Abstract Title Page
Not included in page count.

Title:

The Generalized Regression Discontinuity Design: Using multiple assignment variables and cutoffs to estimate treatment effects

Author(s):

Vivian C. Wong, School of Education and Social Policy, Northwestern University
Peter M. Steiner, Institute for Policy Research, Northwestern University; Institute for Advanced Studies, Vienna, Austria
Thomas D. Cook, Institute for Policy Research, Northwestern University

Abstract Body

Background/context:
In regression-discontinuity design (RDD), units must be assigned to treatment and comparison conditions solely on the basis of a cutoff score on a continuous assignment variable. The assignment variable is any measure taken prior to the treatment intervention, and there is no requirement that the measure be reliable. Units that score on one side of the cutoff score are assigned to the treatment while units that score on the other side are assigned to the comparison. Treatment effects then are estimated by examining the displacement of the regression line at the cutoff point determining program receipt.

In many recent examples in Education where RDDs have been applied, multiple assignment and cutoff variables were available for determining treatment and control conditions. For example, students may be assigned to remedial education interventions based on missing a reading cutoff, a math cutoff, or both (Jacob and Lefgren, 2004). Under No Child Left Behind, schools may miss adequate yearly progress (AYP) if they fail to meet *one of 39 possible criteria*, including percent proficiency and participation rate requirements for each subgroup, and attendance and graduation rate requirements for the whole school and districts. Within each subgroup for each subject area, schools have multiple methods for meeting state percent proficiency requirements. In addition to reaching states' annual proficiency targets, schools may make AYP cutoff by falling just within the "confidence interval" around the state threshold, by averaging proficiency rates of students across multiple years, or by reducing the percentage of non-proficient students from the prior year by 10%. These alternative methods for making AYP are not simply "misallocated cases" in RDD, but they are "exemption rules" that are completely observed by the researcher, and are systematically and uniformly applied to all schools within the state.

Researchers have handled multiple assignment mechanisms in RDD in one of two ways. They may choose a single assignment variable and cutoff, and define treatment effects based on this assignment mechanism alone. This was the case in Jacob and Lefgren's (2004) study, where assignment to treatment was based on students' performance on a reading achievement test and the school district's minimum threshold for reading scores. Alternatively, they may pool scores across different assignment variables by centering at each unit's respective cutoff score. Gill, Lockwood, Martorell, Setodji, and Booker (2007) employed this approach in an RDD study examining the effects of No Child Left Behind policy on student achievement scores. The RD cutoff was based on the criterion on which schools achieved their lowest scores relative to the cutoff. Schools missed AYP if they failed to meet the proficiency requirement for any one subgroup or subtest. By examining only subgroups in subject areas that scored lowest compared to the cutoff, researchers were able to determine whether schools made AYP or not. The RD analysis was then conducted to examine whether a discontinuity existed at the cutoff of the dimension that represented the lowest score for each school.

The first approach suffers from several limitations. The most important is that treatment contamination may occur when alternative assignment mechanisms are simply dropped. In the Jacob and Lefgren (2004) example, students that missed the math cutoff but not the reading would be assigned to the comparison group. But these students may have received intervention services that would affect their outcome scores, and thus possibly underestimate treatment

effects. Jacob and Lefgren (2004) handled this concern by dropping students who made the reading cutoff but missed the math. However, the strategy reduces the number of students included in the sample, and limits generalization of results. The problems are exacerbated if correlation in the treatment assignment for reading and math is low. The second approach – employed by Gill et al. (2007) – avoids these concerns and is one that we explore further in the paper. However, the concern here is heterogeneous treatment effects may be obscured by pooling various assignment variables and cutoffs into a single analysis. The paper introduces and explores a third option for handling multiple assignment mechanisms in RDD. We call it the "multivariate approach" and discuss it below.

Purpose/objective/research question/focus of study:
This paper introduces a generalization of the regression-discontinuity design. Traditionally, RDD is considered in a two-dimensional framework, with a single assignment variable and cutoff. Treatment effects are measured at a single location along the assignment variable. However, this represents a specialized (and straight-forward) application of the design; a more generalized and flexible conceptualization of RDD allows researchers to examine treatment effects along a multi-dimensional frontier using multiple assignment variables (such as math and reading scores) and cutoffs. In Section 1 of this paper, we present the generalized RDD by describing its required components, the treatment effects estimated, and advantages and limitations of the design. In Section 2, we describe two analytic approaches for estimating treatment effects for the generalized RDD. The first is the "multivariate approach," which estimates treatment effects along a multi-dimensional frontier via a regression model. This approach is based on the assumption that the researcher can model the selection mechanism completely if all assignment variables and their respective cutoffs are known and observed (as is the case for many accountability policy studies that use RDD). The second is an extension of an approach originally used by Gill et al. (2007), which we call the "centering" approach. We show that both the multivariate and centering approaches yield identical average treatment effect estimates, though they have distinct advantages and limitations. In Section 3, we present an application of the "multivariate" and "centering" approaches in an RDD example that evaluates the impacts of missing AYP under NCLB on student with disability (SWD) achievement scores. We also discuss scenarios when neither approach is appropriate for estimating treatment effects.

Setting:
Since the focus of this paper is methodological, we limit our discussion of "setting" here.

Population/Participants/Subjects:
In Section 3, we examine the effects of not making AYP on achievement scores for SWDs in Texas and Pennsylvania. We restrict both state samples to include only elementary and middle schools that were in danger of missing AYP for the SWD subgroup for the first time. Thus, our samples include only schools that 1) had an eligible SWD subgroup, 2) were not already in improvement status under NCLB, 3) were an elementary or middle school, and 4) made AYP the prior year.

Intervention/Program/Practice:
The applied section of this paper examines the impacts of schools missing AYP for the first time

on student achievement scores. Since the focus of this paper is methodological, we limit our discussion of the intervention here.

Research Design:
Quasi-experimental approach: the regression-discontinuity design

Data Collection and Analysis:
In Section 3, we used 2006-07 AYP data from Texas and Pennsylvania and simulated outcome data. The AYP data are available to the public, and are published on states' Departments of Education websites. The analysis sample consisted of 608 schools for Texas, and 1101 schools for Pennsylvania. For the RD design, we choose "percent proficient" scores for the SWD subgroup as the assignment variables, and states' proficiency thresholds as the cutoffs. The key validity threat for this study as an RD design is treatment misallocation due to multiple assignment mechanisms and exemption rules. To address this challenge, we propose two analytic strategies: the multivariate and centering approaches.

Findings/Results:
The paper shows that the multivariate and centering approaches yield identical results in a generalized RD design, and that these estimates are unbiased average causal estimates when the selection process is known and observed completely.

Our paper shows applications of the two proposed approaches, and we present our result here. For the multivariate approach, we examine estimating treatment effects along a discontinuity frontier with two assignment variables (reading and math scores) and cutoffs. In this approach, treatment effects would be estimated using the following simplified RD model (to keep notation simple we omit coefficients and error terms):

[1] $Y_R \sim RAV + MAV + Treatment$

where Y_R is school *i*'s outcome in reading, RAV is the percentage of SWDs proficient in reading, MAV is the percentage of SWDs proficient in math, and treatment is 1 if the school is in treatment (for reading, math, or both subject areas) and 0 if school is in the comparison. We assume constant and linear treatment effects along the mathematics and reading frontiers but this need not be the case. In Figure 1, we use Texas AYP data to show a two-dimensional plot of the reading assignment variable on the X axis and the math assignment variable on the Y axis. Schools are depicted by the dark blue dots on the XY plane, with those that score below one or both of the two cutoffs in the treatment group and those that score above both cutoffs in the comparison group. Thus, the red shaded area indicates schools that missed for reading only, math only, or in both subject areas. The blue shaded area indicates the position of schools on the plane that missed for neither subjects—our comparison schools. Note that we do not include any exemption rule schools in this plot.

<Insert Figure 1 about here>

In Figure 2, we show how (simulated) treatment effects could be estimated across a discontinuity frontier. Here, actual outcomes of comparison schools are on the blue part of the surface (plotted

on the Z axis) and treatment schools are in the red parts of the surface. In theory, treatment effects could be estimated by looking at the size of the discontinuities across the frontier where treatment units meet comparisons. This ranges from schools that scored very high on reading proficiency but were near the cutoff for math, to schools that scored high on math proficiency but were near the reading cutoff.

<Insert Figure 2 about here>

For the centering approach, we use data from Pennsylvania to illustrate how the approach would be implemented. More than 95% of schools in Pennsylvania made AYP in 2006–07 (for the SWD subgroup) by meeting the state cutoff or one of the following three exemption rules: a 95% confidence interval, safe harbor target, or 75% confidence interval for the safe harbor target. We restricted the dataset to include only schools that made AYP via the state cutoff or the three exemption rules identified above, and schools that did not make AYP at all. The centering procedure was carried out as follows. For each school, we calculated adjusted thresholds for each exemption rule, for each subject area. Thus, for safe harbor (SH), we examined the school's prior year subgroup performance to calculate the effective cutoff for the subgroup. For confidence interval (CI), we used the state cutoff and number of SWDs in the school to calculate the 95% confidence interval target. For the safe harbor confidence interval, we used the school's SWD performance in 2007 and 2008 and the number of SWDs in 2007 and 2008 to calculate the 75% confidence interval safe harbor target. We then chose a single cutoff for each school by taking the minimum threshold value generated by the state cutoff, confidence interval, safe harbor, and confidence interval for safe harbor rules. The school's percent proficient value was then centered based on the new minimum threshold value. The procedure was applied to each school, for each subject area. Because schools had to meet requirements in reading *and* in mathematics, the policy introduced two possible assignment variables: a centered percent proficient in reading, and a centered percent proficient in mathematics. We combined both selection mechanisms into a single assignment variable by choosing the subject area with the *minimum* centered proficiency score.

We present a series of scatterplots showing the location of schools relative to the cutoff, before and after centering. We used simulated gain scores for our outcomes because we expect that the response function for gain scores would be easier to model, and because of advantages in power. Figure 3 plots schools' *uncentered* reading assignment scores against their gain scores. As the red line indicates, the state cutoff is 63 percent and most schools are located far below the cutoff. In fact, few treatment schools score near the cutoff at all. Schools that made AYP via the confidence interval-safe harbor rule appear to score lowest, followed by those that made AYP through the safe harbor rule. Schools that made AYP via the confidence interval rule score appear to score closest to the cutoff. Less than a quarter of schools in the sample made AYP by meeting the state AMO target. The graph also plots separate treatment and comparison lowess lines of the assignment variable against gain scores. The solid line is a lowess of treatment schools, while the dotted line is a lowess of comparison schools. Note that the lowess for treatment schools does not even reach the cutoff, while the lowess for comparison schools overlaps strongly with schools on the treatment side. The intercept difference in the treatment and comparison lowess reflect the treatment effect that was included in creating our simulated gain scores.

<Insert Figures 3 & 4 about here.>

Figure 4 shows the same sample of schools, but now plotting the *centered* assignment scores against test score gains. Centering produces a plot that is much closer to the traditional RD plot, where the cutoff clearly delineates treatment from comparison schools. Treatment effects are then measured by the size of the discontinuity at the cutoff. For example, a simple estimation equation would look like the following:

[2]   Y ~ Treatment + C_AV

where Y is the outcome (gain scores) and C_AV is the centered assignment variable for reading or math. Because the assignment variable is now centered, there is no need to include control covariates for the multiple assignment and exemption rules. We assume constant and linear treatment effects in the model, but this may not necessarily be the case.

Conclusions:
This paper presents two strategies for analyzing the generalized RDD. It also illustrates why these approaches may be especially useful for evaluating accountability policies in Education, and the advantages and limitations of each approach.

Our results indicate that both approaches yield identical average causal effects when the selection mechanism is completely observed and modeled. However, each approach poses significant limitations. For the multivariate approach, we found that very few schools scored at the extreme ends of the frontier and in fact most schools were located near the intersection of the math and reading cutoffs (see Figure 1). This suggests that the reading and math assignment variables are highly correlated, and that treatment effects will only be reliably estimated around the point in the surface where the reading and math cutoffs intersect. We will need many more schools along the border of the frontier to reliably estimate heterogeneous treatment effects. Second, correct specification of the RD model introduces serious challenges because it requires the correct modeling of the response surface and identification of *all* interaction terms that may affect treatment effect estimates. Third, interpretation of treatment effects may be difficult in the multivariate approach. Because two assignment variables are included in the model, we interpret treatment effects along a discontinuity frontier, but where along the frontier should treatment effects be estimated? The challenge will be when the most reliably estimated effects are not those with the most substantive relevance.

In general, we find that the centering procedure provides the advantage of returning the analysis to a two-dimensional RD context, where assumptions about the response function are more easily probed and treatment effects can be observed directly at the cutoff. However, centering may yield biased results if the procedure introduces discontinuities in some third variables at the cutoff. As a result, careful consideration of the assignment process and careful probing of the assumptions are required. In addition, interpretation of treatment effects for the centering approach remains ambiguous due to the fact that many site-specific cutoffs are collapsed into one. Thus, heterogeneous treatment effects may be obfuscated due to the pooling of multiple cutoffs and assignment variables.

Appendixes

Appendix A. References

Gill, B., Lockwood, J. R., Martorell, F., Setodji, C. M., & Booker, K. (2008). State and local implementation of the No Child Left Behind Act. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service.

Jacob, B., & Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. *Review of Economics and Statistics, LXXXVI*, 226-244.

Appendix B. Tables and Figures

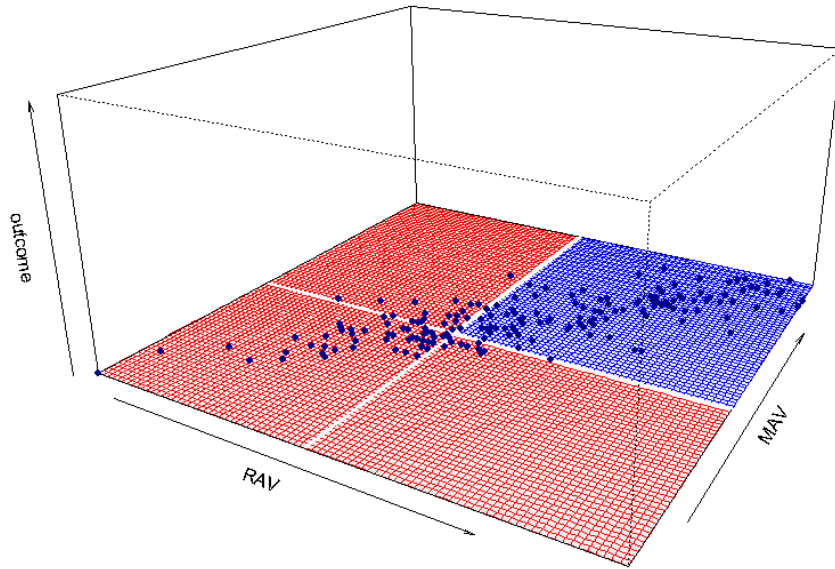Figure 1. Plot of schools' reading and math assignment scores.

Figure 2: Plot of schools' math and reading assignment scores with treatment effect.
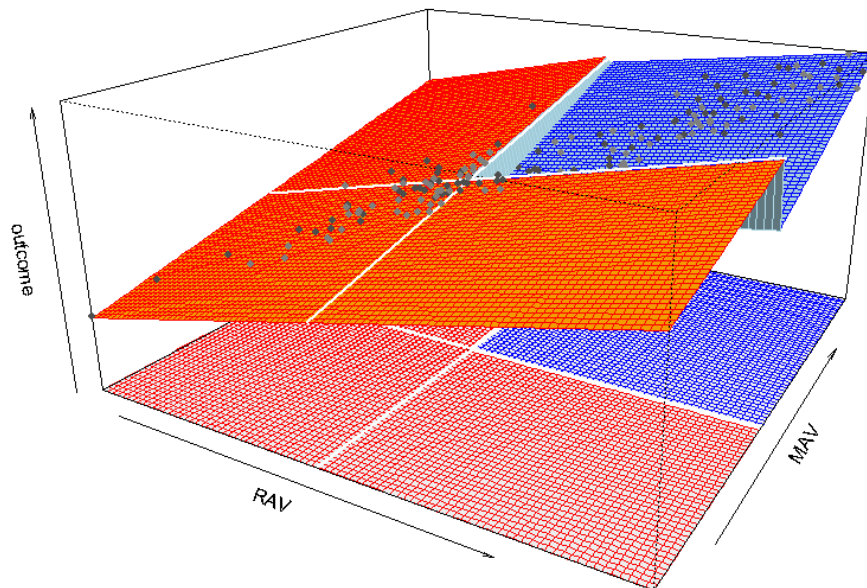
Figure 3: Scatterplot of uncentered reading assignment variable against gain scores.
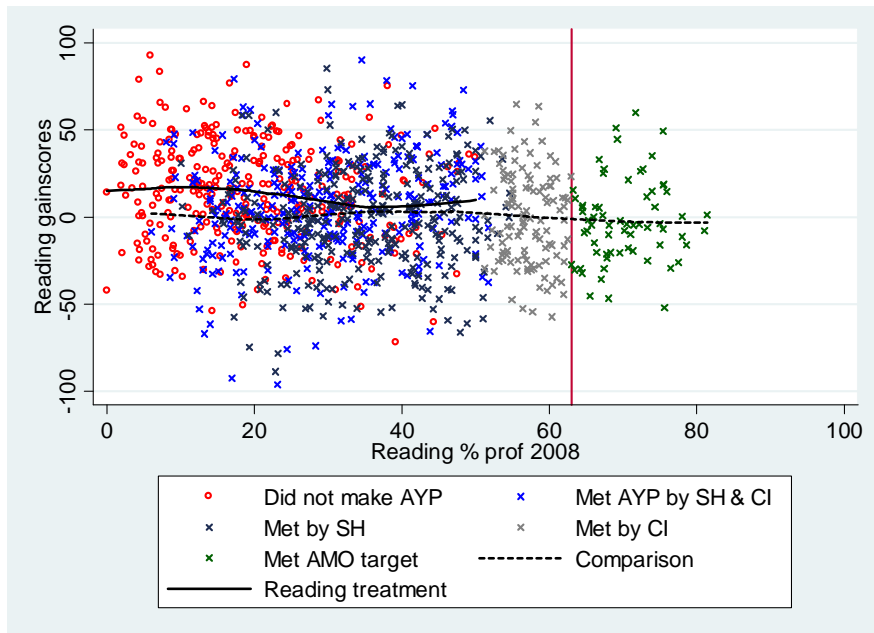

Figure 4: Scatter plot of centered assignment variable against gain scores.