



PROJECT MUSE®

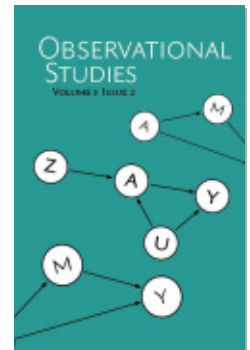
The Choice of Neighborhood in Regression Discontinuity Designs

Matias D. Cattaneo, Gonzalo Vazquez-Bare

Observational Studies, Volume 3, Issue 2, 2017, pp. 134-146 (Article)

Published by University of Pennsylvania Press

DOI: <https://doi.org/10.1353/obs.2017.0002>



➔ *For additional information about this article*

<https://muse.jhu.edu/article/793386/summary>

The Choice of Neighborhood in Regression Discontinuity Designs

Matias D. Cattaneo

*Department of Economics and Department of Statistics
University of Michigan
Ann Arbor, MI 48104, US*

cattaneo@umich.edu

Gonzalo Vazquez-Bare

*Department of Economics
University of Michigan
Ann Arbor, MI 48104, US*

gvazquez@umich.edu

This version: January 22, 2017

The seminal paper of [Thistlethwaite and Campbell \(1960\)](#) is one of the greatest breakthroughs in program evaluation and causal inference for observational studies. The originally coined *Regression-Discontinuity Analysis*, and nowadays widely known as the Regression Discontinuity (RD) design, is likely the most credible and internally valid quantitative approach for the analysis and interpretation of non-experimental data. Early reviews and perspectives on RD designs include [Cook \(2008\)](#), [Imbens and Lemieux \(2008\)](#) and [Lee and Lemieux \(2010\)](#); see also [Cattaneo and Escanciano \(2017\)](#) for a contemporaneous edited volume with more recent overviews, discussions, and references.

The key design feature in RD is that units have an observable running variable, score or index, and are assigned to treatment whenever this variable exceeds a known cutoff. Empirical work in RD designs seeks to compare the response of units just below the cutoff (control group) to the response of units just above (treatment group) to learn about the treatment effects of interest. It is by now generally recognized that the most important task in practice is to select the appropriate neighborhood near the cutoff, that is, to correctly determine which observations near the cutoff will be used. Localizing near the cutoff is crucial because empirical findings can be quite sensitive to which observations are included in the analysis. Several neighborhood selection methods have been developed in the literature depending on the goal (e.g., estimation, inference, falsification, graphical presentation), the underlying assumptions invoked (e.g., parametric specification, continuity/nonparametric specification, local randomization), the parameter of interest (e.g., sharp, fuzzy, kink), and even the specific design (e.g., single-cutoff, multi-cutoff, geographic).

We offer a comprehensive discussion of both deprecated and modern neighborhood selection approaches available in the literature, following their historical as well as methodological evolution over the last decades. We focus on the prototypical case of a continuously distributed running variable for the most part, though we also discuss the discrete-valued case towards the end of the discussion. The bulk of the presentation focuses on neighborhood selection for estimation and inference, outlining different methods and approaches according to, roughly speaking, the size of a typical selected neighborhood in each case, going from the largest to smallest neighborhood. [Figure 1](#) provides a heuristic summary, which we

Figure 1: Choice of neighborhood (single-cutoff sharp RD)

discuss in detail throughout this article. This ordering among neighborhood selectors is not strict, but it does reflect typical empirical results and may hold under reasonable assumptions and conditions. Furthermore, this ordering follows roughly the historical evolution in the empirical and methodological RD literatures. To complement the discussion, we also reflect briefly on neighborhood selection for several falsification and validation approaches that have recently been proposed in the RD literature.

Our main methodological discussion and recommendations apply not only to the most standard single-cutoff sharp RD design but also more generally to many other RD settings such as fuzzy RD designs (e.g., [Hahn et al., 2001](#)), kink RD designs (e.g., [Card et al., 2015, 2017](#)), geographic RD designs (e.g., [Keele and Titiunik, 2015](#); [Keele et al., 2017](#)), multi-cutoff RD designs (e.g., [Cattaneo et al., 2016b](#)), derivative estimation and stability testing (e.g., [Dong and Lewbel, 2015](#); [Cerulli et al., 2017](#)), distributional treatment effects (e.g., [Shen and Zhang, 2016](#)), and density discontinuity designs (e.g., [Jales and Yu, 2017](#)). Adapting the main discussion to these other RD settings is not difficult because our main methodological points are conceptual, and hence not directly tied to any specific RD setup (i.e., only the underlying technicalities or specific features of the problem considered would change, not the general message).

The last section summarizes the implications of our methodological points in the form of concrete recommendations for practice. This section builds on the most recent, and still rapidly expanding, methodological literature on RD designs. Our recommendations are given in general terms so they can be followed in most, if not all, empirical settings employing any regression discontinuity design.

Choosing a Neighborhood

No matter the approach taken (parametric, nonparametric, local randomization) or specific goal (estimation, inference, falsification, graphical presentation) when selecting a neighborhood around the RD cutoff, researchers must impose assumptions, explicitly or implicitly, which they deem reasonable and applicable for the empirical problem at hand. Therefore, it is rarely the case that a method strictly dominates everything else: at the core of the underlying reasoning often lays a trade off between efficiency and robustness, where some methods will be more “efficient” under the assumptions imposed, but more sensitive to violations of these assumptions, while other methods will be more “robust” to such violations but usually at the cost of some loss in precision.

We do rank approaches because we take a stand on the efficiency-robustness trade off: since empirical researchers never know the features of the underlying data generating process, and pre-testing for such features (when possible) can lead to other methodological and practical problems in terms of estimation and inference, we favor procedures that are valid under weaker assumptions, that is, we prefer more *robust* methods. From this robustness perspective, a clear ranking among most neighborhood selectors emerges naturally, as we discuss precisely in this section.

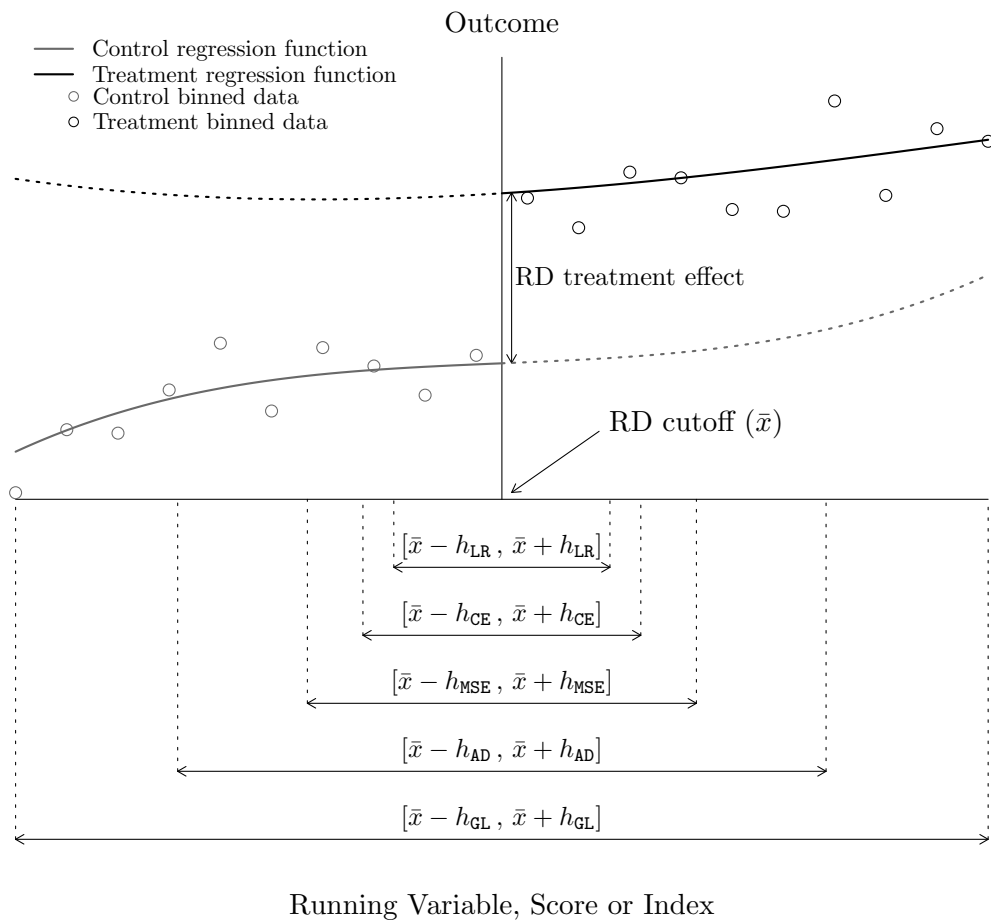


Figure 1: Choice of neighborhood (single-cutoff sharp RD)

Ad-Hoc Neighborhood

We classify as ad-hoc neighborhood selection those approaches that do not employ the data at all to select the neighborhood or, at least, not in a systematic and objective way. These methods were quite popular in the early stages of the RD design developments, but are nowadays widely viewed as inappropriate for the analysis and interpretation of RD designs. We discuss them here not only because they were the first used, but also because they give a natural introduction to the modern approaches outlined further below.

The very first (ad-hoc) method for selecting a neighborhood around the RD cutoff was to employ the full support of the data together with a linear regression model for estimation and inference, which traces back all the way to [Thistlethwaite and Campbell \(1960\)](#). Later, once the crucial role that global extrapolation plays in this approach was fully appreciated, practitioners moved towards either (i) selecting a “smaller” neighborhood in an arbitrary way (and still use linear regression), or (ii) employing higher-order polynomial regression (and still use the full support of the data). These two approaches were popular for some time in early empirical work employing RD designs. [Figure 1](#) offers a graphical schematic of these methods: h_{GL} stands for the “global” or full support approach, where usually a higher-order polynomial is used, and h_{AD} denotes the ad-hoc “local” neighborhood, where the researcher chooses the bandwidth in arbitrary manner. This smaller ad-hoc, parametric linear regression neighborhood is depicted as “large” relative to other modern methods discussed below because in our experience most empirical applications and/or real datasets we have reanalyzed employing the latter methods typically exhibited this pattern. In other words, ad-hoc neighborhoods were usually chosen to be large relative to what automatic, data-driven methods would have selected instead.

Obvious concerns with methods that select a neighborhood around the RD cutoff in an ad-hoc way are: (i) lack of objectivity, (ii) lack of comparability, and (iii) lack of control over the researcher’s discretion. In contrast, all of the data-driven procedures that we discuss below avoid these issues, and hence they provide at least a useful benchmark for empirical work exploiting regression discontinuity designs.

Another important, but more subtle, worry related to ad-hoc neighborhood selection methods relates to the underlying assumptions imposed when conducting estimation and inference, which many times are not even explicitly acknowledged by practitioners. To be specific, underlying any of the ad-hoc methods commonly encountered in empirical work there is a crucial assumption: the regression function is correctly specified or, at least, any misspecification error is small enough to be ignored. This parametric approach to RD designs gives practitioners justification to employ standard least squares results when conducting estimation and inference. While such parametric approach is, of course, correct when the regression functions are correctly specified, in general there is no reason for the *unknown* conditional expectations to have the exact (or close enough) parametric form postulated, and hence misspecification errors can be a serious concern. Furthermore, it is now well recognized that employing higher-order polynomial approximations over a large support is highly detrimental, when the goal is to learn something about a boundary point as in RD designs, because such an approach leads to counterintuitive weighting of observations ([Gelman and Imbens, 2014](#)) and erratic behavior of the estimator near the boundary (usually known as the Runge’s phenomenon, see [Calonico et al., 2015](#), for more discussion).

Finally, some empirical researchers have used ad-hoc neighborhood selectors based on data-driven procedures from the nonparametric literature, such as those related to bandwidth selection for kernel-based density estimation (Wand and Jones, 1995) or local polynomial estimation at an interior point (Fan and Gijbels, 1996). While these approaches are data-driven, they are also ad-hoc in sense that they are not tailored to RD designs, and hence they can lead to invalid (or at least suboptimal) estimation and inference procedures. These approaches are not very popular in modern empirical work employing RD designs, nor are they recommended or theoretically justified, and therefore we do not discuss them further.

The concerns and criticisms outlined above have led modern researchers to employ fully data-driven, objective neighborhood selectors to conduct estimation and inference in RD designs. Ad-hoc methods are nowadays deprecated and dismissed among most well trained practitioners and methodologists. If used, they are typically presented as supplementary evidence after reporting results based on the data-driven methods discussed next, which enjoy demonstrably optimality and/or robustness properties.

Local Polynomial Neighborhood: MSE-Optimal Point Estimation

In this and related approaches, the neighborhood takes the form $[\bar{x} - h, \bar{x} + h]$ and hence is determined by a choice of bandwidth h . Imbens and Kalyanaraman (2012, IK hereafter) were the first to propose an objective neighborhood selector specifically tailored for RD designs. They developed a Mean Squared Error (MSE) optimal bandwidth choice for the local-linear regression point estimator in sharp and fuzzy RD designs. This result was later extended to (i) general local polynomial point estimators, (ii) kink RD designs, (iii) clustered data, (iv) inclusion of pre-intervention covariates, and (v) different bandwidth choices on the left and on the right of the cutoff, in a sequence of more recent papers (Calonico et al., 2014; Bartalotti and Brummet, 2017; Calonico et al., 2016c).

The MSE-optimal bandwidth takes the form $h_{\text{MSE}} = C_{\text{MSE}} \cdot n^{-1/(2p+3)}$, where n denotes the total sample size available, p denotes the polynomial order used for estimation ($p = 1$ for linear regression), and the constant C_{MSE} involves several known and unknown quantities that depend on objects such as the kernel function, p , the parameter of interest, the asymptotic bias and variance of the estimator, the evaluation point (in multi-cutoff or geographic RD designs), and even whether additional pre-intervention covariates were included in the estimation. This approach is also depicted in Figure 1.

Given a sample size n , the infeasible MSE-optimal neighborhood $[\bar{x} - h_{\text{MSE}}, \bar{x} + h_{\text{MSE}}]$ will be larger as the value of the unknown constant C_{MSE} increases. This constant, in turn, will become larger whenever the variability of the estimator and/or model increases near the cutoff (e.g., p is larger, the conditional variance of the outcome is larger or the density of observations near the cutoff is smaller) and whenever the parametric approximation improves near the cutoff (i.e., less misspecification bias). In practice, h_{MSE} is constructed by first forming a preliminary estimator \hat{C}_{MSE} of the unknown constant C_{MSE} , leading to the estimated bandwidth $\hat{h}_{\text{MSE}} = \hat{C}_{\text{MSE}} \cdot n^{-1/(2p+3)}$, and therefore the selected neighborhood around the RD cutoff \bar{x} takes the form $[\bar{x} - \hat{h}_{\text{MSE}}, \bar{x} + \hat{h}_{\text{MSE}}]$. IK proposed a first-generation plug-in rule leading to a bandwidth selector \hat{h}_{MSE} , based on a simple reference model and (possibly inconsistent) plug-in estimators. An improved, second-generation bandwidth selector was

later developed by [Calonico et al. \(2014, 2016c\)](#), which enjoys demonstrably superior finite and large sample properties relative to the original IK's bandwidth selector. See, e.g., [Wand and Jones \(1995\)](#) for a discussion of first- and second-generation bandwidth selectors, and their statistical properties.

In this MSE-optimal point estimation approach, only observations with their running variable lying within the selected neighborhood $[\bar{x} - \hat{h}_{\text{MSE}}, \bar{x} + \hat{h}_{\text{MSE}}]$ are used for estimation of the RD treatment effect. This estimator is fully data-driven, objective and optimal in a mean squared error sense, which makes it highly desirable for empirical work, at least as a benchmark estimate. Employing second-generation plug-in bandwidth selectors lead to superior performance of the MSE-optimal RD treatment effect estimator in finite and large samples.

At the same time, the MSE-optimal point estimator cannot be used directly for inference, that is, for constructing confidence intervals, conducting hypothesis tests or assessing statistical significance. At the core of the argument lays a fundamental logical inconsistency: the neighborhood $[\bar{x} - \hat{h}_{\text{MSE}}, \bar{x} + \hat{h}_{\text{MSE}}]$ is selected for MSE-optimal point estimation and hence balances bias-squared and variance in a way that makes, by construction, inference invalid when the same observations and RD estimator are used. There is no way out of this logical inconsistency: if one assumes that the misspecification bias is not present (i.e., $\text{bias} = 0$), then h_{MSE} is necessarily not well defined because $C_{\text{MSE}} \propto 1/\text{bias}$. In other words, to be able to employ \hat{h}_{MSE} in the first place, one needs to assume the existence of a misspecification error (bias), but it is this very same bias that makes inference invalid when the MSE-optimal point estimator is used for inference purposes.

The invalidity of inference procedures based on the MSE-optimal point estimator was ignored for some time among practitioners. [Calonico et al. \(2014\)](#) highlighted the detrimental consequences of ignoring this misspecification bias and, to solve this inferential problem, proposed a new inference approach based on bias correction of the point estimate, coined *robust bias correction*. The idea behind this method, which allows employing the MSE optimal bandwidth and point estimator, is to adjust the MSE-optimal RD point estimator by estimating its bias and also to adjust the variance estimator used for Studentization purposes when conducting inference. For example, when compared to conventional confidence intervals based on ad-hoc neighborhood selection that rely on standard least squares results, robust bias correction adjusts this confidence interval by recentering (bias correction) and rescaling (robust variance estimator) it. The robust bias corrected RD confidence intervals are fully compatible with employing observations with score lying inside the MSE-optimal selected neighborhood $[\bar{x} - \hat{h}_{\text{MSE}}, \bar{x} + \hat{h}_{\text{MSE}}]$, while still giving valid inference methods.

Furthermore, [Calonico et al. \(2016b,a\)](#) recently showed that robust bias correction gives demonstrably superior inference when compared to alternative methods employing smaller neighborhoods than $[\bar{x} - \hat{h}_{\text{MSE}}, \bar{x} + \hat{h}_{\text{MSE}}]$, that is, when shrinking \hat{h}_{MSE} (known as under-smoothing).

In sum, although the MSE-optimal neighborhood can be used for optimal point estimation, standard least squares inference methods cannot be used for inference, and robust bias corrected confidence intervals and related procedures should be used instead. Estimation and robust bias-corrected inference employing the MSE-optimal neighborhood is more robust to the presence of misspecification bias because it does not rely on strong functional

form assumptions about the unknown conditional expectations. As a consequence, these empirical methods are preferred to those relying on ad-hoc neighborhood selectors.

Local Polynomial Neighborhood: CE-Optimal Robust Bias-Corrected Inference

The MSE-optimal neighborhood $[\bar{x} - \hat{h}_{\text{MSE}}, \bar{x} + \hat{h}_{\text{MSE}}]$ is quite popular in empirical work because it gives an optimal RD treatment effect estimator. As discussed above, the same neighborhood can be used for inference when robust bias correction techniques are employed. However, this neighborhood need not be optimal when the goal is inference. Indeed, [Calonico et al. \(2016b,a\)](#) showed that a different, smaller neighborhood must be used when the goal is constructing optimal confidence intervals in the sense of having the smallest coverage error (CE) probability.

To be more precise, the CE-optimal neighborhood around the RD cutoff is $[\bar{x} - h_{\text{CE}}, \bar{x} + h_{\text{CE}}]$ with $h_{\text{CE}} = C_{\text{CE}} \cdot n^{-1/(p+3)}$ and C_{CE} another unknown constant, fundamentally different from C_{MSE} , which needs to be estimated in practice because it also involves unknown quantities. This new neighborhood offers robust bias corrected confidence intervals with demonstrably superior optimality properties for inference, when compared to those confidence intervals constructed using the MSE-optimal neighborhood $[\bar{x} - h_{\text{MSE}}, \bar{x} + h_{\text{MSE}}]$. It follows that $[\bar{x} - h_{\text{CE}}, \bar{x} + h_{\text{CE}}] \subset [\bar{x} - h_{\text{MSE}}, \bar{x} + h_{\text{MSE}}]$, in large samples, because $h_{\text{CE}} < h_{\text{MSE}}$. The same logic also applies to their estimated versions. [Figure 1](#) depicts the CE-optimal choice.

Therefore, in empirical applications, the MSE-optimal neighborhood $[\bar{x} - \hat{h}_{\text{MSE}}, \bar{x} + \hat{h}_{\text{MSE}}]$ can be used for MSE-optimal RD treatment effect point estimation, and the CE-optimal neighborhood $[\bar{x} - \hat{h}_{\text{CE}}, \bar{x} + \hat{h}_{\text{CE}}]$, with \hat{h}_{CE} denoting a data-driven implementation of h_{CE} , can be used to form CE-optimal robust bias corrected confidence intervals. Employing observations with their score within the CE-optimal neighborhood for point estimation purposes is theoretically allowed but not advisable because the resulting RD treatment effect estimator will have too much variability.

As is the case for the MSE-optimal estimation and robust bias-corrected methods discussed previously, the CE-optimal inference methods are more robust than those based on ad-hoc neighborhood selectors because they optimally trade off misspecification bias underlying the local polynomial approximations to the unknown regression functions, variability of the test statistic (not just the point estimator), and other features of the underlying unknown data generating process.

Local Randomization Neighborhood

The neighborhood selection approaches outlined so far are all related, one way or another, to local or global polynomial regression approximations of the unknown conditional expectations. As such, these methods are based on extrapolation towards the cutoff point \bar{x} , using either observations near the cutoff but within the selected neighborhood or simply using all observations in the sample. An alternative approach for identification, estimation and inference in RD designs is based on the idea of local randomization, which assumes that there exists a neighborhood around the cutoff where the underlying data generating process is one (approximately) mimicking a randomized controlled trial (RCT). This heuristic idea

was originally put forward by Lee (2008), and formally developed in Cattaneo et al. (2015), Cattaneo et al. (2017), Sekhon and Titiunik (2017), and references therein.

From this point of view, neighborhood selection is quite different because substantially different assumptions are placed on the underlying data generating process. In other words, none of the neighborhood selectors discussed previously can be used within the local randomization framework because it would be very difficult to rationalize their validity. Cattaneo et al. (2015, 2017) introduced a new neighborhood selection approach: instead of optimizing a point estimator in a mean squared error sense or a confidence interval in a coverage error sense, their idea is to employ pre-intervention covariates and optimize in the sense of minimizing the statistical evidence against the local randomization assumption. To be more precise, the proposal is to conduct a sequence of “balance” or “placebo” tests of no treatment effect on exogenous covariates known to be unaffected by treatment near the RD cutoff, for different proposed neighborhoods, and then select the largest neighborhood that is compatible with local randomization (i.e., the largest neighborhood for which the null hypothesis is not rejected). Under regularity conditions, this method will select a valid neighborhood, which will tend to be smaller than the true neighborhood because no correction for multiple testing is used. Since by construction the neighborhoods are nested, not using multiple testing corrections is appropriate from a robustness perspective in this case.

This neighborhood selection method based on pre-intervention covariate balance tests is similar in spirit to procedures commonly used in the *matching* literature to select a *matched* sample when analyzing observational data under a conditional independence or ignorability assumption (e.g., Imbens and Rubin, 2015). Despite the similarities, the RD local randomization neighborhood selection method is different in that it explicitly exploits the structure of the RD design by localizing near the cutoff and crucially relying on balance tests in a sequence of nested windows.

While the neighborhood selector described above, and subsequent inference procedures, could be implemented via standard large sample estimation and inference methods for RCTs, Cattaneo et al. (2015, 2017) propose to employ randomization inference methods, which are finite sample valid. The main rationale underlying this proposal is at the heart of the specific setting of RD designs: a local randomization assumption in RD designs is most likely to hold, or at least give a good approximation, in a very small neighborhood around the RD cutoff where usually very few observations are available for estimation and inference. Therefore, randomization inference methods, or other analogous finite sample valid methods such as permutation inference, are most useful in the RD context because large sample approximations are unlikely to provide a good enough approximation. Applying the above neighborhood selector to several applications, we have systematically found very small neighborhoods. Thus, based on the methodological arguments and empirical evidence, Figure 1 depicts the local randomization neighborhood as the smallest of all the possible neighborhoods available for estimation and inference in RD designs.

Local randomization methods are fundamentally different from local polynomial methods, both in assumptions and implementation, and therefore they provide a useful robustness check whenever both methods can be used. Furthermore, another important advantage of local randomization methods is that they can handle discrete running variables without any additional assumptions, and randomization inference methods are again most natural whenever the sample size is small. In contrast, local polynomial methods would require

additional parametric assumptions to be valid when the running variable is discrete. This fact is neither surprising nor worrisome, however, since when the running variable is actually discrete there is no need for extrapolation to begin with. It is much more natural and useful to simply consider only the observations having their running variable at the closest discrete value(s) relative to the RD cutoff, on either side, and then use them to conduct estimation and inference. This, of course, changes slightly the parameter of interest, though this is quite natural whenever the running variable has a discrete distribution.

Falsification/Validation Neighborhood

Our discussion so far has focused on neighborhood selection around the RD cutoff for estimation and inference, explicitly relying on different assumptions (i.e., parametric modeling, nonparametric modeling, local randomization). In this subsection, we briefly discuss the related issue of neighborhood selection for falsification/validation of RD designs.

There are two basic falsification/validation methods in the RD literature: (i) tests looking at the continuity of the density of the running variable, and (ii) tests looking at the absence of RD treatment effects on pre-intervention covariates and “placebo” or unaffected outcomes. Both of these approaches also require “localizing” around the RD cutoff. [Calonico et al. \(2015\)](#) discuss related graphical falsification and presentation methods using RD plots, which we do not discuss here to conserve space.

Continuity in the density of the running variable was originally proposed by [McCrary \(2008\)](#), and is by now extremely popular in empirical work. This test is usually understood as providing evidence, or lack thereof, of units having intentionally changed or manipulated their score value near the cutoff. [Cattaneo et al. \(2016a\)](#) recently developed a more robust, nonparametric local polynomial inference method that avoids selecting multiple tuning parameters when implementing this density test. In their approach, the neighborhood is selected in a data-driven, objective way with the explicit goal of minimizing the MSE of the density estimators used to construct the test statistic. It is not possible to determine whether this MSE-optimal neighborhood will be larger or smaller than any of the neighborhoods described previously, because the objective and estimation methods are quite different (i.e., density estimation vs. conditional expectation estimation). What is clear is that the neighborhood for the density test should not be equal, in general, to any of the other neighborhoods: i.e., it should be chosen explicitly for the goal at hand, falsification testing based on local polynomial density estimation. In addition, [Frandsen \(2017\)](#) also developed a “continuity in density” testing approach for the case of discrete running variable. For this method, at present, there is no optimal way of choosing a neighborhood beyond some ad-hoc selection, though the procedure allows for very few “observations” (mass points) near the cutoff because it relies on finite sample inference methods (formally justified by some large sample approximations). Again, there is no reason why the “neighborhood” used for this density test with discrete running variable should coincide with any of the other neighborhoods, and in general it will not.

The density test is quite useful and intuitive because it exploits some of the specific features of RD designs. The second falsification/validation method commonly used in practice is more standard, in the sense that it is directly imported from common practice in other experimental and non-experimental settings. Specifically, this second method seeks to test

whether there is evidence of an RD treatment effect on covariates and outcomes that should (or, at least, are assumed to) be unaffected by the treatment. This approach is conceptually analogous to testing for a treatment effect on pre-intervention covariates in the context of RCTs, and can be implemented using directly the modern local polynomial and randomization inference methods described in the previous sections for RD estimation and inference. As an alternative, [Canay and Kamat \(2016\)](#) have recently proposed a permutation inference approach for falsification testing based on comparing the whole distribution of treatment and control groups, which is also justified via some large sample approximations near the cutoff. The authors conduct neighborhood selection using a rule-of-thumb based on a simple reference model, which leads to yet another neighborhood to be used in applications when implementing their method.

Recommendations for Practice and Final Remarks

To conclude, we offer some practical recommendations for empirical work. We build on the methodological points put forward above, and hence only offer very brief takeaway methodological points:

1. Always employ RD optimal data-driven neighborhood (bandwidth or window) selectors, at least as a benchmark or starting point. This gives objectivity and robustness because it incorporates explicitly empirical features such as density of observations, variability of the data, or curvature of the unknown regression functions, in a principled way.
2. Employ data-driven neighborhood (bandwidth or window) selectors according to the specific goal and assumptions imposed, which should also be explicitly stated and explained. There is no one neighborhood selector appropriate for all objectives when using local polynomial approximations, and even for local randomization methods sensitivity analysis with respect to the neighborhood used is very important.
3. Do not employ the *same* neighborhood for different outcome variables, pre-intervention covariates (if conducting falsification testing), estimation and inference procedures, or falsification methods. Using the same neighborhood for different goals, outcomes or samples disregards the specific empirical features (e.g., number of observations near the cutoff, variability or curvature), and will lead to unreliable empirical results due to invalidity of the methods employed.

[Thistlethwaite and Campbell \(1960\)](#) introduced one of the best non-experimental methods for the analysis and interpretation of observational studies. In recent years many methodological and theoretical developments not only have extended the basic regression discontinuity design to many other settings, but also have provided major improvements in terms of presentation, estimation, inference and falsification for empirical practice. In this discussion, we focused on arguably the most important and challenging part of analyzing and implementing RD designs: neighborhood, bandwidth or window selection around the RD cutoff. Much methodological progress has been achieved in recent years regarding this important task, making RD designs even more credible and robust in applications.

Acknowledgments

We thank our close collaborators and colleagues, Sebastian Calonico, Max Farrell, Michael Jansson, Xinwei Ma, and Rocio Titiunik, whose ideas and criticisms over the years have shaped this discussion. We also thank Justin McCrary and David McKenzie for recent energizing discussions and excellent comments on RD design methodology. Cattaneo gratefully acknowledges financial support from the National Science Foundation through grant SES-1357561.

R and Stata software packages implementing the main neighborhood (i.e., bandwidth or window) selectors discussed above are available at:

<https://sites.google.com/site/rdpackages>

References

- Bartalotti, O. and Brummet, Q. (2017). Regression discontinuity designs with clustered data. In Cattaneo, M. D. and Escanciano, J. C., editors, *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, volume 38)*. Emerald Group Publishing, *forthcoming*.
- Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2016a). Coverage error optimal confidence intervals for regression discontinuity designs. Working paper, University of Michigan.
- Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2016b). On the effect of bias estimation on coverage accuracy in nonparametric inference. arXiv:1508.02973.
- Calonico, S., Cattaneo, M. D., Farrell, M. H., and Titiunik, R. (2016c). Regression discontinuity designs using covariates. Working paper, University of Michigan.
- Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326.
- Calonico, S., Cattaneo, M. D., and Titiunik, R. (2015). Optimal data-driven regression discontinuity plots. *Journal of the American Statistical Association*, 110(512):1753–1769.
- Canay, I. A. and Kamat, V. (2016). Approximate permutation tests and induced order statistics in the regression discontinuity design. Working paper, Northwestern University.
- Card, D., Lee, D. S., Pei, Z., and Weber, A. (2015). Inference on causal effects in a generalized regression kink design. *Econometrica*, 83(6):2453–2483.
- Card, D., Lee, D. S., Pei, Z., and Weber, A. (2017). Regression kink design: Theory and practice. In Cattaneo, M. D. and Escanciano, J. C., editors, *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, volume 38)*. Emerald Group Publishing, *forthcoming*.
- Cattaneo, M. D. and Escanciano, J. C. (2017). *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, volume 38)*. Emerald Group Publishing, *forthcoming*.

- Cattaneo, M. D., Frandsen, B., and Titiunik, R. (2015). Randomization inference in the regression discontinuity design: An application to party advantages in the u.s. senate. *Journal of Causal Inference*, 3(1):1–24.
- Cattaneo, M. D., Jansson, M., and Ma, X. (2016a). Simple local regression distribution estimators with an application to manipulation testing. Working paper, University of Michigan.
- Cattaneo, M. D., Keele, L., Titiunik, R., and Vazquez-Bare, G. (2016b). Interpreting regression discontinuity designs with multiple cutoffs. *Journal of Politics*, 78(4):1229–1248.
- Cattaneo, M. D., Titiunik, R., and Vazquez-Bare, G. (2017). Comparing inference approaches for rd designs: A reexamination of the effect of head start on child mortality. *Journal of Policy Analysis and Management*, forthcoming.
- Cerulli, G., Dong, Y., Lewbel, A., and Poulsen, A. (2017). Testing stability of regression discontinuity models. In Cattaneo, M. D. and Escanciano, J. C., editors, *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, volume 38)*. Emerald Group Publishing, forthcoming.
- Cook, T. D. (2008). “waiting for life to arrive”: A history of the regression-discontinuity design in psychology, statistics and economics. *Journal of Econometrics*, 142(2):636–654.
- Dong, Y. and Lewbel, A. (2015). Identifying the effect of changing the policy threshold in regression discontinuity models. *Review of Economics and Statistics*, 97(5):1081–1092.
- Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman & Hall/CRC, New York.
- Frandsen, B. (2017). Party bias in union representation elections: Testing for manipulation in the regression discontinuity design when the running variable is discrete. In Cattaneo, M. D. and Escanciano, J. C., editors, *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, volume 38)*. Emerald Group Publishing, forthcoming.
- Gelman, A. and Imbens, G. W. (2014). Why high-order polynomials should not be used in regression discontinuity designs. NBER working paper 20405.
- Hahn, J., Todd, P., and van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209.
- Imbens, G. and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2):615–635.
- Imbens, G. W. and Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies*, 79(3):933–959.
- Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

- Jales, H. and Yu, Z. (2017). Identification and estimation using a density discontinuity approach. In Cattaneo, M. D. and Escanciano, J. C., editors, *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, volume 38)*. Emerald Group Publishing, *forthcoming*.
- Keele, L., Lorch, S., Passarella, M., Small, D., and Titiunik, R. (2017). An overview of geographically discontinuous treatment assignments with an application to children’s health insurance. In Cattaneo, M. D. and Escanciano, J. C., editors, *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, volume 38)*. Emerald Group Publishing, *forthcoming*.
- Keele, L. J. and Titiunik, R. (2015). Geographic boundaries as regression discontinuities. *Political Analysis*, 23(1):127–155.
- Lee, D. S. (2008). Randomized experiments from non-random selection in u.s. house elections. *Journal of Econometrics*, 142(2):675–697.
- Lee, D. S. and 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.
- Sekhon, J. and Titiunik, R. (2017). On interpreting the regression discontinuity design as a local experiment. In Cattaneo, M. D. and Escanciano, J. C., editors, *Regression Discontinuity Designs: Theory and Applications (Advances in Econometrics, volume 38)*. Emerald Group Publishing, *forthcoming*.
- Shen, S. and Zhang, X. (2016). Distributional regression discontinuity: Theory and applications. *Review of Economics and Statistics*, *forthcoming*.
- Thistlethwaite, D. L. and Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex-post facto experiment. *Journal of Educational Psychology*, 51(6):309–317.
- Wand, M. and Jones, M. (1995). *Kernel Smoothing*. Chapman & Hall/CRC, Florida.