrative rules determining treatment assignments jump at some thresholds. Thistlethwaite and Campbell (1960) estimate the effect of receiving the National Merit Scholarship, which is awarded to students who score above a threshold at a test. Angrist and Lavy (1999) exploit the limitation to 40 students in a classroom. Lalive (2008) uses the jump in maximum duration of unemployment benefits at age 50. The main reason for the popularity of the RDD is the widespread existence of discrete rules that makes discontinuity-based identification credible.

*Corresponding Author. Email: mellyblaise@gmail.com, Address: Department of Economics, Schanzeneckstrasse 1, Postfach, 3001 Bern, Switzerland. Replication programs and data on https://sites.google.com/site/blaisemelly/home/research-papers/rdd-data-and-codes.

†Email: rafael.lalive@unil.ch. Address: University of Lausanne, Faculty of Business and Economics, Batiment Internef, CH-1022 Chavannes-pres-Renens.

While the first condition is the presence of a discontinuity in the treatment assignment at the threshold, the second condition is the absence of a discontinuity in the potential outcomes at the threshold. Like an exogeneity assumption required for a matching strategy or an exclusion restriction for an instrumental variable strategy, this second assumption is fundamentally not testable. However, we can assess its plausibility, for instance by checking that (i) there is no jump in the density of the running variable at the threshold, which would be a sign that individuals could precisely manipulate its value, (ii) the pre-determined covariates do not jump at the threshold and (iii) the outcome does not jump at other values (placebo tests).

There exist already many review articles about the RDD, see, among others, Imbens and Lemieux (2008), Lee and Lemieux (2010) and Cattaneo et al. (2020). Angrist and Pischke (2009) and Frölich and Sperlich (2019) provide excellent textbook treatments of this topic. Our review differs from previous work in the following aspects: (i) While we do not go deeply into statistical theory, which can be found in the referenced papers, we motivate and provide the intuition behind the most advanced techniques involved. (ii) We cover topics that have been recently studied and that we consider to be highly relevant such as optimal inference, discrete running variables, distributional effects, estimation in the presence of covariates, and the regression kink design. (iii) We illustrate the results by applying them to data and we provide codes in Stata and R to replicate the results. (iv) While the RDD often offers a credible identification strategy, its main weakness lies in the inherently local nature of the identified parameters. We discuss how to structural economic models to interpret and extrapolate from the effects identified with a RDD strategy.

The outline of this chapter is as follows. Section 2 presents the fundamental results about the RDD: identification of the average treatment effect in the sharp and fuzzy RDD, estimation strategies, implementation issues, statistical tests, and falsification strategies. Section 3 considers interesting extensions: quantile treatment effects, discrete running variables, continuous treatments, estimated thresholds, weak identification, role of covariates, extrapolation, multiple thresholds, multiple running variables, and the regression kink design. Section 4 illustrates the most important results with the job search application in Lalive (2008). Section 5 discusses how to use reduced form RDD estimates to structurally pin down underlying behavioral parameters.

2 Fundamentals

In this section we consider the standard RDD setting: the running variable X is continuous, the treatment variable D is binary, and we are interested in estimating the average treatment effect of D on Y. We start in Section 2.1 with the sharp design before generalizing the results to the fuzzy design in Section 2.2. In Sections 2.3 and 2.4 we discuss estimation, hypothesis tests and confidence intervals. In Section 2.5 we discuss various possibilities to falsify the validity of the design. In Section 3 we will consider a variety of divergences from this standard RDD setting.

2.1 Sharp design

We are interested in the effect of a binary treatment D on an outcome Y. We use the potential outcome notation introduced by Neyman (1923) and popularized by Rubin (1974) to define causal effects. Let $Y(0)$ denote the potential control outcome and $Y(1)$ the potential treated outcome. We are interested in estimating the average treatment effect (ATE), $E[Y(1)-Y(0)]$, but we cannot assume that the treatment has been randomized and we do not have a valid external instrumental variable. On the other hand, we know that the treatment D is a deterministic function of the running variable X: There is a cutoff c such that all units with X below the cutoff are not treated and all units with X above it are treated: $D = 1(X > c)$.

In Lalive (2008), which we use as an illustrative example in Section 4, the outcome Y is the duration of the unemployment spell. The treatment D is equal to 1 if the maximum duration of unemployment benefits is 30 weeks and D is equal to 0 if this maximum duration is 209 weeks.[1] Only individuals older than 50 and living in certain regions are treated. We can exploit either the age discontinuity, in which case X is the age in years and $c = 50$, or the geographical discontinuity, in which case X is the distance to the border and $c = 0$.[2]

The discontinuity in itself is not sufficient to identify the ATE: there is no value of X for which we observe the control and the potential outcomes. If $X < c$ we observe $Y = Y(0)$ and if $X > c$ we observe $Y = Y(1)$. Since we allow X to be arbitrarily correlated with the potential outcomes, we cannot identify the average treatment effect for any population without further restriction. The crucial assumption for the validity of the RDD strategy is the continuity of $E[Y(0)|X = x]$ and $E[Y(1)|X = x]$ at $x = c$. The discontinuity of the treatment status together with the continuity of the potential outcomes in the running variable imply that we can identify nonparametrically the average treatment effect exactly at the threshold. If we consider only observations arbitrarily close to the threshold, the observed outcome displayed by those just below the threshold is $E[Y(0)|X = c]$ while for those just above the threshold it is $E[Y(1)|X = c]$. In other words, we identify the average treatment effect but only for the population with $X = c$, which is an infinitesimal population because X is continuous.

Figure 1: Expected values of the potential and observed outcomes

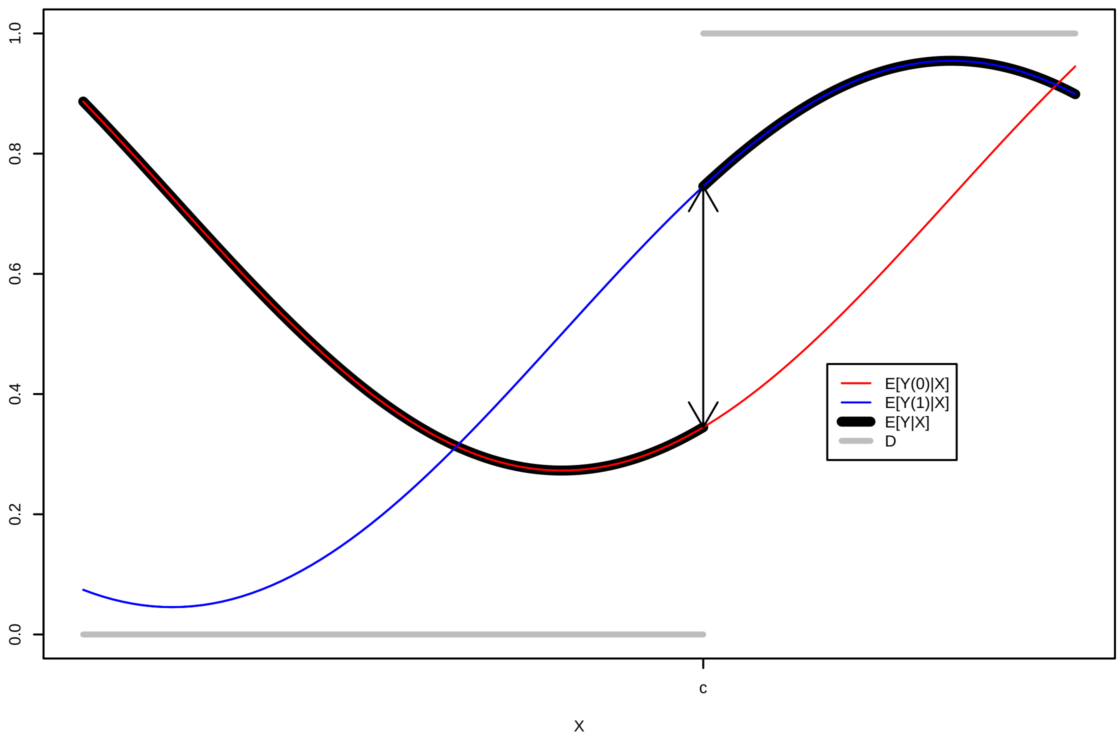


Figure 2.1 provides an artificial illustration. The running variable X is shown on the x-axis. The treatment D is 0 below the threshold c and 1 above it. In red we can observe the average control outcome and in blue the average treated outcome. Of course, with real data we can only observe either the control or the treated outcome at the same value of X; the average observed outcome is shown with the black line. The identified average treatment effect corresponds to

[1] We can consider this as an example of a continuous treatment that is a deterministic function of the running variable. See Section 3.3 for the more complex case of continuous treatments that are still random after we condition on the running variable.

[2] This could be considered as an application with multiple running variables, see Section 3.9 for more details.

the arrow at $X = c$. In this case we identify a positive average treatment effect at the threshold but the average causal effect is negative for low values of X. However, we cannot learn about the ATE away from the threshold without further assumptions. We discuss in Section 3.7 under which conditions it is possible to extrapolate the ATE away from the threshold and in Section 5 we use structural economic models to interpret the estimated effects.

2.2 Fuzzy design

In some cases, individuals are allowed to deviate from the rules. For instance, in Angrist and Lavy (1999), classes cannot have more than 40 students but school districts can decide to decrease the size of the classes even if the law does not force them. Thus, some school districts will not react to a change in the cohort size from 40 to 41 while other school districts will react by adding a new class. In another application, Jacob and Lefgren (2004) estimate the effect of attending a remedial summer school. Students who scored below a threshold at a test should participate. However, exceptions are possible, for instance if the child cannot attend the summer school or for students who passed the exams but were retained because of course failure.

In those cases, that are referred as fuzzy RDD, D is no longer a deterministic function of the running variable but we can observe a jump in the treatment probability at the cutoff: $\lim_{x \downarrow c} Pr\left(D = 1|X = x\right) > \lim_{x \uparrow c} Pr\left(D = 1|X = x\right)$, where the first term is the limit from above while the second term is the limit from below. The sharp RDD is obviously a special case of the fuzzy RDD with a jump in the treatment probability from 0 to 1 at the threshold. In both cases, we maintain the same continuity assumption for the expected value of the potential outcomes at the threshold. In the fuzzy design, the treatment is no longer locally randomized due to self-selection into the treatment, but the discontinuity represents a valid local instrumental variable.

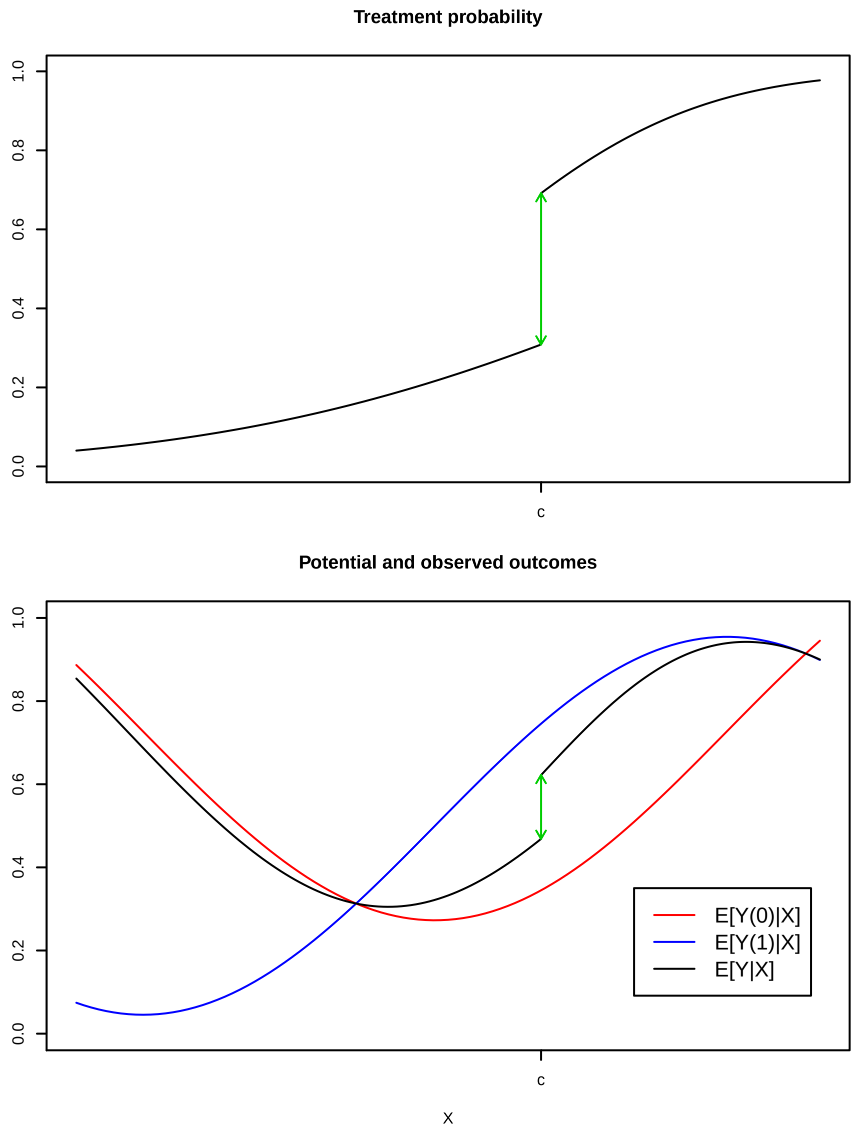
Locally, in the neighborhood of $X = c$, we have a binary treatment D and a binary instrument $1(X > c)$. This corresponds to the instrumental variable framework of Wald (1940). The expression

$$\frac{\lim_{x \downarrow x_0} E\left[Y|X = x\right] - \lim_{x \uparrow x_0} E\left[Y|X = x\right]}{\lim_{x \downarrow x_0} E\left[D|X = x\right] - \lim_{x \uparrow x_0} E\left[D|X = x\right]} \equiv \rho \tag{2.1}$$

identifies a causal parameter but there are two ways to interpret it. (i) We assume that the treatment effect is the same for all observations at the threshold. Then, $\rho = E[Y_i(1) - Y_i(0)|X = c]$. (ii) We allow for heterogeneous treatment effects but we impose the monotonicity restriction of Imbens and Angrist (1994). In the RDD context, this assumption consists in assuming that there are no individuals who would be treated if they were just below the threshold but would not be treated just above it. Thus, there are only three types of individuals: individuals who are treated just below and just above the threshold, individuals who are not treated just below and just above the threshold, and local compliers who are not treated if they are just below the threshold and are treated if they are just above it. With this assumption, Hahn et al. (2001) show that the Wald representation identifies the average treatment effect for these local compliers: $\rho = E[Y(1) - Y(0)|X = c, complier]$.

Figure 2.2 shows an artificial example of a fuzzy RDD. The top panel shows the treatment probability as a function of the running variable. There is a discontinuity at $X = c$ but the probability is not exactly 0 below this cutoff and is not exactly 1 above it. The bottom panel displays the expected values of both potential outcomes and of the observed outcome as functions of the running variable. In the fuzzy design, the expected value of the observed outcome at $X = x$ is a weighted average of the treated and control potential outcomes with

Figure 2: A fuzzy regression discontinuity design



weights $\mathrm{P}(D = 1|X = x)$ and $\mathrm{P}(D = 0|X = x)$, respectively. Thus, the size of the jump in $E[Y|X = x]$ at $x = c$ does not correspond to the average treatment effect.

However, under the assumptions mentioned above, we know that this jump is only due to the jump in the proportion of treated observations at $x = c$. Therefore, the fuzzy RDD is constructed out of two sharp RDD: one for the treatment probability and one for the outcome. The Wald representation in (2.1) consists in dividing the discontinuity for the outcome, which corresponds to the reduced form in instrumental variable models, by the discontinuity for the treatment, which corresponds to the first stage of the two-stage least squares estimator. If the discontinuity in $\mathrm{P}(D = 1|X = x)$ is small, we have a problem of weak identification, which we discuss in Section 3.5.

2.3 Estimation

Both in the fuzzy and in the sharp RDD, the identified parameters are functions of the conditional expected values at the boundary of the support of the running variable such as $\lim_{x \downarrow c} E[Y|X = x]$. Parametric estimation of these parameters incur a high risk of bias when the assumed functional form is not exact because we evaluate the function outside of the support of covariates in the estimation sample. Therefore, nonparametric estimation should be preferred.

There are two families of nonparametric estimation strategies: global and local estimators. Global estimators apply a high-dimensional (e.g. polynomial) model to the whole sample. Local estimators apply a low-dimensional (often a linear) model only to the part of the sample that is close to the threshold. When the objective is to estimate nonparametrically the whole function, these two strategies are asymptotically equivalent. However, in the RDD, we are only interested in the value of the function at one point, which is at the boundary of the support. In this case, following Gelman and Imbens (2018), we recommend using a local linear estimator. The global estimators give implicitly a high weight to observations far from the threshold, which is counter-intuitive. In addition, they are sensitive to the number of terms used in the polynomial approximation and we do not have a good way to select the optimal number of terms.

Operationalizing the local linear estimator requires choosing the weighting function (called kernel) and the bandwidth that determines the width of the estimation window. The first choice is relatively straightforward because the triangular kernel function is known to be optimal, see Cheng et al. (1997), or near-optimal, see Armstrong and Kolesár (2020), depending on the set-up. In practice, the results are rarely sensitive to this choice.

The choice of the bandwidth is a much more difficult problem and can strongly affect the results. On one hand, the bandwidth must not be too low, otherwise the effective estimation sample will be so small that we will not learn anything about the treatment effects (huge standard errors). On the other hand, the bandwidth must not be too large, otherwise the estimator will be biased like a parametric estimator. We want to find the bandwidth that balances the squared bias and the variance of the estimator and, therefore, minimizes the mean squared error (MSE) of the estimator.

More formally, consider the local linear estimator of $\lim_{x \downarrow c} E\left[Y|X=x\right]$ implemented with a bandwidth h and applied to a sample of size n. The MSE of this estimator is asymptotically

$$MSE = B^2 + V \approx h^4 \cdot C_B^2 + \frac{1}{n \cdot h} \cdot C_V \tag{2.2}$$

where B is the bias and V the variance of the local linear estimator while C_B and C_V are constants that depend on the kernel function and on the data generating process. If we minimize the MSE as a function of the bandwidth, we obtain the optimal bandwidth $h^* = \left(\frac{C_V}{4 \cdot n \cdot C_B^2}\right)^{1/5}$. To use this result we must estimate the constants C_V and C_B. While relatively straightforward estimators of C_V exist, C_B is proportional to $\lim_{x \downarrow x_0} \frac{\partial^2 E[Y|X=x]}{\partial x^2}$, the second-order derivative of the true function at the discontinuity. This parameter is often estimated with a new local regression, e.g. a local quadratic or cubic regression, which requires choosing a new bandwidth and making additional smoothness assumptions, see e.g. Imbens and Kalyanaraman (2012).

This issue concerns all nonparametric estimators. The particularity of the RDD estimators is that they are differences between two local linear estimators, one on the left and one on the right of the threshold. As a consequence, the relevant variance will be the sum of the variances (because the samples are independent) and the relevant bias will be the difference between the bias on the right and the bias on the left of the cutoff. If the second-order derivatives are the same on both sides, it follows that the 'optimal' bandwidth, which is derived under the assumption that it converges to zero, is infinite—a theoretical contradiction. In practice, the estimated optimal bandwidth can be very large and it will perform badly when the function is not globally linear. For this reason, Imbens and Kalyanaraman (2012) suggest to use a regularized bandwidth: a term is added in the denominator to make sure that the bandwidth converges to zero asymptotically.[3]

[3] It is possible to use a different bandwidth on the right and on the left of the threshold. Arai and Ichimura

Armstrong and Kolesár (2018, 2020) suggest another approach. Instead of estimating the second-order derivatives on both sides of the threshold, they assume that we can bound them.[4] Under this condition, they are able to derive (and consistently estimate) the bandwidth that minimizes the maximum MSE of the local linear estimator over all functions with bounded second-order derivative. An apparent drawback of this approach is the necessity to choose a bound for the second-order derivative of the functions. In reality, no method can avoid imposing–explicitly or implicitly–restrictions on the smoothness of the true function. For instance, estimating the second-order derivative requires assuming that the third order derivative is bounded and the choice of a pilot bandwidth. In Section 4, we discuss the choice of the bound on the second-order derivative in the context of an application.

We conclude this subsection with a few words about estimation of fuzzy RDD. In a fuzzy RDD, the parameter of interest can be written as a ratio of two sharp RDD. Therefore, the common approach consists in taking the ratio of two sharp RDD that are estimated separately. This can be implemented using a weighted two-stage least estimator if the same bandwidth is used for the numerator and denominator. Most researchers use the same bandwidth for the four local linear regressions. The delta method provides an asymptotic approximation of the MSE that is similar to the MSE of the sharp RDD. Thus, the previously described methods to select the bandwidth can be used for fuzzy RDD. Imbens and Kalyanaraman (2012), Calonico et al. (2014) and Armstrong and Kolesár (2020) extend their results to this setup.

2.4 Hypothesis tests and confidence intervals

When a rate-optimal bandwidth is used, e.g. the regularized bandwidth of Imbens and Kalyanaraman (2012), its modification by Calonico et al. (2014) or the minimax bandwidth of Armstrong and Kolesár (2020), then the local linear RDD estimator is asymptotically normally distributed. Inference is nevertheless not trivial because of the presence of a bias term; the asymptotic distribution of the t-statistic is not centered at zero:

$$\frac{\hat{\rho} - B - \rho}{sd(\hat{\rho})} = \frac{\hat{\rho} - \rho}{sd(\hat{\rho})} - \frac{B}{sd(\hat{\rho})} \sim N(0, 1). \tag{2.3}$$

If we neglect the bias, the confidence intervals will undercover the true parameter and the hypothesis tests will over reject a correct null hypothesis. Three different ways to solve this problem have been suggested: (i) select the bandwidth such that $B/sd(\hat{\rho}) \to 0$, (ii) estimate the bias B and make inference based on the bias-corrected statistic $(\hat{\rho} - \hat{B} - \rho)/sd(\hat{\rho} - \hat{B})$, which is centered at 0, (iii) bound the bias/standard deviation ratio and use the critical value that corresponds to the worst-case. In the following we provide more details on these approaches.

First, we can use a bandwidth that converges to zero faster than the optimal bandwidth (undersmoothing). In this case, the bias vanishes more quickly than the variance such that standard inference tools can be used. A practical disadvantage is that different estimators must be used for point estimation and hypothesis tests. A more fundamental weakness is that the confidence intervals obtained with undersmoothing will shrink slowly. In other words, it will be more difficult to reject incorrect null hypotheses.

Secondly, Calonico et al. (2014) suggest to use a rate-optimal bandwidth but they estimate the asymptotic bias and subtract it from the estimated effect. They show that it is important

(2018) exploit this possibility and show that it is in theory possible to remove the bias up to an arbitrary order by choosing adequately the bandwidths. However, there is no optimal solution to the problem of minimizing the mean squared error. For this reason, they modify the objective function and only try to remove the first-order term of the bias.

[4]Imbens and Wager (2019) assume that we can bound the derivative of the function not only at threshold but over the whole support of the running variable. They suggest an alternative method that is fully data-driven but does not allow for a closed-form solution.

to take into account the fact that the bias has been estimated when we compute the standard errors. Thus, they use a bias-corrected estimator with inflated standard errors. Unfortunately, this procedure is also not rate-efficient for principally the same reason as undersmoothing. In the case of the local linear estimator, Calonico et al. (2014) show that their robust bias-corrected procedure amounts to running a local quadratic regression with a bandwidth that is efficient for a local linear regression, i.e. a bandwidth that is smaller than the efficient one for local quadratic regression.

Third, Armstrong and Kolesár (2018, 2020) suggest to use a rate-optimal bandwidth without bias-correction but to take the bias into account by using larger critical values. When the second-order derivative is bounded and we use a MSE optimal bandwidth, then, in the worst-case, $|B|/sd(\hat{\rho}) = 0.5$. The reason is that the optimal bandwidth balances the bias squared and the variance such that none of them dominate asymptotically. It follows that, in the worst-case, $|\hat{\rho} - \rho|/sd(\rho) \sim |N(0.5, 1)|$. Thus, instead of using the traditional critical values, we must use the $1 - \alpha$ quantile of $|N(0.5, 1)|$. This amounts, for instance, to using 2.18 instead of 1.96 to build 95% confidence intervals. They show that no other approach to inference can substantively reduce the length of the confidence intervals while still maintaining coverage. We recommend this method for its simplicity and optimality.

For the fuzzy RDD, the asymptotic distribution of the estimator can be derived using the delta method. The first-order linear approximation can also be written as in equation (2.3), such that we can apply the same methods as for the sharp RDD. However, this approximation performs poorly in finite samples when identification is weak, i.e. when the discontinuity in the treatment probability is small; we discuss this issue in Section 3.5.

2.5 Falsification tests

In the RDD, like in any other causal design, identifying assumptions cannot be formally tested because they restrict counterfactual outcomes, which are not observable by definition. For instance, in the sharp design, the continuity assumption at $X = c$ cannot be tested because we observe only $Y(0)$ on the left of c and only $Y(1)$ on the right of c.[5] If an assumption can be tested, then it implies that it could be weakened until it becomes untestable. However, the credibility of assumptions can be assessed by testing restrictions that are stronger than what is strictly required but that are closely related. Often, the rationale for an identifying assumption has other implications that are testable.

For instance, it seems unlikely that $E[Y(0)|X = x]$ and $E[Y(1)|X = x]$ are continuous exactly at $x = c$ but are discontinuous at other values. Thus, the researcher may be willing to impose the continuity assumption over the whole support of the running variable. This stronger assumption is now testable. Tests can be implemented with the same procedures that have been developed to estimate the treatment effect, see Section 2.3. The idea consists in estimating the ATE at cutoffs that are different from c while using only observations on one side of the threshold to avoid that the true ATE contaminates this test. Under the null hypotheses of continuity, there is no effect (discontinuity) at the placebo thresholds.

Considering that the continuity assumption is difficult to interpret and not testable, Lee (2008) suggests to motivate it using the following model. He assumes that the running variable X is determined by two unobservable variables: one that is determined endogenously by the individuals and a noise that is revealed at a later stage. Thus, while the individuals can influence their probability of treatment, they do not have exact control over it. For instance, in Thistlethwaite and Campbell (1960), the running variable is a test score. The students have

[5] Actually, we only need an even weaker assumption: the conditional ATE $E[Y(1) - Y(0)|X = x]$ must be continuous at $x = c$. The continuity of the conditional ATE is obviously not testable at any x because we do not observe both potential outcomes at any x.

some control over their test score; they can work very hard or intentionally respond incorrectly to the questions. However, it is very unlikely that they can precisely manipulate their score to be just above or just below the threshold. This would require knowing with certainty all the responses. Lee (2008) shows that even a small random noise in the running variable is enough to satisfy the continuity assumption. In this framework, the RDD identifies a weighted ATE where the weights are proportional to the ex ante probabilities that $X = c$, see also Lee and Lemieux (2010).

In addition, this model produces two testable assumptions that are very often used to assess the validity of the RDD. First, the density of the running variable must be continuous at the threshold. Again, this is not a necessary (nor sufficient, even if Lee (2008) model is sufficient) condition for the validity of the RDD. There may be exact manipulation of the running variable that may be independent from the potential outcomes. For instance, Angrist et al. (2019) argue that the presence of a discontinuity in the distribution of the running variable (school enrollment) is due to school administrators who simply maximize their budget (and therefore the number of classes) and is independent from the potential test scores of the students.[6]

McCrary (2008) suggests a simple way to test the continuity of the running variable at the threshold. The idea consists in obtaining an histogram of the running variable and estimating its density by running local linear regressions separately on the left and on the right of the threshold. The null hypothesis that the density is the same on both sides can then be tested with a Wald test. Cattaneo et al. (2019) suggest a similar test that avoids pre-binning, Otsu et al. (2013) develops an empirical likelihood testing procedure, and Bugni and Canay (2020) proposes a sign test.

If there is a discontinuity in the running variable and we do not want to assume that the manipulation of the running variable is independent from the potential outcomes, we lose point identification of the average treatment effect. Gerard et al. (forthcoming) show that we can nevertheless bound the ATE. They assume that there are two types of units: those who manipulate the running variable and those we satisfy the standard RDD assumptions. The size of the discontinuity in the density identifies the proportion of units who manipulate the running variable. We can then bound the treatment effect by allowing these units to take the highest and lowest feasible values.[7]

A second testable implication of the model of Lee (2008) is the continuity in the distribution of pre-determined covariates at the threshold. If the agents cannot precisely control the value of the running variable around the threshold, then they are unable to sort (in particular) on the basis of variables that were determined prior to the realization of the assignment variable. This corresponds to the common practice in the analysis of randomized experiments of comparing the baseline characteristics of the treated and control groups. This test can be implemented simply by applying the RDD estimator with a pre-determined covariate as outcome. Canay and Kamat (2018) note that Lee's model implies the continuity of the whole distribution of the covariates and not only of their expected value. Therefore, they suggest a permutation test based on the whole distribution. Alternatively, the estimators of the whole distribution discussed in Section 3.1 can be used.

These two tests (continuity of the density of the running variable and continuity of the distribution of pre-determined covariates) can also be applied to assess the validity of fuzzy RDD. In addition, Arai et al. (2019) suggest a specification test for the heterogeneous fuzzy RDD that imposes monotonicity in the first stage. In a fuzzy RDD, there are agents that are

[6]In some cases, the discontinuity in the running variable may be explained by covariates, see Section 3.6 for more details.

[7]Manipulation of the running variable leads to bunching in the density of the running variable. Bunching can be used to estimate key behavioral parameters. See Kleven (2016) for an overview of the literature, and Chetty et al. (2011) or Saez (2010) for early applications of the bunching approach to estimating behavioral parameters.

(not) treated on both sides of the threshold. The observed outcome distribution should therefore be continuous at the threshold for these units. However, we do not observe them separately but only as part of a mixture with compliers. The implied restrictions on the observed outcome distribution are testable but not all deviations from the null hypothesis will be detectable even asymptotically.[8]

3 Extensions

In this section, we consider a variety of divergences from the standard RDD setting analyzed in the previous section: distributional effects of the treatment, discrete running variables, continuous treatments, estimated thresholds, weak discontinuities, role of covariates, extrapolation, multiple thresholds, multiple running variables, regression kink design.

3.1 Distributional and quantile treatment effects

Section 2 provides only results for average treatment effects. However, ATE provide only a limited view of the treatment effects. For instance, the treatment may not affect the average outcome while it harms half of the population and helps the other half. In some applications where the outcome is earnings, policy makers may be interested especially in the effects at the lower tail because they care about poor individuals. In other applications where the outcome is the unemployment duration, they may be especially interested in the upper tail of the distribution because they care about long-term unemployment. If we slightly strengthen the continuity assumption by imposing it on the whole distribution of both potential outcomes, then the RDD allows identifying and estimating these two distributions at the threshold.

By definition, the conditional distribution is a conditional expected value: $F_Y(y|x) = E[1(Y \leq y)|X = x]$, where $F_Y(y|x)$ denotes the conditional cumulative distribution function of Y evaluated at y given $X = x$ and $1(\cdot)$ is the indicator function. It follows that we can estimate the effects on the distribution by applying standard results for the outcome $1(Y \leq y)$ instead of Y. In the sharp RDD, the distribution of the control potential outcomes is identified as $F_{Y(0)}(y|c) = \lim_{x\uparrow c} E\left[1(Y \leq y)|X = x\right]$ and similarly for the treated outcome $F_{Y(1)}(y|c) = \lim_{x\downarrow c} E\left[1(Y \leq y)|X = x\right]$. The local linear estimator can be used to estimate these two distributions. A separate regression must be estimated for each level y. Note that this method can be used without modification for continuous, discrete and mixed outcomes.[9]

In the fuzzy RDD things are slightly more complicated because there are control and treated units on both sides of the cutoff. Frandsen et al. (2012) show that the distribution of both potential outcomes are identified using two separate Wald representations for $1(Y \leq y)$ interacted with the treatment status. We provide here the expression for the treated outcome:

$$F_{Y(1)}(y|c) = \frac{\lim_{x\downarrow c} E\left[1(Y \leq y)D|X = x\right] - \lim_{x\uparrow c} E\left[1(Y \leq y)D|X = x\right]}{\lim_{x\downarrow c} E\left[D|X = x\right] - \lim_{x\uparrow c} E\left[D|X = x\right]} \tag{3.1}$$

[8]In this case also, it is possible to weaken the identifying assumptions while preserving identification. We only need to assume that the conditional ATE for the compliers $E[Y(1) - Y(0)|X = x, complier]$ is continuous at the threshold, which is fundamentally untestable.

[9]The conditional distribution can also be estimated with a local binary regression estimator, such as logit or probit. It ensures that the estimated distribution function lies between 0 and 1. However, the local linear and binary estimators have exactly the same asymptotic distribution.

More structure can be enforced in the form of a local parametric model for Y. For instance, a local ordered probit model can be assumed for an ordered outcome, a local multinomial logit model for an unordered outcome, and a local Cox proportional hazard model for a duration outcome. This additional structure will improve the precision of the estimator under correct specification but can lead to a bias under misspecification.

Also in this case we can use estimators very similar to the estimators introduced to estimate ATE in fuzzy RDD.

In both sharp and fuzzy designs, the distribution functions of $Y(0)$ and $Y(1)$ are identified at the threshold. Therefore, we can also identify any function of these two distributions. For instance, when we are interested in the effect of the treatment on inequality, we can identify its effect on the variance, the coefficient of variation or the Gini coefficient of the outcome. The quantile treatent effect is a popular parameter that provides an intuitive way to report the effect of a treatment on the distribution of the outcome. The τ quantile treatment effect at the cut-off point (QTE) is simply

$$F^{-1}_{Y(1)}(\tau|c) - F^{-1}_{Y(0)}(\tau|c); \tag{3.2}$$

that is, the difference between the τ quantile of the treated and control outcome distributions, where $0 < \tau < 1$. Instead of a single QTE, we may also report the whole quantile treatment effect function. It allows testing for the homogeneity of the effects (the treatment only shift the location of the distribution of the outcome). Under an additional rank preservation assumption, the τ quantile treatment effect can be interpreted as the individual effect for units at the τ quantile of the outcome distribution. In that case, we can also recover the joint distribution of both potential outcomes at the threshold.

Frandsen et al. (2012) show that the sharp and fuzzy RDD allows identifying distributional treatment effects. They suggest a local linear estimator for the whole distributions, use a plug-in approach to estimate any functional of the distributions, and show uniform convergence of the whole estimated distribution and quantile functions. Shen and Zhang (2016) propose uniform tests based on the estimated distribution functions. Qu and Yoon (2018) suggest a quantile regression based estimator for the sharp RDD and provide tools for functional inference. While all these papers use undersmoothing for inference; Chiang et al. (2019) propose a bias-correction for the estimated quantile treatment effects.

3.2 Discrete running variables

When the running variable is discrete, it is no longer possible to find treated and control units with values of the running variable that are arbitrarily close. Since the running variable can have a direct effect on the outcome, we lose nonparametric point identification. In an influential paper, Lee and Card (2008) suggest to treat the specification error as random. In practice, they suggest to use clustered standard errors at the level of the running variable. Kolesár and Rothe (2018) show formally, in simulations, and in applications, that the clustered standard errors do *not* improve the quality of inference. In many relevant cases, they will even exacerbate the under-coverage of the confidence intervals. We strongly recommend against using this procedure.

Instead, we must accept that the treatment effects are not nonparametrically identified. If we allow $E|Y(0)|X = x]$ and $E[Y(1)|X = x]$ to vary arbitrarily between two points in the support of a discrete X, then the treatment effect is unbounded. Assuming continuity of these functions at the threshold, which is enough to identify the effect for continuous running variables, does not help for discrete running variables. We need to impose additional restrictions on the true functions. A natural assumptions consists in bounding the second-order derivative of both conditional expected values with respect to x. Kolesár and Rothe (2018) shows that the ATE is partially identified under this assumption. For inference, they note that the t-statistic can be decomposed as

$$\frac{\hat{\rho} - \rho}{sd(\hat{\rho})} + \frac{B}{sd(\hat{\rho})} \tag{3.3}$$

where the first term is standard normally distributed and the second is bounded. It follows that honest confidence intervals can be computed using the results in Armstrong and Kolesár (2020).

The only difference with the continuous case is the non-vanishing length of the confidence interval.

Noack and Rothe (2020) suggest a similar procedure for the fuzzy design. We discuss it below in Section (3.5) because it is also robust against weak identification.

3.3 Continuous treatments

If the treatment D is a deterministic function of the running variable X, then we can simply redefine the treatment of interest as a binary variable equal to 0 if D takes the value just below the cutoff and 1 if D takes the value just above the cutoff. The treatment effect of this particular change in D is identified using the standard sharp design. This corresponds for instance to the application in Lalive (2008).

If D is not a deterministic function of the running variable but there is a jump in $E[D|X = x]$ at $x = c$, then we can use the local Wald ratio for the fuzzy RDD defined in (2.1). Theorem 1 in Angrist et al. (2000) implies that this ratio identifies a weighted average of the derivative of $Y(d)$ with respect to d. In the case where $Y(d)$ is linear, then the local Wald ratio identifies the slope of the function. Dong et al. (2019) impose additional structure in the form of rank preservation for the treatment variable below and above the threshold. Under this assumption, they are able to utilize not only the mean change in the treatment but all changes along its distribution.

3.4 Estimated thresholds

Usually, we know the threshold c at which the treatment probability is discontinuous because it appears in a law or some sort of written rules. However, there are cases where the cutoff point is unknown. In van der Klaauw (2002) the discontinuity point was not disclosed to avoid manipulations. In Card et al. (2008) white residents leave a neighborhood as soon as the minority share reaches a tipping point, which is generally unknown.

Porter and Yu (2015) consider these cases. In the sharp design, they assume that the treatment itself is not observed.[10] They estimate the jump in the outcome at each possible value for the threshold. Then, they estimate the location of the discontinuity as the value that maximizes the size of this jump. Interestingly, this estimator of the discontinuity point is super-efficient: it converges so quickly to the true discontinuity point that inference can be performed as if we knew it. In the fuzzy design, the treatment is observed but not the threshold. The location of the discontinuity point can be estimated using the same procedure but applied to the treatment probability.

3.5 Weak discontinuities

The fuzzy RDD estimator is an instrumental variable estimator where the size of the jump in the treatment probability measures the strength of the instrument. Conventional methods for inference are unreliable when the discontinuity is small, i.e. when the instrument is weak. This problem is potentially magnified by the small effective sample size in nonparametric RDD settings. Feir et al. (2016) document the resulting size distortions in theory and in applications. The root of this problem is the division by the estimated first-stage discontinuity, which might be close to zero.

Feir et al. (2016) suggest inference tools that are robust against weak identification. They adapt an idea of Anderson et al. (1949) to the RDD: If the true ATE is equal to ρ_0, then we

[10] If the treatment is observed in a sharp design, the estimation problem will degenerate to the case of a known discontinuity.

should find no discontinuity in $E[Y - \rho_0 D|X = x]$ at $x = c$. We can test this null hypothesis using a sharp RDD, which circumvents the need to divide the reduced-form effect by the first-stage effect. We can then construct confidence intervals as the union of all the ρ_0 that are not rejected by these simple t-tests. These confidence intervals are valid even when there is no discontinuity in the treatment probability; of course, they will be unbounded in such a case because we do not have any source of random variation. Feir et al. (2016) assume that the running variable is continuous and they use undersmoothing. Noack and Rothe (2020) combine the insights of Anderson et al. (1949) and Armstrong and Kolesár (2020). They suggest confidence intervals that are robust against weak identification, allow for discrete running variables (see Section 3.2), and are more efficient than the confidence intervals based on under-smoothing. Their procedure requires specifying a bound on the second derivative of the functions.

3.6 RDD with covariates

In many applications, in addition to X, D, and Y, we also observe some covariates W. We may want to include these covariates in the estimation procedure for two reasons: to increase precision or to recover identification. First, covariates can increase the precision of the estimator even if they are not needed for identification. This is well-known in the context of randomized experiments and explains why covariates are often added in that case. In the RDD setting, covariates are not required for identification if their distribution is balanced at the cutoff, which is exactly an implication of the model of Lee (2008) and is often used as a falsification test, see Section 2.5. Second, when the distribution of the covariates is discontinuous at the cutoff, we can restore identification by including all the covariates that are discontinuous at the cutoff and correlated with the potential outcomes . For instance, Black (1999) exploits school district borders (geographic discontinuity) to estimate the impact of school quality on housing prices. One issue is that the quality of the houses also changes at the school border because different types of households are interested in them. Therefore, she controls for the characteristics of the houses when he estimates the effect of school quality on house prices. In other applications where it is hoped that the RDD assumptions are satisfied without covariates, estimators that incorporate control variables can be used as robustness checks or as falsification tests.[11]

Calonico et al. (2019) suggest an estimator that is straightforward to implement but achieves only the first objective. They consider the case where the estimator without covariates is consistent and the covariates are balanced at the cutoff. They recommend to simply include the covariates linearly in the traditional RDD estimation procedure. This estimator is nonparametric only in X but it is consistent even if the effects of the covariates are misspecified. It is very simple to implement, does not require the choice of new smoothing parameters, and can easily accommodate continuous and discrete covariates. They show that including covariates (weakly) reduces the variance of the estimator if the slopes on the covariates are identical on both sides of the cutoff. On the other hand, this estimator does not restore identification when the covariates are imbalanced.

Frölich and Huber (2018) suggest an estimator that achieves both objectives at the price of a more complex procedure. They allow the distribution of the covariates to be discontinuous at the threshold. In a first stage, they estimate the ATE conditionally on the control variables using smoothing both in X and W. This requires choosing another bandwidth parameter but permits analyzing the heterogeneity of the effects with respect to W. This first step estimator suffers from the curse of dimensionality in the number of continuous variables. In a second stage, they estimate the unconditional ATE by integrating over the distribution of W, which avoids the curse of dimensionality. Their estimator can restore consistency if the covariates are

[11]See Section 6.2.1 in Frölich and Sperlich (2019) for a more detailed discussion.

imbalanced at the cutoff and it reduces the estimation variance if the covariates are balanced at the cutoff.

3.7 Extrapolation

The RDD identifies only local effects, i.e. the effect for units exactly at the threshold. Formally, the population for which the effects are identified has a zero probability mass. In addition, fuzzy RDD with heterogeneous effects identifies only the effect for the local compliers. Naturally, this motivates researchers to explore strategies to generalize the identified effects.

Dong and Lewbel (2015) show that, under very slightly stronger assumptions, the RDD also identifies nonparametrically the derivative of the treatment effect with respect to the running variable at the cutoff. Under a policy invariance assumption, this derivative provides the change in the treatment effect due to a marginal change in the threshold. Consider for instance an application where this derivative is very small such that we cannot reject that the treatment effect is constant. It means that a marginal change in the threshold c would not change significantly the estimated effects. This certainly increases the credibility of extrapolations away from the cutoff.

Another approach consists in using the RDD to test assumptions that identify the treatment effect for the whole (or at least a larger) population. Angrist and Rokkanen (2015) test whether controlling for observable control variables is enough to capture the correlation between the running variable and the potential outcomes. This assumption has testable implications below the cutoff for the control outcome and above it for the treated outcome. If this assumption holds, the sharp and fuzzy RDD identify the average treatment effect for the whole population and the compliers, respectively. Battistin and Rettore (2008) and Bertanha and Imbens (2019) consider only the fuzzy design.[12] They note that it is possible, at the threshold, to test the assumption that the potential outcomes are independent from the types (compliers, always-takers and never-takers), possibly after conditioning on some control variables. If this assumption is correct, it implies that (i) the ATE for the compliers is equal to the ATE for the whole population and (ii) the treatment is exogenous. Thus, it allows extrapolating the identified treatment effect. A common weakness of the methods discussed in this paragraph is that we can never completely verify identifying assumptions even asymptotically. The lack of rejection increases the credibility of the assumptions but cannot prove their validity.

3.8 Multiple thresholds

In some applications, the threshold c at which the treatment probability is discontinuous may be different for different units. For instance, different regions may have a different cutoff for college admission. In another application, the vote share needed to win an election may be different in different electoral districts. In such a case, researchers can analyze each threshold separately with standard methods. They have to be careful that each sample includes only one threshold. Alternatively, they can normalize the running variable such that the cutoff is the same for all units, for instance by subtracting the original threshold from the observed running variable. Cattaneo et al. (2016) show that the pooled RDD estimator converges to a weighted average of cutoff-specific treatment effects for the populations facing different cutoffs. This estimand is difficult to interpret when the treatment effects are heterogenous.

Bertanha (forthcoming) considers a setup where the number of observations and the number of cutoffs grow to infinity. He assumes that the treatment effect does not depend on the cutoff after we take the value of the running variable into account, i.e. all cutoffs affect the same population. Under these assumptions, he shows that it is possible to weight the local effects in

[12] Battistin and Rettore (2008) analyze only cases with one-sided perfect compliance.

such a way that it consistently estimates the ATE for the whole population. Cattaneo et al. (forthcoming) do not require that the number of cutoffs increases with the sample size and allow the cutoff to directly affect the treatment effect. Instead, they make a common bias assumption across cutoffs, which is similar to the common trend assumption in the difference-in-differences setup. This assumption allows extrapolating the causal treatment effects to values of the running variables different from observed cutoffs.

3.9 Multiple running variables

Treatment assignment rules can depend on several running variables. For example, in Jacob and Lefgren (2004), students are assigned to a summer school if their test score in reading *or* their test score in mathematics (or both) is below a threshold. Similarly, school graduation requirements often include minimum grades in several fields of study. Changes in policies at geographic boundaries represent another popular example, see e.g. Lalive (2008), Dell (2010), and Keele and Titiunik (2015). In that case, latitude and longitude are the running variables that define the treatment assignment sets.

With multiple running variables, the discontinuity becomes a boundary and it is possible to estimate the local ATE at any point of the boundary (with a positive density of the running variable), see Imbens and Zajonc (2011). The estimated ATE function allows analyzing the heterogeneity of the effects with respect to the running variables but it is imprecisely estimated because it suffers from the curse of dimensionality. Therefore, in most applications, an integrated ATE is estimated. This parameter can be estimated by averaging the conditional ATE or by reducing the problem to a one-dimensional running variable. This second alternative is simpler to implement and, therefore, much more common in the literature. Usually, the distance to the nearest boundary is used as a running variable and the problem simplifies to a standard one-dimensional RDD. Jacob and Lefgren (2004) keep only students who scored above the reading threshold such that the problem simplifies to a simple RDD with the mathematics test score.

3.10 Regression kink design

The RDD exploits a discontinuity or jump in the treatment function $E[D|X = x]$ at a known cutoff c. The regression kink design (RKD) exploits a jump or discontinuity in the derivative of the treatment function $dE[D|X = x]/dx$ at a known value c. The RKD is simply a RDD in the first derivative. If we plot $E[D|X = x]$ as a function of x, we observe a kink in this function, which accounts for the name of this design. Card et al. (2015) for continuous treatments and Dong (2016) for binary treatments show that, under appropriate conditions,

$$\frac{\lim_{x\downarrow c} dE[Y|X = x]/dx - \lim_{x\uparrow c} dE[Y|X = x]/dx}{\lim_{x\downarrow c} dE[D|X = x]/dx - \lim_{x\uparrow c} dE[D|X = x]/dx} \tag{3.4}$$

identifies the average effect of a marginal increase in the treatment (probability). We have a sharp (fuzzy) RKD if the treatment is a deterministic (random) function of the running variable.

Local quadratic regression is recommended to estimate RDK (compared to local linear regression for RDD). Calonico et al. (2014) derive the asymptotic distribution of this estimator. Many of the results are similar for RDD and RDK but researchers must be aware that the rate of convergence of the RDK estimator is slower because it is more difficult to estimate the derivative than the level of a function. Various extensions have already been developed; for instance, Chiang and Sasaki (2019) introduce quantile RKD, Ganong and Jäger (2018) suggest a permutation test, and Hansen (2017) considers a setup with an unknown threshold.

4 Applied RDD

In this section, we apply the most commonly used procedures and some extensions to the data first analyzed by Lalive (2008). On our companion website, we provide the data as well as the Stata and R codes that allow replicating all the results. We make use of the 'rdrobust' and 'rddensity' Stata packages, described in Calonico et al. (2017) and Cattaneo et al. (2018), respectively. We also employ the 'RDHonest' R package written by Kolesár (2020).

In June 1988, the Austrian government extended the maximal duration of unemployment benefits from 30 to 209 weeks for job seekers aged 50 or older when entering unemployment, if they had lived for at least 6 months in certain regions of Austria. This regulation creates two discontinuities: one at an age of 50 in the treated regions and one at the geographical borders between regions for individuals older than 50. We exploit these two sharp discontinuities, separately for women and men, to estimate the effect of a longer maximal duration of unemployment benefits on unemployment duration.

Figure 3: Unemployment duration for men at the age threshold

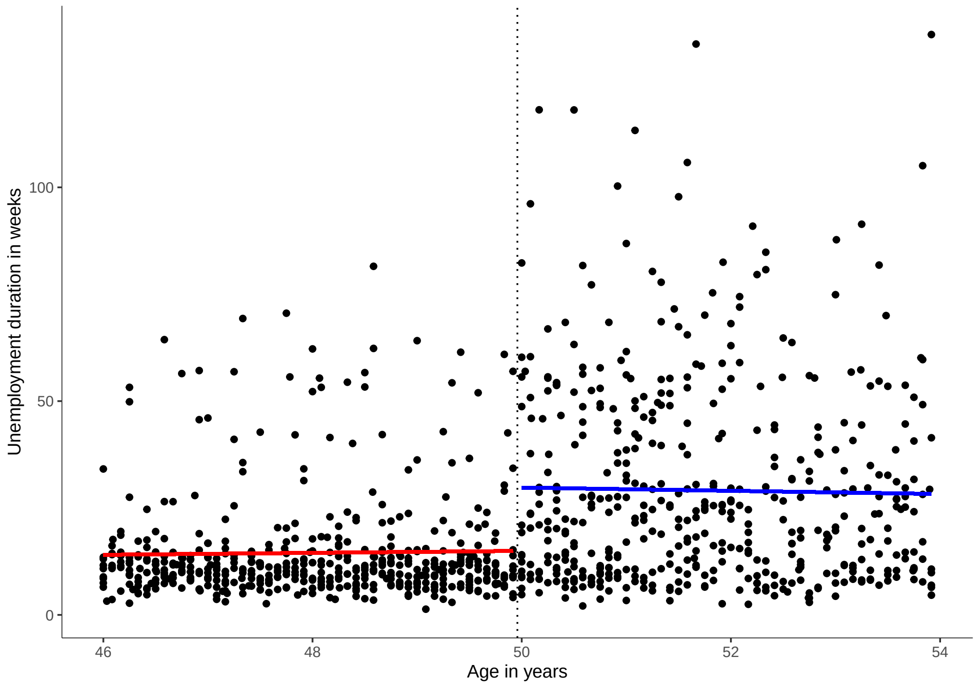


Figure 3 shows unemployment duration and age for 9734 men in the treated regions. Each dot represents 10 observations to avoid overloading the graphic. The lines show the fitted values of global linear regressions on the left and right of the thresholds. A few characteristics of the data are noticeable. First, the running variable is not perfectly continuous because we observe age only in month. We will check the results of methods assuming continuity with those of the procedures of Kolesár and Rothe (2018). Second, the outcome distribution is heavily right-skewed. In fact, the length of the unemployment spell is above the average for less than 15% of the observations. In such a case, the ATE will be dominated by the effect at the upper tail of the distribution. We will also report quantile treatment effects to analyze the heterogeneity of the effects. Third, the curvature of the functions does not seem to be very high. The results should not be very sensitive to the bandwidth choice.

Table 1 provides the estimated ATE of extended unemployment benefit duration as well as 95% confidence intervals using several methods. Panel A displays the results for men using age (in years) as the running variable. Even if there is some disagreement about the optimal choice of the bandwidth, the point estimates are stable with values ranging from 12.99 to 15.88.

Table 1: ATE of extended unemployment benefit duration

Method	Bandwidth	ATE	S.E.	C.I.	
Panel A: Men, age discontinuity					
CCT	0.95	14.9	5.48	4.17	25.63
CCT bias-corrected	0.95	15.88	6.57	3.01	28.75
IK	1.73	12.99	3.94	5.28	20.71
IK bias-corrected	1.73	14.47	5.9	2.9	26.05
Bias-aware, M=1	3.66	14.51	2.63	8.95	20.08
Bias-aware, M=8	1.54	13.21	4.21	4.43	21.99
Bias-aware, M=16	1.16	14.04	4.89	3.86	24.21
Panel B: Women, border discontinuity					
CCT	11.04	95.84	29.35	38.32	153.35
CCT bias-corrected	11.04	93.62	33.03	28.88	158.36
IK	39.65	41.51	6.43	28.9	54.12
IK bias-corrected	39.65	50.09	13.41	23.82	76.37
Bias-aware, M=0.05	27.12	48.60	8.89	24.15	73.04
Bias-aware, M=0.1	21.06	47.34	11.84	14.97	79.72
Bias-aware, M=0.2	17.57	42.93	15.84	-0.94	86.79

All estimates are based on the local linear estimator with a triangular kernel function. In each panel, the first line (CCT) provides the estimates obtained with the Calonico et al. (2014) bandwidth and the confidence intervals neglecting the bias. The second line (CCT bias-corrected) provides the bias-corrected results using the same bandwidth. The third line (IK) provides the estimates obtained with the Imbens and Kalyanaraman (2012) bandwidth and the confidence intervals neglecting the bias. The fourth line (IK bias-corrected) provides the bias-corrected results using the same bandwidth. The fifth, sixth, and seventh lines provide the results for the bandwidths that minimize the largest MSE over all functions with bounded second-order derivative for 3 different bounds M, as suggested in Armstrong and Kolesár (2020) and Kolesár and Rothe (2018). The corresponding confidence intervals take the maximal possible bias into account by using larger critical values.

While the point estimates on the first and third lines converge at the fastest possible rate, the effective coverage rate of the corresponding confidence intervals is below the theoretical one due to the bias. The bias-corrected confidence intervals suggested by Calonico et al. (2014) on the second and fourth lines cover the true effect with the correct probability but they are longer.

The "bias-aware" procedures for inference suggested in Armstrong and Kolesár (2020) are particularly relevant because they are optimal and accommodate discrete running variables without modification. They require the choice of M, an upper bound for the second-order derivative of the conditional average outcome on the left and right of the threshold. Intuitively, if we set $M = 0$, we are assuming that the true function is linear while a large value allows for a high curvature. More precisely, Kolesár and Rothe (2018) show that the largest deviation from a line on a segment of X of length 1 is bounded by $M/8$. In our application, given that the first-order derivative is close to zero, setting $M = 8$, which allows the true function to deviate by 1 from the line on each interval of length 1, seems to be a conservative bound.

Armstrong and Kolesár (2018) show that, without further restrictions, we cannot use the data to estimate M. However, a lower bound for the second-order derivative can be estimated from the data. In our case, only very small values of M are rejected by the data. Armstrong and Kolesár (2020) assume that local smoothness of the function is no smaller than its smoothness at large scales. This assumption justifies the following data-driven procedure: estimate a global flexible parametric model and set M equal to the largest second-order derivative over the support of X.[13] When we estimate cubic regressions on both sides of the cutoff the largest estimated second-order derivative attain 14. Higher-order polynomials give much higher bounds but they are very imprecisely estimated due to the limited support of X in our sample (from 46 to 54 years old).

For these reasons, Table 1 reports the results for $M = 1, 8, 16$. Mechanically, when M, the bound on the second derivative, increases, then the potential bias increases and the optimal bandwidth decreases to reduce this bias. It also follows that the bias-aware confidence interval gets wider when M increases. In Panel A of Table 1, the ATE is significantly different from 0 for all considered M because the effect is large enough.[14]

As discussed in Section 2.5, an important falsification test for a RDD consists in testing the continuity of the density of the running variable at the threshold. Figure 4 provides the results of this test as implemented by Cattaneo et al. (2019). For men, we do not find any evidence of a discontinuity. For women, however, there is a very clear jump in the density at the threshold. As discussed in Lalive (2008), this could reflect the fact that women have access to special income support at 54 years old (i.e. when their unemployment benefits are exhausted in the treated region). Thus, in the treated regions, they have a strong incentive to remain employed until they attain the cutoff of 50 years. For this reason, we will focus on the geographic discontinuity to identify the effects for women.

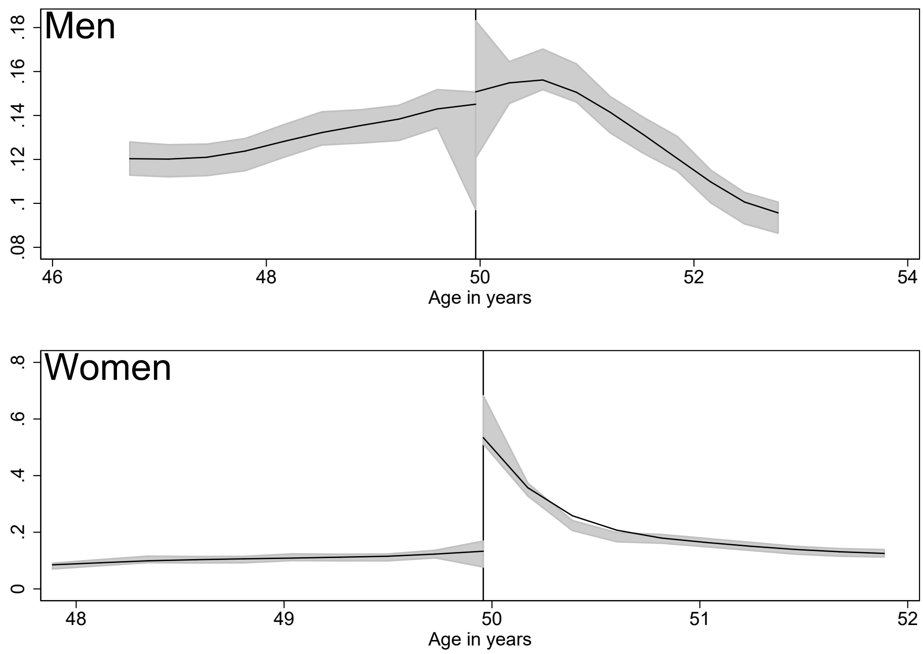
Figure 5 provides unemployment duration and distance to the border (in minutes of driving) for 7091 women between 50 and 54 years old. While we do not find a discontinuity in the density of the running variable, this density is low close to the regional border. This implies that the choice of the bandwidth might be more delicate; in particular, we cannot choose a very small bandwidth.

Panel B of Table 1 reports the results for women using the regional discontinuity. The ATE tends to be higher than for men but less precisely estimated. To select the bound M, we estimate a quartic polynomial regression on both sides of the cutoff. The largest estimated second-order derivative in the support of the data is equal to 0.1. Therefore, we consider the results with M set to this value as our main results. The corresponding confidence interval

[13]See Appendix E in the Online Supplemental Material of Armstrong and Kolesár (2020) for more details.

[14]Kolesár and Rothe (2018) present results with the same dataset, the same bounds, but a uniform kernel function.

Figure 4: Test of the continuity of the density at the age threshold



[14.97, 79.72] shows a significant, positive ATE. The treatment effect is barely insignificant if we double the upper bound.

Finally, motivated by Figures 3 and 5, we estimate the quantile treatment effects of the extended unemployment benefit duration. The top-left panel of Figure 6 shows that the treatment effects for men are very close to zero (and not statistically different from it) for the bottom 80% of the distribution.[15] We find large and significant effects only at the highest quantiles. The bottom-left panel shows that a very large proportion of the male population leaves anyway unemployment in less than 30 weeks. Thus, it is not surprising that an increase of the potential duration of the benefit from 30 to 209 weeks did not affect these individuals. The results for women on the right-side of Figure 6 are similar but less pronounced. The estimated quantile effects are close to zero for the bottom half of the distribution.

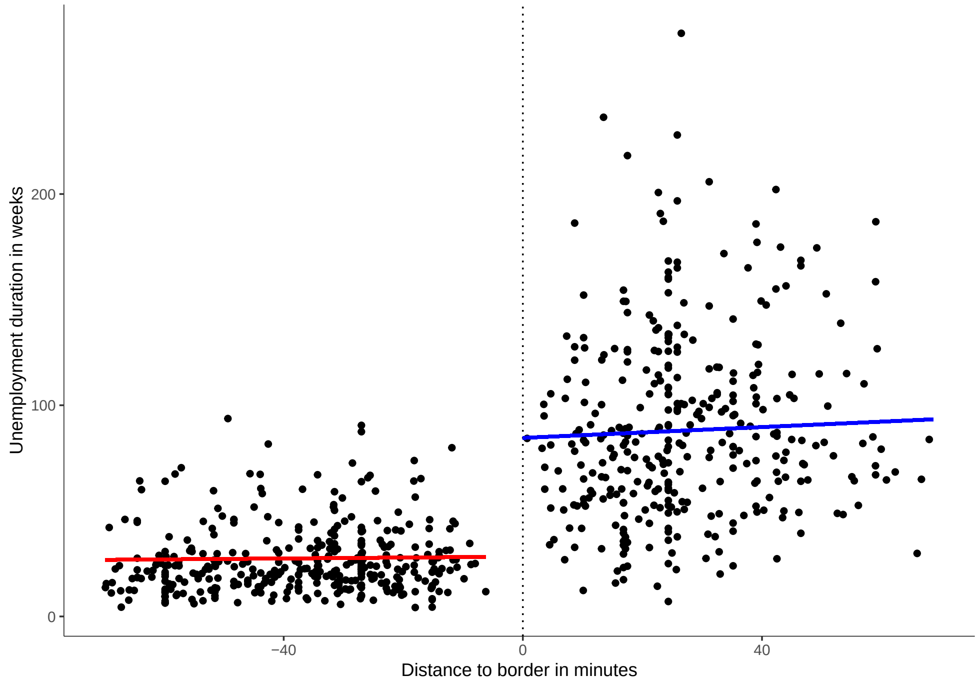
5 Interpreting RDD

This section discussed interpretation of the estimated treatment effects. RDD is a powerful and credible research design, and it delivers compelling estimates of the treatment effect. Beyond estimation, researchers are interested in how they can interpret the treatment effect. A reduced form interpretation, one that does not rely on additional structure, is that the RDD estimate is an average of the effects of the treatment on the outcome. This reduced form interpretation of the treatment effect is informative for answering policy relevant questions such as how much the outcomes change due to implementing a treatment, whether the benefits of the treatment justify its cost, but the treatment effect does not provide information on the underlying behavioral mechanism, and the associated utility or cost parameters.

In some contexts, the RDD treatment effects on two distinct outcomes provide information on optimal policy without more explicit structure. Schmieder et al. (2012) study age discontinuities in the potential benefit duration in the German UI system and compare the RDD estimates on two durations. The first duration is the time job seekers need to find a job,

[15]See Schmidt and Zhu (2016) for a similar result using a different method.

Figure 5: Unemployment duration for women at the regional border

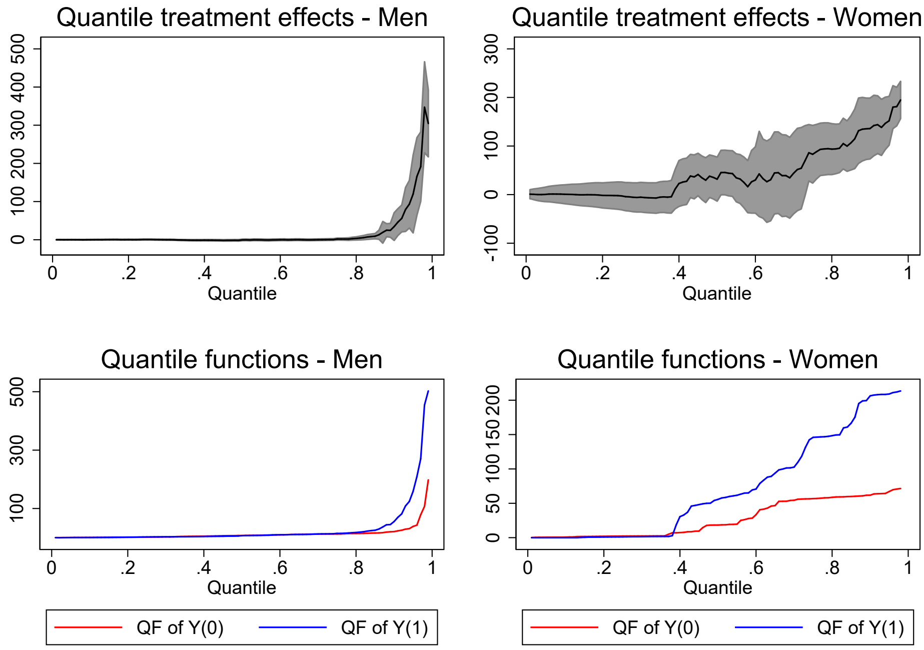


counted from the end of the job prior to entering unemployment. This duration provides information on the costs of benefit extensions inflicted by behavioral responses of job seekers. The second duration captures the additional days of benefit coverage to job seekers, which reflect the insurance value of extending benefits. The paper reports estimates of how job search and benefit payments respond to prolonged UI benefits over a period of 20 years, and finds that payment coverage rises sharply during recessions while behavioral costs are almost acyclical. This suggests that extensions of benefit durations in recessions, as is common in the U.S., is optimal. Comparing RDD estimates on these two distinct outcomes provides simple and intuitive rationale for optimal unemployment insurance policy over the cycle.

Linking RDD estimates to behavior often requires to spell out a behavioral framework. In the context of the job search setting of Section 4, suppose, for instance, that job offers pay a wage w, arrive at poisson rate $\lambda \times e$, where e is search effort, (bounded between 0 and 1), and λ is the baseline arrival rate of offers. Job seekers receive unemployment benefits, $b = \overline{b}$, during R periods, and social assistance $b = \underline{b}$ afterwards ($\overline{b} > \underline{b}$). Search is costly, and the per period utility from benefits and searching for jobs is $b(1-e)$, so benefits are worth more when devoting little effort to search, e.g. consumption and leisure are complements. In this simple context, job seekers either search at full intensity, or not at all. Job seekers who search for jobs compare the expected gain from taking a job, $\lambda(w - \overline{b})$, with the cost of searching at full intensity ($e = 1$), which is $\overline{b}$. When the search cost is still high, in the covered part of the spell, when benefits are still high, job seekers will not look for a job if $\lambda(w - \overline{b}) < \overline{b}$. But once benefits run out, both the benefits from searching, and the costs of searching increase, and more job seekers will look for jobs. This simple framework rationalizes the finding that extended benefits only affect the higher quantiles of the job search duration distribution (Figure 6).

RDD estimates can provide powerful information to estimate key parameters of the underlying model. Lacetera et al. (2012) argue that inattention explains a salient phenomenon in the used cars market. The values of a used car declines with the number of miles it has been driven so far, but cars that have, e.g. 10'002 miles on their odometers, sell at substantially lower prices than cars with , e.g. 9'989 miles on their odometer, so there is a discontinuity in price. Lacetera et al. (2012) claim that this can be rationalized by buyers having left digit

Figure 6: Quantile treatment effects



bias, e.g. buyers tend to focus on the left-most digit of a number. The paper supposes that the perceived odometer reading is $\hat{m} = d_H 10^H + \sum_{j=1}^{\infty}(1-\theta)d_{H-j}10^{H-j}$, where d_H is the highest digit in a number, and θ is the inattention parameter. A buyer with $\theta = 1$ would perceive an odometer of 10'002 to be 10'000 for the car with miles, and 9'000 for the car with 9'989 miles. The perceived difference is 1'000 miles instead of the actual difference of 13 miles on the cars' odometers. Using RDD, Lacetera et al. (2012) show that there are striking discontinuities in the selling price of cars that are just on either side of a (multiple of) 10K miles threshold. Using non-linear least squares, the paper estimates the inattention parameter θ to be 0.31 (with standard error 0.01), leveraging identification from the (multiples of) 10K miles thresholds where the perceived valuation of a car differs strongly from its actual value.[16]

This example interprets the reduced form RDD estimates through the lens of stylized models. An alternative is to use the data to estimate the key underlying structural parameters. DellaVigna et al. (2017) argue that reference dependence drives the pattern of job search. Newly registered job seekers find that the unemployment benefit is much lower than their income on the job, which is their reference point, and dedicate a lot of time to job search to escape unemployment. Also, job seekers who approach benefit exhaustion look for jobs intensively, and gradually reduce job search intensity as they get used to even lower income after exhausting benefits. The paper tests this hypothesis in a reform in Hungary that reduces the unemployment benefit payments in the latter parts of the unemployment spell. The empirical exit rates are very much in line with the prediction, and structural estimates indicate that the model with reference dependence fits the data much better than a standard model with e.g. habit formation.[17]

Structural estimation can be helpful in situations where the key question can not be answered with the data alone. Lalive et al. (2014) discuss job search during parental leave and

[16]Englmaier et al. (2018) find a similar pattern in a European Auction, but document in addition that prices vary discontinuously with registration year. Strittmatter and Lechner (2020) study sorting in the used car market after the VW manipulation scandal, and find strong inflow of cars that were supposed to be manipulated and price reductions on these cars.

[17]DellaVigna (2018) provides a guide to estimating structural models with behavioral features.

focus on the specific role of the guarantee to be able to return to the previous job, job protection, and cash benefits being paid to mothers who care for their children. Austria reformed these two elements in the 1990s and early 2000s, with the day of birth of a child determining the type of policy that that family was exposed to – day of birth RDD. Unfortunately, the reforms changed the time women were eligible for benefits several times, but not the time until job protection is guaranteed, so to separate the role of benefits and job protection. To better understand the role of these two policy parameters, the paper sets up a model of return to work from parental leave, and estimates its main parameters over a time period without any policy change. Using model estimates, the paper predicted behavior during the actual parental leave reforms to validate the model estimates. The paper then uses the model estimates to simulate job search on parental leave without benefits, or job protection. The paper finds that cash benefits ensure large take-up of parental leave, and job protection ensures job continuity after the end of parental leave. Both policy parameters complement each other to create time for parents to care for children while maintaining a smooth return to work once parental leave is over.[18]

RDD can provide essential insights into decisions with spillovers. Fu and Gregory (2019) study the context in New Orleans after Hurricane Katrina devastated large segments of the city. To support rebuilding, Louisiana created the Road Home (RH) rebuilding program which offered large subsidies for households whose damage index was above a threshold. Basic reduced form evidence shows that households above the threshold were more likely to engage in rebuilding efforts, but also their neighbors, whose incentives to rebuild were not directly modified through the grand, were more actively rebuilding their homes. The paper develops an equilibrium model of rebuilding after Katrina that takes grants and spillovers into account. Based on this model Fu and Gregory (2019) provide estimates of counterfactual policies, optimal policy, and assess the welfare benefits of the RH grants. This study illustrates how compelling reduced form evidence motivates an explicit equilibrium framework to answers key questions that can not be addressed with reduced form assessments alone. Yet the strength of the reduced form estimates is the key building block for the more comprehensive behavioral model.

6 Conclusions

The RDD is one of the most credible and transparent identification strategy. We have reviewed recent developments in the estimation of and inference on treatment effects in the RDD. The literature provides clear guidance about the method–local linear regression–and the weighting function–triangular kernel–to use. On the other hand, the choice of the bandwidth remains a serious challenge. Researchers would like to have a fully data-driven and objective procedure to choose the smoothing parameter. However, as shown in Kamat (2018) and Bertanha and Moreira (2020), it is impossible to distinguish a null from an alternative hypothesis if we do not restrict the class of models, for instance by bounding the first-order derivative of the potential outcome functions. Thus, we cannot avoid making assumptions about the true functions that we want to estimate. Armstrong and Kolesár (2018) suggest to explicitly and transparently assume that the first-order derivative of the potential outcome functions is bounded by a known value. With this assumption, the optimal bandwidth is easily estimated and the efficient confidence intervals take a simple form. This approach is also applicable when the running variable is discrete. The bound on the first-order derivative has, thus, many useful implications but it is not obvious how to choose it in an application. There is a natural temptation to estimate this

[18] This approach is inspired from the "ex ante" evaluation approach, which consists in modeling the behavior structurally, and predicting the effects of actual and counterfactual policy changes. Todd and Wolpin (2020) survey approaches to estimation that rely both on randomized experiments and structural modeling.

bound from the data but, without further assumption, a data-driven choice of the bound would jeopardize the good properties of the resulting inference. This unavoidable choice will remain the crux of any RDD implementation.

The traditional RDD identifies the average treatment effect in a set-up that consists of a unique continuous running variable, a known unique threshold, a binary treatment that is a deterministic function of the running variable, and no covariates. We have also reviewed the most interesting extensions of this standard RDD set-up. Instead of focusing on the average effect, quantile treatment effects provide the treatment effect on the whole distribution of the outcome. They allow analyzing the heterogeneity of the effect and can inspire the appropriate interpretation of the findings. When there are several thresholds or several running variables, we can either normalize the variables such that it fits into the standard set-up or we can analyze the heterogeneity of the effects. When the threshold is not known, it can be estimated. Including covariates may be needed to recover the identification of causal effects or may be included to increase the precision of the estimates. Finally, we have seen that the regression kink design is a fuzzy RDD in first differences.

In Section 5, we have discussed how reduced form RDD can be combined with theoretical economic models to estimate deep structural parameters. One possibility highlighted consists in interpreting reduced form evidence through the lens of theoretical models, even without actually estimating them. An alternative is to use the data to estimate the key underlying structural parameters exploiting an RDD for identification. Structural estimation of models can provide answers that can not be addressed with pure reduced form estimation.

Nearly 60 years after its first use, RDD is very much alive. On the theoretical side, the debate about optimal inference is open and extensions to new set-ups are being proposed. On the applied side, the RDD has helped researchers study important questions in nearly all domains of economics.

References

Anderson, T. W., Rubin, H., et al., 1949. Estimation of the parameters of a single equation in a complete system of stochastic equations. The Annals of Mathematical Statistics 20 (1), 46–63.

Angrist, J., Graddy, K., Imbens, G., 2000. The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. Review of Economic Studies 67, 499–527.

Angrist, J. D., Lavy, V., 1999. Using maimonides' rule to estimate the effect of class size on scholastic achievement. The Quarterly journal of economics 114 (2), 533–575.

Angrist, J. D., Lavy, V., Leder-Luis, J., Shany, A., 2019. Maimonides' rule redux. American Economic Review: Insights 1 (3), 309–24.

Angrist, J. D., Pischke, J.-S., 2009. Mostly harmless econometrics: An empiricist's companion. Princeton university press.

Angrist, J. D., Rokkanen, M., 2015. Wanna get away? regression discontinuity estimation of exam school effects away from the cutoff. Journal of the American Statistical Association 110 (512), 1331–1344.

Arai, Y., Hsu, Y., Kitagawa, T., Mourifie, I., Wan, Y., 2019. Testing identifying assumptions in fuzzy regression discontinuity designs. Tech. rep., cemmap working paper.

Arai, Y., Ichimura, H., 2018. Simultaneous selection of optimal bandwidths for the sharp regression discontinuity estimator. Quantitative Economics 9 (1), 441–482.

Armstrong, T. B., Kolesár, M., 2018. Optimal inference in a class of regression models. Econometrica 86 (2), 655–683.

Armstrong, T. B., Kolesár, M., 2020. Simple and honest confidence intervals in nonparametric regression. Quantitative Economics 11 (1), 1–39.

Battistin, E., Rettore, E., 2008. Ineligibles and eligible non-participants as a double comparison group in regression-discontinuity designs. Journal of Econometrics 142 (2), 715–730.

Bertanha, M., forthcoming. Regression discontinuity design with many thresholds. Journal of Econometrics.

Bertanha, M., Imbens, G. W., 2019. External validity in fuzzy regression discontinuity designs. Journal of Business & Economic Statistics, 1–39.

Bertanha, M., Moreira, M. J., 2020. Impossible inference in econometrics: Theory and applications. Journal of Econometrics.

Black, S., 1999. Do 'better' schools matter? parental valuation of elementary education. Quarterly Journal of Economics 114, 577–599.

Bugni, F. A., Canay, I. A., 2020. Testing continuity of a density via g-order statistics in the regression discontinuity design. Journal of Econometrics.

Calonico, S., Cattaneo, M. D., Farrell, M. H., Titiunik, R., 2017. rdrobust: Software for regression-discontinuity designs. The Stata Journal 17 (2), 372–404.

Calonico, S., Cattaneo, M. D., Farrell, M. H., Titiunik, R., 2019. Regression discontinuity designs using covariates. Review of Economics and Statistics 101 (3), 442–451.

Calonico, S., Cattaneo, M. D., Titiunik, R., 2014. Robust nonparametric confidence intervals for regression-discontinuity designs. Econometrica 82 (6), 2295–2326.

Campbell, D. T., Stanley, J. C., 1963. Experimental and quasi-experimental designs for research on teaching. In: Gage, N. (Ed.), Handbook of research on teaching. Rand McNally & Co., Chicago.

Canay, I. A., Kamat, V., 2018. Approximate permutation tests and induced order statistics in the regression discontinuity design. The Review of Economic Studies 85 (3), 1577–1608.

Card, D., Lee, D. S., Pei, Z., Weber, A., 2015. Inference on causal effects in a generalized regression kink design. Econometrica 83 (6), 2453–2483.

Card, D., Mas, A., Rothstein, J., 2008. Tipping and the dynamics of segregation. The Quarterly Journal of Economics 123 (1), 177–218.

Cattaneo, M. D., Jansson, M., Ma, X., 2018. Manipulation testing based on density discontinuity. The Stata Journal 18 (1), 234–261.

Cattaneo, M. D., Jansson, M., Ma, X., 2019. Simple local polynomial density estimators. Journal of the American Statistical Association, 1–7.

Cattaneo, M. D., Titiunik, R., Vazquez-Bare, G., 2020. The regression discontinuity design. In: Curini, L., Franzese, R. J. (Eds.), Handbook of Research Methods in Political Science and International Relations. Sage Publications.

Cattaneo, M. D., Titiunik, R., Vazquez-Bare, G., Keele, L., 2016. Interpreting regression discontinuity designs with multiple cutoffs. The Journal of Politics 78 (4), 1229–1248.

Cattaneo, M. D., Titiunik, R., Vazquez-Bare, G., Keele, L., forthcoming. Extrapolating treatment effects in multi-cutoff regression discontinuity designs. Journal of the American Statistical Association.

Cheng, M.-Y., Fan, J., Marron, J. S., et al., 1997. On automatic boundary corrections. The Annals of Statistics 25 (4), 1691–1708.

Chetty, R., Friedman, J. N., Olsen, T., Pistaferri, L., 2011. Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from danish tax records. The quarterly journal of economics 126 (2), 749–804.

Chiang, H. D., Hsu, Y.-C., Sasaki, Y., 2019. Robust uniform inference for quantile treatment effects in regression discontinuity designs. Journal of Econometrics.

Chiang, H. D., Sasaki, Y., 2019. Causal inference by quantile regression kink designs. Journal of Econometrics.

Dell, M., 2010. The persistent effects of peru's mining mita. Econometrica 78 (6), 1863–1903.
DellaVigna, S., 2018. Structural behavioral economics. NBER Working Papers 24797, National Bureau of Economic Research, Inc.
URL https://EconPapers.repec.org/RePEc:nbr:nberwo:24797
DellaVigna, S., Lindner, A., Reizer, B., Schmieder, J. F., 05 2017. Reference-Dependent Job Search: Evidence from Hungary*. The Quarterly Journal of Economics 132 (4), 1969–2018.
URL https://doi.org/10.1093/qje/qjx015
Dong, Y., 2016. Jump or kink? regression probability jump and kink design for treatment effect evaluation. Unpublished manuscript.
Dong, Y., Lee, Y.-Y., Gou, M., 2019. Regression discontinuity designs with a continuous treatment. Tech. rep., Research Institute of Economy, Trade and Industry (RIETI).
Dong, Y., Lewbel, A., 2015. Identifying the effect of changing the policy threshold in regression discontinuity models. Review of Economics and Statistics 97 (5), 1081–1092.
Englmaier, F., Schmöller, A., Stowasser, T., 2018. Price discontinuities in an online market for used cars. Management Science 64 (6), 2754–2766.
URL https://doi.org/10.1287/mnsc.2016.2714
Feir, D., Lemieux, T., Marmer, V., 2016. Weak identification in fuzzy regression discontinuity designs. Journal of Business & Economic Statistics 34 (2), 185–196.
Frandsen, B. R., Frölich, M., Melly, B., 2012. Quantile treatment effects in the regression discontinuity design. Journal of Econometrics 168 (2), 382–395.
Frölich, M., Huber, M., 2018. Including covariates in the regression discontinuity design. Journal of Business & Economic Statistics, 1–13.
Frölich, M., Sperlich, S., 2019. Impact evaluation. Cambridge University Press.
Fu, C., Gregory, J., 2019. Estimation of an equilibrium model with externalities: Post-disaster neighborhood rebuilding. Econometrica 87 (2), 387–421.
Ganong, P., Jäger, S., 2018. A permutation test for the regression kink design. Journal of the American Statistical Association 113 (522), 494–504.
Gelman, A., Imbens, G., 2018. Why high-order polynomials should not be used in regression discontinuity designs. Journal of Business & Economic Statistics, 1–10.
Gerard, F., Rokkanen, M., Rothe, C., forthcoming. Bounds on treatment effects in regression discontinuity designs with a manipulated running variable. Quantitative Economics.
Hahn, J., Todd, P., van der Klaauw, W., 2001. Identification and estimation of treatment effects with a regression-discontinuity design. Econometrica 69, 201–209.
Hansen, B. E., 2017. Regression kink with an unknown threshold. Journal of Business & Economic Statistics 35 (2), 228–240.
Imbens, G., Kalyanaraman, K., 2012. Optimal bandwidth choice for the regression discontinuity estimator. The Review of economic studies 79 (3), 933–959.
Imbens, G., Wager, S., 2019. Optimized regression discontinuity designs. Review of Economics and Statistics 101 (2), 264–278.
Imbens, G., Zajonc, T., 2011. Regression discontinuity design with multiple forcing variables.
Imbens, G. W., Angrist, J., 1994. Identification and estimation of local average treatment effects. Econometrica 62, 467–475.
Imbens, G. W., Lemieux, T., 2008. Regression discontinuity designs: A guide to practice. Journal of econometrics 142 (2), 615–635.
Jacob, B. A., Lefgren, L., 2004. Remedial education and student achievement: A regression-discontinuity analysis. Review of economics and statistics 86 (1), 226–244.
Kamat, V., 2018. On nonparametric inference in the regression discontinuity design. Econometric Theory 34 (3), 694–703.
Keele, L. J., Titiunik, R., 2015. Geographic boundaries as regression discontinuities. Political Analysis 23 (1), 127–155.

Kleven, H. J., 2016. Bunching. Annual Review of Economics 8, 435–464.

Kolesár, M., 2020. RDHonest: Honest inference in regression discontinuity designs. R package version 0.3.2.
URL https://github.com/kolesarm/RDHonest

Kolesár, M., Rothe, C., 2018. Inference in regression discontinuity designs with a discrete running variable. American Economic Review 108 (8), 2277–2304.

Lacetera, N., Pope, D. G., Sydnor, J. R., 2012. Heuristic thinking and limited attention in the car market. American Economic Review 102 (5), 2206–36.

Lalive, R., 2008. How do extended benefits affect unemployment duration? a regression discontinuity approach. Journal of econometrics 142 (2), 785–806.

Lalive, R., Schlosser, A., Steinhauer, A., Zweimüller, J., 2014. Parental leave and mothers' careers: The relative importance of job protection and cash benefits. Review of Economic Studies 81 (1), 219–265.
URL https://EconPapers.repec.org/RePEc:oup:restud:v:81:y:2014:i:1:p:219-265

Lee, D. S., 2008. Randomized experiments from non-random selection in us house elections. Journal of Econometrics 142 (2), 675–697.

Lee, D. S., Card, D., 2008. Regression discontinuity inference with specification error. Journal of Econometrics 142 (2), 655–674.

Lee, D. S., Lemieux, T., 2010. Regression discontinuity designs in economics. Journal of economic literature 48 (2), 281–355.

McCrary, J., 2008. Manipulation of the running variable in the regression discontinuity design: A density test. Journal of econometrics 142 (2), 698–714.

Neyman, J., 1923. On the application of probability theory to agricultural experiments. essay on principles. Statistical Science Reprint, 5, 463–480.

Noack, C., Rothe, C., 2020. Bias-aware inference in fuzzy regression discontinuity designs.

Otsu, T., Xu, K.-L., Matsushita, Y., 2013. Estimation and inference of discontinuity in density. Journal of Business & Economic Statistics 31 (4), 507–524.

Porter, J., Yu, P., 2015. Regression discontinuity designs with unknown discontinuity points: Testing and estimation. Journal of Econometrics 189 (1), 132–147.

Qu, Z., Yoon, J., 2018. Uniform inference on quantile effects under sharp regression discontinuity designs. Journal of Business & Economic Statistics, 1–23.

Rubin, D. B., 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology 66, 688–701.

Saez, E., 2010. Do taxpayers bunch at kink points? American economic Journal: economic policy 2 (3), 180–212.

Schmidt, L. D., Zhu, Y., 2016. Quantile spacings: A simple method for the joint estimation of multiple quantiles without crossing.

Schmieder, J., von Wachter, T., Bender, S., 2012. The effects of extended unemployment insurance over the business cycle: Evidence from regression discontinuity estimates over 20 years. The Quarterly Journal of Economics 127 (2), 701–752.

Shen, S., Zhang, X., 2016. Distributional tests for regression discontinuity: Theory and empirical examples. Review of Economics and Statistics 98 (4), 685–700.

Strittmatter, A., Lechner, M., 2020. Sorting in the used-car market after the volkswagen emission scandal. Journal of Environmental Economics and Management 101, 102305.
URL http://www.sciencedirect.com/science/article/pii/S0095069620300280

Thistlethwaite, D., Campbell, D., 1960. Regression-discontinuity analysis: An alternative to the ex post facto experiment. Journal of Educational Psychology 51, 309–317.

Todd, P. E., Wolpin, K. I., 2020. The best of both worlds: Combining rcts with structural modeling. Tech. rep., University of Pennsylvania.

van der Klaauw, W., 2002. Estimating the effect of financial aid offers on college enrollment: A regression-discontinuity approach. International Economic Review 43, 1249–1287.

Wald, A., 1940. The fitting of straight lines if both variables are subject to error. Annals of Mathematical Statistics 11, 284–300.