The Analysis of the Regression-Discontinuity Design in R

Felix Thoemmes
Wang Liao
Ze Jin
Cornell University

This article describes the analysis of regression-discontinuity designs (RDDs) using the R packages `rdd`, `rdrobust`, *and* `rddtools`. *We discuss similarities and differences between these packages and provide directions on how to use them effectively. We use real data from the Carolina Abecedarian Project to show how an analysis of an RDD can be performed from start to finish.*

The regression-discontinuity design (RDD) was first considered by Thistle-thwaite and Campbell (1960). Despite some initial interest in the method, it never became a very popular design choice (Cook, 2008). There is, however, a current renaissance of the RDD, fueled in part by important theoretical contributions from economics (Angrist & Pischke, 2009; Imbens & Kalyanaraman, 2012). In this literature, the RDD is often considered to be one of the strongest nonrandomized designs with regard to drawing causal conclusions from data. As a result of these advances, there is now also an increased interest in performing RDDs among applied researchers, especially in economics and education (Louie, Rhoads, & Mark, 2016; Melguizo, Bos, Ngo, Mills, & Prather, 2016; Porter, Reardon, Unlu, Bloom, & Cimpian, 2016; Zhang, Hu, Sun, & Pu, 2016).

Introductions to the underlying logic and analysis of the RDD are numerous (Imbens & Lemieux, 2008; Lee & Lemieux, 2010; Schochet et al., 2010; Trochim, 1984); therefore, we will keep our theoretical presentation very short. The RDD is a design that facilitates causal identification of an effect T, on an outcome Y, in the presence of confounding due to unobserved variables. The key feature of RDDs is an assignment variable, X, that (often uniquely) defines treatment assignment T. For example, the treatment could be enrollment in a remedial math class, and the assignment variable is a score on a standardized math test. Here, school administrators assign students to the math class, if and only if, the math score of a particular student is below a certain threshold.

If the assignment variable X deterministically causes the treatment, we refer to this as a sharp RDD. If the relationship between X and T is only probabilistic
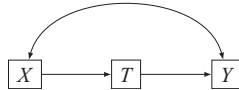
1

FIGURE 1. *Graphical causal model of a regression-discontinuity design.*

(meaning that the probability of receiving the treatment does not switch from 0 to 1 at the cutoff of X but to some other values ranging between 0 and 1, e.g., .1 or .9), we refer to this as a fuzzy RDD. The relationship between X and Y is allowed (and expected) to be confounded by unobserved variables. However, even in the presence of unobserved confounding, it is possible to estimate an unbiased causal effect, if the data are analyzed properly using RDD methods. The exact reasons why the RDD can yield unbiased effects have been spelled out by Shadish, Cook, and Campbell (2002) using the language of the generalized causal inference framework by Campbell, or by Imbens and Lemieux (2008) using the language of potential outcomes. We provide here a simple explanation using graphical causal models (Pearl, 2009).

Consider the graphical model in Figure 1, which consists of variables (nodes in the graph) and paths (arrows in the graph). Directed paths indicate causal relationships, and bidirected paths indicate confounding relationships due to unobserved variables. Consider now that we are interested in the causal effect of T on Y. In the graph in Figure 1, the relationship between T and Y is confounded, due to the presence of X and the (countless) unobserved variables that induce an association between X and Y (the bidirected path). However, the graph also informs us that every unobserved variable influences T only through X and that X alone determines T. In some sense, X is a ''bottleneck'' for all potentially confounding influences between T and Y. As such, conditioning on X is sufficient to deconfound the relationship between T and Y and thus obtain an unbiased causal effect.

Another way to understand why the RDD yields results similar to those of a randomized experiment is to consider the fact that individuals who are right below the cutoff of the assignment variable and right above the cutoff of the assignment variable are expected to be very similar to each other (we may also say that they are exchangeable). They only differ on treatment assignment. Consider the hypothetical data shown in Figure 2, in which every individual who scored lower than 0 on the assignment variable X is assigned to the treatment condition, whereas everyone who scored equal or higher than 0 is assigned to the control condition. The individuals who are closely around this cutoff are indeed comparable with respect to X and presumably other variables as well. They only differ with respect to the treatment assignment. As a result, the data in this area that is close to the cutoff resemble a randomized experiment, which allows the identification of a causal treatment effect.
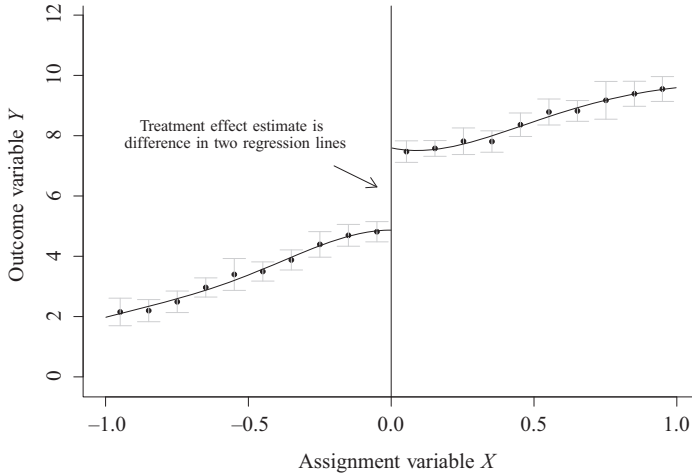
FIGURE 2. *Example plot of data in a regression-discontinuity design, created using the* `rdrobust` *package. The graph shows binned means, with 95% confidence intervals, and an overlaid smoother.*

Conceptual Overview of an RDD Analysis

The statistical analysis of RDDs is described in much detail elsewhere (Imbens & Lemieux, 2008; Lee & Lemieux, 2010; Schochet et al., 2010; Trochim, 1984), and it is not the focus of this article to formally describe these analyses. Nevertheless, it is helpful to at least give a brief overview of the involved analyses, before considering the exact implementation of them in the R packages that are being reviewed here.

We consider a scenario, in which a researcher has identified a variable that acts as an assignment variable in an RDD that allows identification of a causal effect of interest. As a concrete example, a researcher may have identified that a local agency administers an enrichment program at schools, but only to students whose grades are below a certain threshold, for example, a grade point average (GPA) below 2.5. The researcher wants to know whether providing the enrichment has any effect on later learning, and therefore collects data on students' GPAs, later learning outcomes, and whether they have attended the enrichment program.

Assumption Checks

In a first step, the researcher would have to confirm that the design assumptions of the RDD were not violated. In particular, this means confirming that the treatment assignment mechanism behaved as assumed. For example, there may be concerns that students with slightly higher GPAs were also allowed to participate in the enrichment program, by changing their records slightly. Or maybe

3

parents could petition that their child would be allowed to participate even if the child's GPA made it ineligible. Those violations could be visible in discontinuities (essentially "bumps") in the distribution of the assignment variable, here GPA. These discontinuities can be visually inspected or formally tested with a hypothesis test (McCrary, 2008).

Second, one would assume that the treatment effect only occurs at the cutoff of GPA 2.5 and not at other cutoffs. If we would in fact observe treatment effects at other cutoffs, we would feel less confident that the presumed treatment effect is really due to the treatment that was administered differentially at the cutoff. A formal way to explore this is to estimate treatment effects at various other cutoffs and compare them with the effect at the presumed cutoff. This procedure is sometimes referred to as "placebo tests."

Third, the treatment is believed to have an impact on an outcome variable but not on nonoutcome covariates, especially those collected prior to treatment administration. In fact, if we would observe a treatment effect on a pre-treatment covariate, much doubt would be cast on the validity of the RDD. To formally explore this, researchers are encouraged to replace the actual outcome with each of the covariates and redo the RDD analysis. A desired result would be that no effect at the cutoff is found in any of the pretreatment covariates.

Estimation

Assuming that all assumption checks were successful, a researcher may now estimate the treatment effect. The treatment effect of an RDD is quantified in the difference between regression lines right at the cutoff. In the example at hand, the researcher would have to estimate the jump in the regression line relating GPA to later academic outcomes right at the cutoff of a GPA 2.5, at which the treatment assignment switched. This effect estimation is achieved either through parametric regression models (also known as the global approach) or through semi- or nonparametric methods, which often only consider points close to the cutoff through weighting and employing local linear or local polynomial regression (also known as the local approach). If researchers choose to employ the global, parametric approach, it is often suggested that higher order polynomials should be fitted, although there are opposing viewpoints (Gelman & Imbens, 2014). The parametric models are parameterized in such a fashion that one of the resulting coefficients expresses the discontinuity in the regression line, usually achieved through centering the assignment variable and forming interaction terms. If the local, nonparametric approach is chosen, only data points closely around the cutoff are chosen and a local linear (or local polynomial) regression model is fitted on either side of the cutoff. Usually, these models are fitted using weighted regression, giving higher weights to individuals who are closer to the cutoff. A challenging question is how to choose the bandwidth that determines these weights. The most popular choice is a data-driven bandwidth selection algorithm

4

first suggested by Imbens and Kalyanaraman (2009). In Imbens and Kalyanaraman (2012), this selection algorithm is further modified. More recent developments have added additional choices for the bandwidth selection (Calonico, Cattaneo, & Farrell, 2016; Calonico, Cattaneo, & Titiunik, 2015b). Simulation studies (Calonico, Cattaneo, & Titiunik, 2014) suggest that these novel choices have good frequentist coverage properties. As explained in much detail in Calonico, Cattaneo, and Titiunik (2015b), these new choices for bandwidth selection attempt to correct for bias due to undersmoothing (bias that emerges because the functional form of the regression lines at the cutoff is not well approximated) and correct standard errors due to uncertainty in the bias correction.

In the preceding section, we did not differentiate treatment effect estimation for sharp and fuzzy RDDs. While there are differences in the actual analysis, the general logic of measuring a difference in regression lines at the cutoff applies to both. In the case of a fuzzy RDD, an additional step is performed such that the actual observed treatment assignment is regressed on the assignment variable conditioning on the cutoff. Then, an RDD is performed on the predicted treatment assignment, as opposed to the actual observed treatment assignment. This so-called two-stage least squares procedure is identical to the instrumental variables estimator.

Sensitivity Checks

The choice of the parametric model in the global approach or the choice of bandwidth for the local approach is critical, and different choices will yield different estimates of the treatment effect. Because of this model dependency on these choices, it is often suggested to perform some sensitivity checks and explore other modeling options and in doing so bound the treatment effect. Some packages provide these checks automatically, but in theory, the user could always perform the checks manually by simply changing the particular parametric model, or the specific bandwidth, and then estimate the effect again.

To conclude, these three conceptual steps—assumption checks, estimation, and sensitivity checks—complete the analysis of an RDD. We now turn to the implementation of this analysis in R.

R **Packages**

We first provide an overview of all available R packages that can analyze RDDs and specifically focus on the three analytic steps for an RDD outlined above. We describe and compare their features, followed by an applied example.

rdd

The rdd package (Dimmery, 2016) was the first published package to perform a full-featured analysis of an RDD. The package was last updated on March 2016, and appears to be still under active development. The current version of

5

rdd has all essential features (summarized and compared to other packages in Table 1) to conduct an RDD. The rdd package relies on a set of existing R packages to perform several tasks. It uses the lmtest (Zeileis & Hothorn, 2002) package to perform inference, the AER (Kleiber & Zeileis, 2008) package for two-stage least squares regression for the fuzzy RDD, the sandwich (Zeileis, 2004) package for robust standard errors, and the Formula (Zeileis & Croissant, 2010) package for general handling of model specifications.

Assumption checks. The rdd package performs the McCrary test (McCrary, 2008) to assess potential discontinuities at the cutoff of the assignment variable. By default, it produces a *p* value and an associated plot. The rdd package does not have built-in functions to perform placebo tests or tests on non-outcome covariates. However, these can be performed by manually changing the cutoff or manually replacing the outcome variable with a non-outcome covariate.

Estimation. The rdd package allows the estimation of a treatment effect using the local, nonparametric approach. By default, it uses the Imbens–Kalyanaraman (Imbens & Kalyanaraman, 2009) bandwidth selection (from hereon we refer to this simply as IK) to determine the weights of the local linear regression at the cutoff. The way that the effect is estimated in the rdd package is by running a weighted linear regression, with weights derived from the IK bandwidth selection. The so-called kernel (a weighting function that follows a particular distribution) for the local linear regression is by default the triangular kernel (a choice recommended by Imbens & Kalyanaraman, 2012), but it can be changed to a variety of other choices. The package also allows estimation of a treatment effect for a fuzzy RDD and allows the same choices for bandwidth and kernel as described above. The package also allows the inclusion of potential covariates. Usually, it is not necessary to include covariates in a sharp RDD, but doing so can potentially yield smaller standard errors. In the rdd package, covariates are entered (as linear terms) in the regression equation of a sharp RDD or are entered (again as linear terms) in both regressions that are necessary to estimate a fuzzy RDD. Lastly, the package can compute robust standard errors using the sandwich package in R. The rdd package provides a scatterplot of the assignment and the outcome variable (binned means, not actual data points), along with a smooth approximation of the relationship between these two variables.

Sensitivity checks. The rdd package reports by default, and without any user input, treatment effects using three local linear regressions with a computed bandwidth per the IK selection, a double of the preceding bandwidth, and a half of the preceding bandwidth, respectively.

6

TABLE 1.
Differences Between RDD Packages

Package	`rdd`	`rdrobust`	`rddtools`
Package dependencies	(1) `sandwich` (2) `lmtest` (3) `AER` (4) `Formula`	N/A	(1) `sandwich` (2) `lmtest` (3) `AER` (4) `Formula` (5) `np`
RDD designs	(1) Sharp (2) Fuzzy	(1) Sharp (2) Fuzzy	(1) Sharp (2) Fuzzy
Coefficient estimator	Local linear regression	Local polynomial regression	Local polynomial regression
Bandwidth selectors	Imbens and Kalyanaraman (2009)	(1) Calonico, Cattaneo, and Titiunik (2014) (2) Imbens and Kalyanaraman (2012) (3) Ludwig and Miller (2007) (4) Calonico, Cattaneo, and Farrell (2016) (5) Calonico, Cattaneo, Farrell, and Titiunik (2016)	(1) Imbens and Kalyanaraman (2012) (2) Ruppert, Sheather, and Wand (1995)
Kernel functions	(1) Triangular (2) Rectangular (3) Epanechnikov (4) Gaussian	(1) Triangular (2) Rectangular (3) Epanechnikov	(1) Triangular (2) Rectangular (3) Gaussian
Sharp RDD estimate	`lm {stats}`	Matrix computation	`lm {stats}`
Fuzzy RDD estimate	`ivreg {AER}`	Matrix computation	`ivreg {AER}`
Bias correction	N/A	Local polynomial regression	N/A
Covariate options	Include	Include	(1) Include (2) Residual
Standard error estimate	`vcovHC {sandwich}`	Matrix computation	`vcov {stats}`
Clustered error estimate	`sandwich {sandwich}`	Matrix computation	`sandwich {sandwich}`
Assumption testing	McCrary sorting	N/A	(1) McCrary sorting (2) Equality of covariates distribution (3) Equality of covariates mean

Note. N/A = not applicable; RDD = regression-discontinuity design.

7

rddtools

The `rddtools` package (Stigler & Quast, 2015) was first released in 2013 and has been continuously updated to the current version, last updated on July 2015. The package shares the same dependencies as the `rdd` package but also depends on other packages, such as `np` (Hayfield & Racine, 2008), and uses the test by McCrary (2008) as implemented in the `rdd` package.

Assumption checks. The `rddtools` package has extensive capabilities for assumption checks. It offers the McCrary test (through calling the `rdd` package) and automatic placebo tests. For the latter, the package reestimates the treatment effect at user-specified cutoffs and presents a convenient graph that shows treatment effects as a function of the cutoffs. The package does not offer an automatic option to estimate treatment effects of non-outcome covariates, but this could be done manually by changing the outcome variable and reestimating the model.

Estimation. The `rddtools` package offers nonparametric estimation for both sharp and fuzzy RDDs using the newer IK bandwidth selection algorithm (Imbens & Kalyanaraman, 2012). In addition, it offers an alternative algorithm based on Ruppert, Sheather, and Wand (1995). Besides the local estimation approach (where both local linear and local polynomial models can be estimated), the package can also perform global estimation (using parametric models). The offered parametric models are quite flexible and include not only the simple interactive model but also the higher order polynomials or other parametric models (e.g., probit models, if the outcome is binary). As in `rdd`, `rddtools` allows the inclusion of covariates and the estimation of robust standard errors. It is also possible to request a basic plot of the RDD that consists of binned means of the outcome variable along values of the assignment variable with an overlaid parametric function.

Sensitivity checks. The `rddtools` package offers a comprehensive sensitivity check and automatically reestimates the nonparametric, local treatment effect based on different bandwidths. The result of these sensitivity tests is returned as a graph that shows the treatment effect as a function of the chosen bandwidth. Seeing similar treatment effects across a wide range of bandwidths in the plot would bolster faith in the treatment effect.

rdrobust

The `rdrobust` package (Calonico, Cattaneo, Farrell, & Titiunik, 2016) is the latest addition among published packages but arguably the most comprehensive one. The latest version of this package was released in August 2016. Unlike the packages discussed so far, `rdrobust` does not depend on other packages.

Assumption checks. The `rdrobust` package itself does not offer any assumption checks. However, the authors of the package are currently in the process of

8

releasing additional `R` packages and provide downloads of `R` functions (not packages) on their website https://sites.google.com/site/rdpackages/home. The `rddensity` function (Cattaneo, Jansson, & Ma, 2015) performs advanced tests for discontinuities in the assignment variable that go beyond McCrary (2008). Because they are not part of the `R` package, we do not consider them in detail and simply provide a reference for the reader.

Estimation. Just like all previous packages, `rdrobust` allows for the nonparametric estimation of both sharp and fuzzy RDDs. It offers an extremely comprehensive array of bandwidth selections—a total of over 10 different selection algorithms (Calonico et al., 2016; Calonico et al., 2014; Ludwig & Miller, 2007). It also offers bandwidth selection based on cross-validation. It allows the user to choose local linear or local polynomial regression for treatment effect estimation. It is the only package that adds a bias correction due to possible undersmoothing and adjusts standard errors based on this correction. Like other packages, it is capable of the inclusion of covariates and reporting robust standard errors. The package also provides well-formatted plots of an RDD with binned means and overlaid regression smoothers (Calonico, Cattaneo, & Titiunik, 2015a) and confidence intervals for each binned mean.

Sensitivity checks. The `rdrobust` package has no built-in feature for either placebo tests or bandwidth sensitivity tests, although these could be performed manually.

Summary of `R` Packages

All three packages are equally easy to use and do not require specialized knowledge of `R`. Likewise, all three packages are capable of performing the analysis of both sharp and fuzzy RDDs with nonparametric (local) estimation methods. When it comes to the critical choice of bandwidth selection, the `rdrobust` package has the most extensive options. It is the only package that offers a variety of novel bandwidth selection algorithms coupled with robust standard errors and confidence intervals that are reported to have superior coverage. It also provides the most comprehensive plotting options and allows flexible tweaking of these plots. A strength of the `rddtools` package is that it is the only package that automates certain assumption and sensitivity checks. Only `rddtools` provides plots of treatment effects under different bandwidth choices, and it automates the otherwise tedious creation of placebo tests with different cutoffs. In summary, all three packages can be recommended. The `rdd` package provides a basic but sound analysis, `rddtools` excels in the domain of assumption and sensitivity checks, and `rdrobust` is the most advanced package when it comes to nonparametric bandwidth selection and treatment effect estimation.

Finally, for researchers who are less familiar with using packages in `R`, `rddapp` is an `R` Shiny interface to facilitate the analysis of RDDs. The goal

of `rddapp` is to provide researchers with an easy-to-use graphical interface that allows similar analyses as the ones presented here without the need for coding.[1]

Applied Example

For our applied example, we rely on the published data from the Carolina Abecedarian Project and the Carolina Approach to Responsive Education (Ramey, Gallagher, Campbell, Wasik, & Sparling, 2004), which can be accessed online (http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/4091). In this randomized controlled trial, young children were assigned to either a control group or to some early childhood intervention, which started at 6 weeks of age and lasted until the third year of elementary school. Children were followed longitudinally for multiple years and were measured on a variety of cognitive measures and academic achievements. The outcome measure that we chose was the Stanford–Binet IQ score at age 2 that was assessed after almost 2 years of treatment. The data set contained 103 children in the control condition and 73 children in the treatment condition. Due to 18 children with missing data on the outcome, only 158 children in total remained. To form a baseline of comparison, we first estimated a treatment effect based on the randomized controlled trial. The mean difference between the two groups was 9.88 IQ points, a highly significant difference, $t(156) = 5.46$, $p = 1.64 \times 10^{-7}$, and presumably a very important practical impact.

We then engaged in the following thought experiment: What if the authors of the original study would not have randomly assigned children to conditions, but based on a cutoff on a pretreatment variable? We pretended that treatment was assigned based on such a cutoff. We assumed that treatment would only be administered to mothers whose IQ was below the median of the sample (which in this data set happened to be 85). We took a subset of the data and only retained the treated children whose mothers had an IQ below 85 and the untreated children whose mothers had an IQ of 85 or more. In doing so, we created an RDD out of the randomized controlled trial. This procedure allows us to benchmark our results against the results of the randomized controlled trial. Figure 3a shows a scatterplot of the relationship between mother's IQ and child's IQ at 2 years with overlaid locally weighted scatterplot smoothing (LOWESS) smoothers for both treated and untreated children in the original data set. Figure 3b shows the same relationship for the reduced data that mimic an RDD.

A naive (and biased) estimate of the treatment effect on this subset would be to simply compute the mean difference. Here this turned out to be 4.53 IQ points, a nonsignificant result, $t(79) = 1.82$, $p = .072$. We now conduct the RDD analysis using the R packages, with the expectation that such an analysis would recover the treatment effect from the randomized controlled trial. Our complete R code to replicate all results is given in the Appendix, available in the online version of the journal.
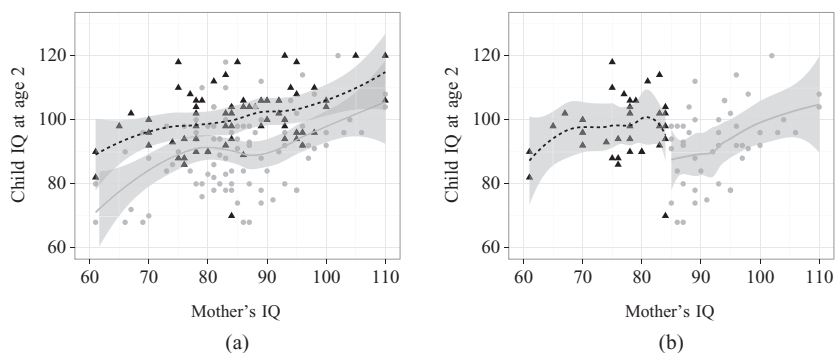
10

FIGURE 3. *Plot of the relationship between assignment variable (mother's IQ) and outcome (child's IQ) for (a) the full data and (b) the reduced data. Treated children are shown as black triangles, untreated children as gray dots. Two overlaid smoothers with 95% confidence interval and corresponding colors are shown.*

Assumptions

We first used the `rdd` package to perform the test by McCrary (2008) to check for any discontinuities in the assignment variable. This discontinuity test was not significant, $z = 1.16$, $p = .244$, indicating no violation of this assumption. A plot is shown in Figure 4. All other packages rely on the same function and therefore we do not report duplicate results.

A second assumption check is that the treatment effect only occurs at the cutoff. We used the `rddtools` package to perform placebo tests. We reestimated the treatment effects (using the local approach) for different cutoffs. Ideally, we would hope to see that all other effects are near zero. Figure 5 displays the effect at the cutoff in blue, effects at lower cutoffs in red, and effects at higher cutoffs in green. All effects are bounded by a confidence interval.

In this instance, it appears that some cutoffs also yield similarly sized effects, some even with reversed sign. However, most of the "placebo" cutoffs yield a confidence interval that substantially covers zero, in comparison to the marginal coverage at the actual cutoff. Therefore, the plot suggests some evidence against potential violations in treatment assignment. In real data sets, we might feel inclined to investigate why we might be seeing at least one large effect in the opposite direction at different cutoffs. No other package reports placebo tests.

Finally, we performed a single test of a non-outcome covariate. We chose the Apgar score (a score that identifies how healthy a newborn is). Because it is measured at birth, it is clearly a covariate that was assessed *prior* to the treatment. An RDD analysis of this non-outcome covariate (using the local approach with default IK bandwidth) found no effect at the cutoff, $z(29) = -.08$, $p = .865$. Other non-outcomes could be similarly analyzed to even further increase our confidence in the observed treatment effect.
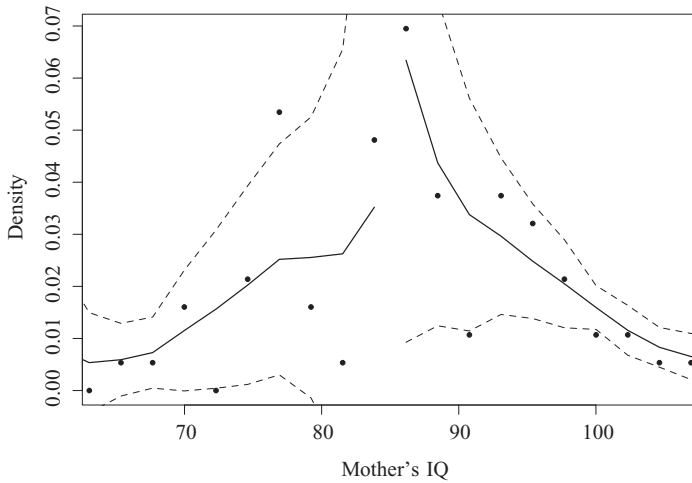
11

FIGURE 4. *Plot of the density of the assignment variable (mother's IQ) created using the* rdd *package. Individual data points show binned means, with 95% confidence intervals, and an overlaid smoother.*
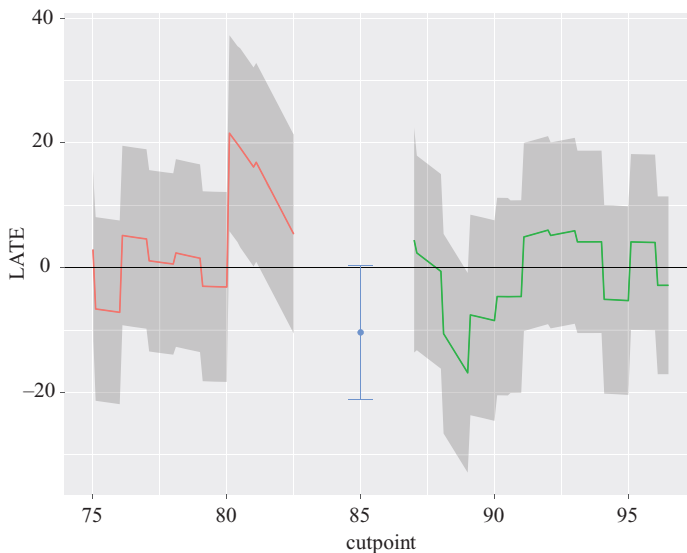

FIGURE 5. *Placebo tests as performed by* rddtools. *The plot is generated using the default settings. LATE = local average treatment effect.*
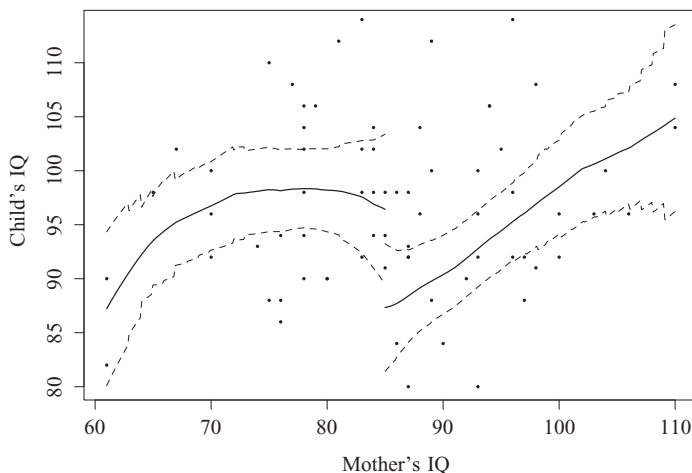
12

FIGURE 6. *Regression-discontinuity design plot generated in* `rdd`*. The plot was generated using the default options, but axes labels were added manually.*

Estimation

We then proceeded to estimate the treatment effect on IQ at 2 years of age. We first plotted our RDD, using both the `rdd` package and the `rdrobust` package. The plot of the `rddtools` is similar to the one from `rdd` and therefore omitted. The plots of the packages are shown in Figures 6 and 7.

Both plots show the assignment variable (mother's IQ) on the horizontal axis and the outcome variable (child's IQ) on the vertical axis. Both graphs show binned data (bins are formed on the assignment variable, and binned means on the outcome are plotted); however, the binning width is based on different defaults.[2] The plot from the `rdd` package displays a smoother (separately estimated for each side of the cutoff), along with a confidence interval around this smoother. In contrast, the plot from the `rdrobust` package displays confidence intervals around individual binned means and by default also draws a horizontal line at the cutoff. Both plots are useful to visually explore the discontinuity of the regression lines at the cutoff, and they can also be helpful to visualize the functional form between the assignment and outcome variable.

We then estimated the treatment effect using a wide variety of choices within each of the packages to demonstrate a comprehensive use of them. We have labeled our models M1–M8 and organized our results in Table 2. The table also includes detailed information about bandwidth choices, kernels, types of regression model, sample sizes, and all inferential statistics of the treatment effect. The first model (M1) was estimated using `rdd`. By default, `rdd` will return three different estimates, one using the IK bandwidth from 2009, and then half and double of that bandwidth, which serves as a sensitivity check. As seen in Table 2,
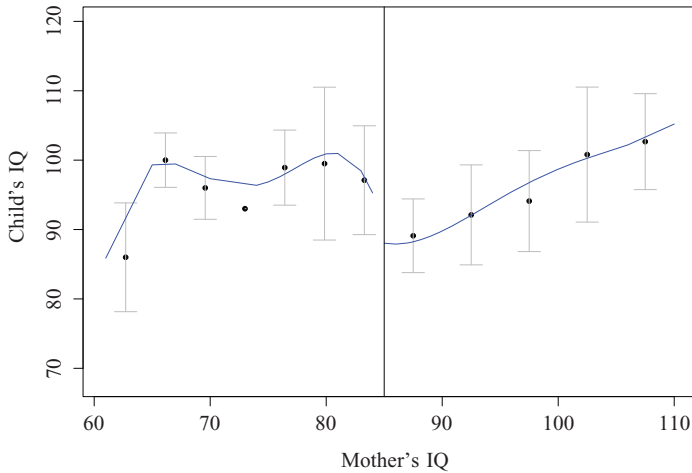
13

FIGURE 7. *Regression-discontinuity design plot generated in* rdrobust.

the point estimate ranged from 5.9 to 10.5 IQ points, but not all estimates were statistically significant. As a note, the point estimates reported in Table 2 are all negative (they are reported like this by the packages). At first sight this seems to indicate that the treatment suddenly has a negative impact on IQ. However, this is based on the fact that the packages simply compare the difference in regression lines at the cutoff, taking the left-hand side of the cutoff as the baseline by default. In our case, individuals on the left-hand side of the cutoff received the treatment, and thus were higher, and hence we observed a *drop* at the cutoff. This result is absolutely congruent with the effects observed in the randomized controlled trial.

Models M2 through M4 were all estimated using the rdrobust package. Recall that the rdrobust package uses a different standard error estimation that is corrected for undersmoothing bias. The models that we estimated are the following: one that used the default of the rdrobust package (M2), one in which we changed the bandwidth to the popular IK method from 2012 (M3), and one in which we used the exact same bandwidth as the one used by the rdd package. Point estimates were similar, ranging from 5.5 to 9.2 IQ points, but standard errors were generally much larger (both with and without bias correction), and all effects failed to reach significance, despite being comparable in magnitude.

Models M5 through M8 were estimated using the rddtools package. Model M5 uses the defaults when requesting the local, nonparametric estimation. Model M6 is the default global, parametric model. Model M7 also uses the global, parametric model, but with a sample size restriction based on the IK bandwidth selection. Note that this model does not use local linear regression

14

TABLE 2.
Estimated Treatment Effects Using Various R *Packages*

| Package (Version Number) | Setting | Bandwidth Selection | | | Estimation | | | | | |
		Procedure	Bandwidth	Kernel	Estimator	N	Estimate	SE	Test	p
M1 rdd (0.57)	Default[a]	IK (2009)	11.095	Triangular	local LR	62	−9.085	5.902	z	.124
		IK (2009)	5.547 (half)	Triangular	local LR	33	−5.014	8.838	z	.571
		IK (2009)	22.189 (double)	Triangular	local LR	77	−10.532	4.827	z	.029
M2 rdrobust (0.93)	Default	CCFT (2016)	5.682/9.264[b]	Triangular	local LR	81	−5.489	8.921	z	.538/.824[b]
M3 rdrobust (0.93)	Custom	IK (2012)	11.449/13.310[b]	Triangular	local LR	81	−9.218	6.064	z	.129/.460[b]
M4 rdrobust (0.93)	Custom	(set as M1)	11.095/11.095[b]	Triangular	local LR	81	−9.085	6.125	z	.138/.569[b]
M5 rddtools (0.4.0)	Default	IK (2012)	19.467	Triangular	local LR	75	−10.371	5.498	z	.059
M6 rddtools (0.4.0)	Default	—	—	—	LR	81	−12.611	3.807	t	.001
M7 rddtools (0.4.0)	Custom	IK (2012)	19.467	—	LR	75	−11.116	4.347	t	.013
M8 rddtools (0.4.0)	Custom	(set as M1)	11.095	—	LR	62	−10.544	5.283	t	.051

Note. RDD assignment is based on mother's IQ, cut at median (85) with lower scores assigned as treated (i.e., a sharp design). LR = linear regression; IK = Imbens and Kalyanaraman; CCFT = Calonico, Cattaneo, Farrell, and Titiunik; RDD = regression-discontinuity design.
[a]Defaults to three different estimates. [b]Bias bandwidth and robust *p* value.

15

with weights, but simply truncates the sample, and then fits a parametric model. Finally, Model M8 mimics the default choices of the `rdd` package and is used for comparison. Across the four estimates by the `rddtools` package, point estimates varied between 10.3 and 12.6 IQ points, with most of the results being significant.

In summary, we can see that treatment effects vary to some extent based on the bandwidth selection. Since different bandwidth selections imply different units that are being considered, it is not surprising that we observed some variability in the estimates. Likewise, we saw that standard errors were generally larger than in the case of the randomized controlled trial (which had a larger sample size). As expected, using narrower bandwidths yielded larger standard errors, and standard errors that were corrected for undersmoothing bias were even larger. At the same time, the general direction and magnitude of the point estimates was consistent with the randomized controlled trial, and generally better than the naive estimate.

Sensitivity checks

Finally, we used the `rddtools` package to quickly perform a more comprehensive assessment of the RDD's sensitivity to the selection of the bandwidth. The package reestimated the treatment effect across various bandwidths and plots results, as shown in Figure 8.

Ideally, we would like to see that the treatment effect remains stable across different choices of bandwidth. As we can see in this plot, the effect remains relatively constant for most choices of bandwidth. With widening bandwidth (toward the right end of the plot), the treatment effect becomes significant (while being of similar magnitude), which is partly due to the increased sample size and the resulting smaller standard error. For very small bandwidths, the estimate becomes highly unstable, which is expected.

Discussion

All packages that we have reported here can be used to analyze an RDD. Many features of the packages are shared and in fact rely on similar underlying dependent R packages. The `rdrobust` package stands out in this regard, as all of its routines are completely independent of other packages. It also features the most advanced and most comprehensive selection of bandwidth selection choices to estimate the treatment effect using the local, nonparametric approach. The `rddtools` package has the most advanced capabilities to quickly and efficiently perform assumption checks and sensitivity checks. It is the only package that can automatically produce plots of placebo tests and plots of changing bandwidth selections.

There are, however, features that are currently missing from all of the packages. First, it is impossible to estimate treatment effects from RDDs with
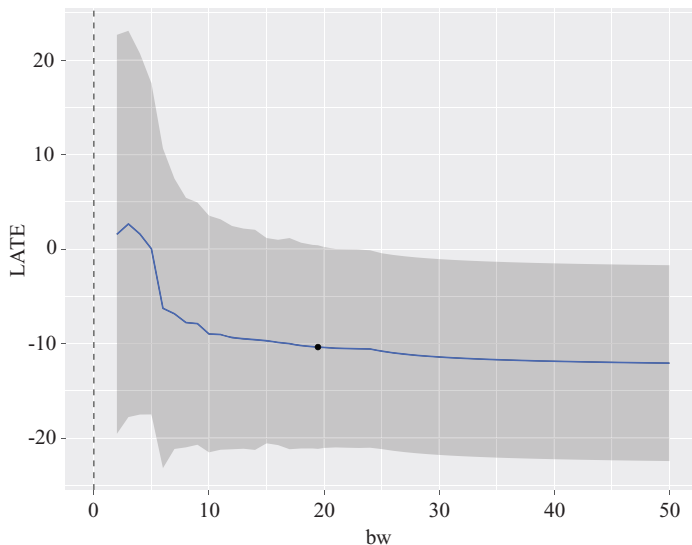
16

FIGURE 8. *Sensitivity analysis as performed by* `rddtools`. *LATE = local average treatment effect; bw = bandwidth.*

multiple assignment variables. Wong, Steiner, and Cook (2013) described situations in which assignment to treatment is not based on a single variable but two (e.g., a remedial classes offered to students who fall below on academic threshold on at least one of the two possible criteria). Wong et al. identify at least four different ways to estimate such effects. Another missing feature is the estimation of statistical power for an RDD. Power considerations are important in RDDs, because statistical power tends to be much lower than in randomized controlled trials (Cappelleri, Darlington, & Trochim, 1994; Goldberger, 1972; Schochet, 2009). Cattaneo, Titiunik, and Vazquez-Bare (2016) have Stata functions and are currently working on putting together an R package to estimate the power for RDDs.

In summary, there are currently several great packages in R that will perform the vast majority of analyses needed for an RDD, and applied researchers who wish to perform such an analysis have a great toolbox at their disposal.

Acknowledgments

Authors' Note

The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

17

Declaration of Conflicting Interests

Funding

Notes

1. The beta version of `rddapp` can be accessed here: https://rddapp.shinyapp s.io/shinyrdd/
2. It appears that the `rdd` package uses many bins by default, which results in plots that are more similar to scatterplots.

References

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.

Calonico, S., Cattaneo, M. D., & Farrell, M. H. (2016). On the effect of bias estimation on coverage accuracy in nonparametric inference. *arXiv preprint arXiv:1508.02973*.

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2016). rdrobust: Robust data-driven statistical inference in regression-discontinuity designs [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=rdrobust (R package version 0.93).

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

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2015a). Optimal data-driven regression discontinuity plots. *Journal of the American Statistical Association*, *110*, 1753–1769.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2015b). rdrobust: An r package for robust nonparametric inference in regression-discontinuity designs. *R Journal*, *7*, 38–51.

Cappelleri, J. C., Darlington, R. B., & Trochim, W. M. (1994). Power analysis of cutoff-based randomized clinical trials. *Evaluation Review*, *18*, 141–152.

Cattaneo, M. D., Jansson, M., & Ma, X. (2015). *rddensity: Manipulation testing based on density discontinuity* (Tech. Rep. Working paper). University of Michigan. Retrieved from http://www-personal.umich.edu/∼cattaneo/papers/Cattaneo-Jans son-Ma_2016_Stata.pdf

Cattaneo, M. D., Titiunik, R., & Vazquez-Bare, G. (2016). *Power calculations for regression discontinuity designs* (Tech. Rep.). Ann Arbor: University of Michigan.

Cook, T. D. (2008). "Waiting for life to arrive": A history of the regression-discontinuity design in psychology, statistics and economics. *Journal of Econometrics*, *142*, 636–654.

Dimmery, D. (2016). rdd: Regression discontinuity estimation [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=rdd (R package version 0.57).

18

Gelman, A., & Imbens, G. (2014, August). *Why high-order polynomials should not be used in regression discontinuity designs* (Working Paper No. 20405). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w20405doi:10.3386/w20405

Goldberger, A. S. (1972). *Selection bias in evaluating treatment effects: Some formal illustrations*. Madison: University of Wisconsin–Madison.

Hayfield, T., & Racine, J. S. (2008). Nonparametric econometrics: The np package. *Journal of Statistical Software*, *27*. Retrieved from http://www.jstatsoft.org/v27/i05

Imbens, G., & Kalyanaraman, K. (2009, February). *Optimal bandwidth choice for the regression discontinuity estimator* (Working Paper No. 14726). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w14726

Imbens, G., & Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*, *79*, 933–959. Retrieved from http://restud.oxfordjournals.org/content/79/3/933.abstract

Imbens, G., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, *142*, 615–635. doi:10.1016/j.jeconom.2007.05.001

Kleiber, C., & Zeileis, A. (2008). *Applied econometrics with R*. New York, NY: Springer-Verlag. Retrieved from http://CRAN.R-project.org/package=AER

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, *48*, 281–355. Retrieved from http://www.jstor.org/stable/20778728

Louie, J., Rhoads, C., & Mark, J. (2016). Challenges to using the regression discontinuity design in educational evaluations: Lessons from the transition to algebra study. *American Journal of Evaluation*, *37*, 381–407. Retrieved from http://aje.sagepub.eom/content/37/3/381.abstract

Ludwig, J., & Miller, D. L. (2007). Does head start improve children's life chances? Evidence from a regression discontinuity design. *The Quarterly Journal of Economics*, *122*, 119–157. doi:10.1162/qjec.122.1.159

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

Melguizo, T., Bos, J. M., Ngo, F., Mills, N., & Prather, G. (2016). Using a regression discontinuity design to estimate the impact of placement decisions in developmental math. *Research in Higher Education*, *57*, 123–151. doi:10.1007/s11162-015-9382-y

Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). New York, NY: Cambridge University Press.

Porter, K. E., Reardon, S. F., Unlu, F., Bloom, H. S., & Cimpian, J. R. (2016). Estimating causal effects of education interventions using a two-rating regression discontinuity design: Lessons from a simulation study and an application. *Journal of Research on Educational Effectiveness*. doi:10.1080/19345747.2016.1219436

Ramey, C. T., Gallagher, J. J., Campbell, F., Wasik, B. H., & Sparling, J. J. (2004). *Carolina Abecedarian Project and the Carolina Approach to Responsive Education (care), 1972-1992* (icpsr04091-v1) [Computer software manual]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [Distributor]. Retrieved from http://doi.org/10.3886/icpsr04091.v1

Ruppert, D., Sheather, S. J., & Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, *90*, 1257–1270.

Schochet, P. (2009). Statistical power for regression discontinuity designs in education evaluations. *Journal of Educational and Behavioral Statistics*, *34*, 238–266.

Schochet, P., Cook, T., Deke, J., Imbens, G., Lockwood, J. R., Porter, J., & Smith, J. (2010). Standards for regression discontinuity designs. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/wwc_rd.pdf

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin Company.

Stigler, M., & Quast, B. (2015). rddtools: Toolbox for regression discontinuity design ('rdd') [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=rddtools (R package version 0.4.0).

Thistlethwaite, D. L., & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, *51*, 309–317. doi:10.1037/h0044319

Trochim, W. M. (1984). *Research design for program evaluation: The regression-discontinuity approach*. Beverly Hills, CA: Sage.

Wong, V. C., Steiner, P. M., & Cook, T. D. (2013). Analyzing regression-discontinuity designs with multiple assignment variables a comparative study of four estimation methods. *Journal of Educational and Behavioral Statistics*, *38*, 107–141.

Zeileis, A. (2004). Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software*, *11*, 1–17.

Zeileis, A., & Croissant, Y. (2010). Extended model formulas in R: Multiple parts and multiple responses. *Journal of Statistical Software*, *34*, 1–13. Retrieved from http://www.jstatsoft.org/v34/i01/

Zeileis, A., & Hothorn, T. (2002). Diagnostic checking in regression relationships. *R News*, *2*, 7–10. Retrieved from http://CRAN.R-project.org/doc/Rnews/

Zhang, L., Hu, S., Sun, L., & Pu, S. (2016). The effect of Florida's bright futures program on college choice: A regression discontinuity approach. *The Journal of Higher Education*, *87*, 115–146. doi:10.1353/jhe.2016.0003

Authors

FELIX THOEMMES is an assistant professor at Cornell University, G62A MVR, Ithaca, NY 14850; email: felix.thoemmes@cornell.edu. His research interests are design- and model-based causal inference, and missing data analysis.

WANG LIAO is a PhD candidate in Communication at Cornell University. 455 Mann Library Building, Ithaca, NY 14850; email: wl483@cornell.edu. His research interests are the roles of communication in emergent structures and processes in groups, and quantitative methods for causal inference in observational data and social networks.

ZE JIN is a PhD student in Statistics at Cornell University, 101 Malott Hall, Ithaca, NY 14850; email: zj58@cornell.edu. His research interests are nonparametric testing and statistical learning.