Abstract Title Page

Not included in page count.

Title:

Analyzing Regression-Discontinuity Designs with Multiple Assignment Variables: A Comparative Study of Four Estimation Methods

Authors and Affiliations:

Vivian C. Wong
Northwestern University, Institute for Policy Research
Peter M. Steiner
University of Wisconsin, Madison, Department of Educational Psychology
Thomas D. Cook
Northwestern University, Institute for Policy Research

Abstract Body

Limit 4 pages single-spaced.

Background / Context:

In a traditional regression-discontinuity design (RDD), units are assigned to treatment and comparison conditions solely on the basis of a single cutoff score on a continuous assignment variable. The discontinuity in the functional form of the outcome at the cutoff represents the treatment effect, or the average treatment effect at the cutoff. However, units are often assigned to treatment on more than one continuous assignment variable. Recent applications of RD designs in education have had multiple assignment variables and cutoff scores available for treatment assignment. For example, Jacob and Lefgren (2004a) and Matsudaira (2008) examined the effects of summer remedial education programs that were assigned to students based on missing a reading score cutoff, a math cutoff or both. Kane (2003) and van der Klaauw (2002) evaluated the effects of college financial aid offers on students' post-secondary school attendance by using measures of income, assets and grade point average (Kane, 2003) or grade point average and SAT scores (van der Klaauw, 2002) as multiple assignment variables in an RD design. Papay, Murnane, and Willett (2010) and Martorell (2004) looked at the effects of failing high school exit exams in two subject areas – English language arts and math – on the probability of students' graduating from high school. Finally, Gill et al. (2007) examined the effects of schools' failure to make Adequate Yearly Progress (AYP) under No Child Left Behind by missing one of 39 possible assignment criteria. All are examples of the multivariate regression discontinuity design (MRDD), where treatment assignment is based on cutoffs for two or more covariates rather than a single point along an assignment variable. MRDDs are not unique to education; they also occur with increasing frequency in other fields of research, such as in the evaluation of labor market programs (Card, Chetty & Weber, 2007; Lalive, Van Ours & Zweimüller, 2006; Lalive, 2008).

Purpose / Objective / Research Question / Focus of Study:

This paper has three purposes. The first is to use potential outcomes notation (Holland, 1986; Rubin, 1974) to define the causal estimand τ_{MRD} for an MRDD with two assignment variables (*M* and *R*) and cutoffs. We show that the frontier average treatment effect τ_{MRD} may be decomposed into a weighted average of two univariate RDD effects, τ_M at the *M*-cutoff and τ_R at the *R*-cutoff. We introduce the term frontier average treatment effect to emphasize that the MRD design estimates treatment effects only for the sub-population of units located *at the cutoff frontier*, as opposed to the average treatment effect for the overall study population. This is analogous to the univariate RD design, where only the average treatment effect *at the cutoff* is estimated. In both cases, the average treatment effect of the study population may be inferred from the local estimates at the cutoff frontiers only when constant treatment effects can be reasonably assumed.

The second purpose of this paper is to provide guidance on the complexities of choosing an appropriate causal estimand of interest. Because each frontier produces a separate impact estimate, treatment effects may be reported individually (τ_M and τ_R) or pooled across multiple frontiers (τ_{MRD}). We show that in most cases, the frontier-specific effects will be the preferred causal estimand over the frontier average treatment effect τ_{MRD} because the latter is not scale-invariant. That is, τ_{MRD} depends crucially on the metric and scaling of the assignment variables. Estimating τ_{MRD} makes sense only if either the frontier-specific treatment effects are homogeneous or the assignment variables' metrics and scales are comparable. In this paper, we elaborate further on issues related to choosing an appropriate causal estimand in MRD designs, and highlight the contexts and conditions required for preferring frontier-specific effects over a pooled effect. Finally, the paper seeks to test four analytic approaches for estimating treatment effects in an MRD design – the frontier, centering, univariate, and instrumental variable (IV) approaches – and to identify the causal estimand(s) produced by each approach.

Significance / Novelty of study:

A regression-discontinuity design with multiple assignment variables raises challenges that are distinct from those identified in a traditional RD design. Treatment effects for an RD design with multiple assignment variables may be identified across multiple cutoff frontiers as opposed to a single point along the assignment variable. Thus, analytic procedures for estimating treatment effects across a multi-dimensional space are more complex and require more observations than approaches for estimating a treatment effect at a single point along the assignment variable. Although Cook et al. (2009), Reardon and Robinson (in press), and Papay, Willett, and Murnane (2011) outline various procedures for estimating treatment effects in an MRD design, the proposed approaches have not been derived formally, nor have they been tested empirically to examine their relative benefits and disadvantages.

Statistical, Measurement, or Econometric Model:

Unlike the traditional RD design, the multivariate regression-discontinuity design (MRDD) has an assignment process that is based on two or more assignment variables. In this paper, we consider only sharp MRDDs with two assignment variables, *R* and *M*, with respective cutoffs r_c and m_c. Units are assigned to treatment if they miss cutoff r_c, m_c, or both. Figure 1 shows that units are assigned to the control condition *C* if they score above both cutoffs ($R_i \geq r_c, M_i \geq m_c$) and to the treatment condition *T* if they score below either cutoff ($R_i < r_c$ or $M_i < m_c$). We partition the treatment assignment space into three subsets: T_1 if units miss only cutoff r_c, T_3 if they miss only cutoff m_c, and T_2 if they miss both cutoffs. Though we partition the treatment space into three subspaces, we assume that all cases receive exactly the same treatment (otherwise, more than one potential treatment outcome needs to be considered). In this design, *R* and *M* may be reading and math test scores (respectively), treatment may be a standardized test preparation course, and assignment to treatment may be based on whether students fail to achieve minimum threshold scores for reading or math. Although this is a fairly specific implementation of an MRDD, the results presented here also apply to MRDDs where treatment and control conditions are swapped. Figure 1 shows the cutoff frontier $F = \{(r,m):(r \geq r_c, m = m_c) \cup (r = r_c, m \geq m_c)\}$ at which the frontier average treatment effect is estimated. Assuming complete treatment compliance, the frontier average treatment effect τ_{MRD} is given by the expected difference in potential outcomes at the cutoff frontier: $\tau_{MRD} = E[Y_i(1) - Y_i(0) \mid (R_i, M_i) \in F]$.

Since the cutoff frontier consists of the *R*-frontier along assignment variable *M*, $F_R = \{(r,m):(r = r_c, m \geq m_c)\}$, and the *M*-frontier along assignment variable *R*, $F_M = \{(r,m):(r \geq r_c, m = m_c)\}$, we can decompose the frontier average treatment effect into a weighted average of the treatment effects at the *R*- and *M*-frontiers. Let the difference in potential outcomes be $G_i = Y_i(1) - Y_i(0)$ and the joint density function for assignment variables *R* and *M* be $f(r,m)$, then, we can define the treatment effect at the cutoff frontier *F* as the weighted average of conditional expectations given the single frontiers F_R and F_M:

$$\tau_{MRD} = E[G_i \mid (R_i, M_i) \in F] = w_R E[G_i \mid R_i \in F_R] + w_M E[G_i \mid M_i \in F_M]$$
$$= w_R \tau_R + w_M \tau_M \ ,$$

where weights w_R and w_M reflect the probabilities for observing a unit at the *R*- or *M*-frontier.

$$w_R = \frac{\int\limits_{m \geq m_c} f(r = r_c, m)\, dm}{\int\limits_{m \geq m_c} f(r = r_c, m)\, dm + \int\limits_{r \geq r_c} f(r, m = m_c)\, dr} \quad \text{and}$$

$$w_M = \frac{\int\limits_{r \geq r_c} f(r, m = m_c)\, dr}{\int\limits_{m \geq m_c} f(r = r_c, m)\, dm + \int\limits_{r \geq r_c} f(r, m = m_c)\, dr} .$$

The conditional expectations represent the treatment effects τ_R and τ_M at the two discontinuity frontiers F_R and F_M since

$$\tau_R = E[G_i \mid R_i \in F_R] = \frac{\int\limits_{m \geq m_c} g(r, m) f(r = r_c, m)\, dm}{\int\limits_{m \geq m_c} f(r = r_c, m)\, dm} \quad \text{and}$$

$$\tau_M = E[G_i \mid M_i \in F_M] = \frac{\int\limits_{r \geq r_c} g(r, m) f(r, m = m_c)\, dr}{\int\limits_{r \geq r_c} f(r, m = m_c)\, dr} ,$$

where $g(r,m) = y_1(r,m) - y_0(r,m)$ is the difference in potential outcomes. Note that τ_R is the

The decomposition of the frontier average treatment effect of an MRDD into a weighted average of unviariate RDD effects, τ_R and τ_M, reveals that the frontier average treatment effect τ_{MRD} depends on weights w_R and w_M. Since the weights are determined by integrating the joint density $f(r,m)$ along frontier *F,* their ratios depend crucially on the metric and scaling of assignment variables *R* and *M*. This is an unpleasant property of MRDD that is of special relevance whenever assignment variables are on a different metric or measurement scale and the treatment effects for frontiers F_M and F_R differ ($\tau_M \neq \tau_R$).

Usefulness / Applicability of Method:

An MRDD with two assignment variables allows the estimation of three different causal quantities: two frontier-specific effects, τ_R and τ_M, and the frontier average treatment effect τ_{MRD}. We will present the following four estimation procedures to estimate treatment effects: the frontier, centering, univariate, and instrumental variable approach. The frontier approach estimates treatment effects by first modeling the discontinuity at the cutoff frontier using parametric, semiparametric or nonparametric procedures, and then by applying appropriate treatment weights to each cutoff frontier to estimate τ . The approach estimates the frontier average treatment effect (τ_{MRD}) and frontier-specific effects (τ_M and τ_R) simultaneously. It is a more flexible extension of an approach introduced by Berk and de Leeuw (1999), which relied on parametric regression estimation of the entire response surface under the assumptions of constant treatment effects and a correctly specified regression model. The frontier approach we propose relaxes these assumptions by allowing for heterogeneous treatment effects along the cutoff frontier. Its limitation, however, is that it estimates the frontier average treatment effect, as opposed to the more general average treatment effect estimated by Berk and de Leeuw's method. In the centering approach, all assignment variables are centered at their respective cutoffs, and each unit is assigned its minimum centered assignment score. The minimum assignment score is used then as the single assignment variable in a traditional univariate RDD to estimate τ_{MRD}. This approach was employed by Gill et al. (2007) in their evaluation of No Child Left Behind. In the univariate approach, researchers choose a single assignment variable and cutoff to estimate a frontier-specific effect, and exclude all observations that are assigned to treatment via the second assignment variable and cutoff. Jacob and Lefgren (2004a) applied this approach in their evaluation of Chicago remedial education programs. Finally, in the IV approach, researchers use at least one assignment mechanism as an instrument for treatment receipt and designate units assigned by the second assignment variable and cutoff as treatment-misallocated cases. Cook et al. (2009) and Reardon and Robinson (in press) propose this approach for analyzing MRDDs, but it has yet to be examined empirically. For each approach, we discuss the causal quantities, theoretical underpinnings, and required assumptions. Through Monte Carlo simulations, we will examine the performance of the four approaches. Overall, we find that the frontier, centering, univariate, and IV approaches succeed in producing unbiased treatment effects when their design and analytic assumptions are met.

Conclusions:

Our analytic and simulation will work highlight the complexities of choosing an appropriate causal estimand in an MRD design. In many cases, the frontier average treatment effect may not have a meaningful interpretation because it does not make sense to pool effects across multiple frontiers. If at one frontier, the estimate indicates no effect and at the other frontier, a significant positive effect, then the average effect across the entire frontier rests on a scale-dependent weighting scheme. In these cases, we recommend that researchers estimate frontier-specific effects because τ_M and τ_R can provide at least upper and lower bounds for the overall treatment effect. In addition, without strong assumptions (e.g., constant treatment effects), the frontier-specific effects τ_M and τ_R is less general than what would be obtained from a traditional univariate RDD with a corresponding assignment variable and cutoff. That is because the cutoff of a traditional RDD is not restricted by the cutoffs of additional assignment variables (e.g., units with $M_i<m_c$ are excluded for estimating treatment effects at F_R). Still, the presence of multiple cutoff-frontiers has the advantage of exploring the heterogeneity of treatment effects along different dimensions. Finally, the frontier-specific and frontier average treatment effect cannot be generalized beyond the sub-population of units that is close to the cutoff frontiers. As with standard RDD, MRDD produces only the treatment effects along the cutoff frontier(s) as opposed to across the entire response surface. Thus, researchers have the onus of communicating to practitioners and policy-makers which causal quantities are evaluated, explaining why these are the causal quantities of interest, and discussing the benefits and limitations of the results.

Appendix A. References

References are to be in APA version 6 format.

References

Aiken, L. S., West, S. G., Schwalm, D. E., Carroll, J., & Hsuing, S. (1998). Comparison of a randomized and two quasi-experimental designs in a single outcome evaluation: Efficacy of a university-level remedial writing program. *Evaluation Review, 22*, 207-244.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. Journal *of the American Statistical Association, 87*, 328-336.

Berk, R., Barnes, G., Ahlman, L., & Kurtz (2010). When second best is good enough: A comparison between a true experiment and a regression discontinuity quasi-experiment. *Journal of Experimental Criminology, 6*(2), 191-208.

Berk, R. A., & de Leeuw, J. (1999). An evaluation of california's inmate classification system using a generalized regression discontinuity design. *Journal of the American Statistical Association, 94*(448), 1045-1052.

Berk, R. A., & Rauma, D. (1983). Capitalizing on nonrandom assignment to treatments: A regression-discontinuity evaluation of a crime-control program. *Journal of the American Statistical Association, 78*(381), 21-27.

Black, D., Galdo, J., & Smith, J. A. (2007). Evaluating the regression discontinuity design using experimental data. Retrieved from economics.uwo.ca/newsletter/misc/2009/smith_mar25.pdf

Buddelmeyer, H., & Skoufias, E. (2004). An evaluation of the performance of regression discontinuity design on PROGRESA. *World Bank Policy Research Working Paper No. 3386; IZA Discussion Paper No. 827*. Retrieved from http://ssrn.com/abstract=434600

Card, D., Chetty, R., & Weber, A. (2007). Cash-on-hand and competing models of intertemporal behavior: New evidence from the labor market. *Quarterly Journal of Economics, 122*(4), 1511-1560.

Cook, T. D., Wong, V. C., Steiner, P. M., Taylor, J., Gandhi, A., Kendziora, K., et al. (2009). *Impacts of School Improvement Status on Students with Disabilities: Feasibility report*. Washington, DC: American Institutes for Research.

Finkelstein, M., Levin, B., & Robbins, H. (1996). Clinical and prophylactic trials with assured new treatment for those at greater risk: II. Examples. *Journal of Public Health, 86*(5), 696-705.

Gill, B., Lockwood, J. R., Martorell, F., Setodji, C. M., & Booker, K. (2007). 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.

Goldberger, A. S. (2008). Selection bias in evaluating treatment effects: Some formal illustrations. In *Modelling and Evaluating Treatment Effects in Economics*, eds. T. Fomby, R. C. Hill, D. L. Millimet, J. A. Smith, E. J. Vytlacil, Amsterdam: JAI Press, pp. 1-31 .

Hahn, J., Todd, P., & van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica, 69*(1), 201-209.

Holland, P. W. (1986). Statistics and causal inference. *Journal of American Statistical Association, 81*(396), 945-960.

Imbens, G. W., & Lemieux, T. (2007). Regression discontinuity designs: A guide to practice. *Journal of Econometrics, 142*(2), 615-635.

Imbens, G., & Kalyanaraman, K. (2010). Optimal bandwidth choice for the regression discontinuity estimator. *Unpublished manuscript*.
Jackson, R., McCoy, A., Pistorino, C., Wilkinson, A., Burghardt, J., Clark, M., et al. (2007). *National evaluation of Early Reading First: Final report*. Washington, DC: U.S. Department of Education, Institute of Education Sciences.
Jacob, B., & Lefgren, L. (2004a). Remedial education and student achievement: A regression-discontinuity analysis. *Review of Economics and Statistics, LXXXVI*, 226-244.
Jacob, B., & Lefgren, L. (2004b). The impact of teacher training on student achievement: Quasi-experimental evidence from school reform efforts in Chicago. *Journal of Human Resources, 39*, 50-79.
Kane, T. J. (2003). A quasi-experimental estimate of the impact of financial aid on college-going. Working Paper 9703, National Bureau of Economic Research.
Lalive, R. (2008). Unemployment benefits, unemployment duration, and post-unemployment jobs: A regression discontinuity approach. *American Economic Review, 97*(2), 108-112.
Lalive, R., Van Ours, J. C., & Zweimüller, J. (2006). How changes in fiscal incentives affect the duration of unemployment. *Review of Economic Studies, 73*(4), 1009-1038.
Lee, D. S., & Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature, 48*(2), 281-355.
Martorell, P. (2004). Do high school graduation exams matter? A regression-discontinuity approach.
Matsudaira, J. (2008). Mandatory summer school and student achievement. *Journal of Econometrics, 142*(2), 829-850.
Papay, J. P., Murnane, R. J., & Willett, J. B. (2010). The consequences of high school exit examinations for low-performing urban students: Evidence from Massachusetts. *Educational Evaluation and Policy Analysis, 32*(1), 5-23.
Papay, J. P., Murnane, R. J., & Willett, J. B. (2011). Extending the regression-discontinuity approach to multiple assignment variables. *Journal of Econometrics, 161*, 203-207.
Reardon, S. F., & Robinson, J. P. (in press). Regression discontinuity designs with multiple rating-score variables. *Journal of Research on Educational Effectiveness.*
Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology, 66*, 688-701.
Shadish, W. S., Galindo, R., Wong, V. C., Steiner, P. M., & Cook, T. D. (in press). A Randomized Experiment Comparing Random to Cutoff-Based Assignment. *Psychological Methods*.
Trochim, W. M. K. (1984). *Research design for program evaluation*. Beverly Hills, CA: Sage Publications.
van der Klaauw, W. (2002). Estimating the effect of financial aid offers on college enrollment: A gegression-discontinuity approach. *International Economic Review, 43*(4), 1249-1287.
Wong, V. C., Cook, T. D., Barnett, W. S., & Jung, K. (2008). An effectiveness-based evaluation of five state pre-kindergarten programs. *Journal of Policy Analysis and Management, 27*(1), 122-154.

Appendix B. Tables and Figures

Not included in page count.

Figure 1. MRDD with two assignment variables R and M