Abstract Title Page
Not included in page count.

Title: An Empirical Comparison of Randomized Control Trials and Regression Discontinuity Estimations

Authors and Affiliations: Felipe Barrera-Osorio, Harvard Graduate School of Education. Deon Filmer, World Bank. Joe McIntyre, Harvard Graduate School of Education.

Abstract Body

Limit 4 pages single-spaced.

Background / Context:

Description of prior research and its intellectual context.

Randomized controlled trials (RCTs) and regression discontinuity (RD) studies both provide estimates of causal effects. A major difference between the two is that RD only estimates local average treatment effects (LATE) near the cutoff point of the forcing variable. This has been cited as a drawback to RD designs (Cook & Wong, 2008). Comparisons of RCT estimates of average treatment effect (ATE) and RD estimates of LATE are rare because few studies have both randomized assignment and a forcing variable. Cook, Shadish, and Wong (2008) and Cook and Wong identify only three studies that allow the comparison of RCT and RD estimates, namely studies by Aiken (1998), Skoufias (2004), and Black and Smith (2005), though Shadish, Galindo, Wong, Steiner, and Cook (2011), Green, Leong, Kern, Gerber, and Larimer (2009), and Gleason, Resch, and Berk (2012) provide more recent examples. Wing and Cook (2011) also provide general advice for using control groups to improve RD estimates. In this study, we extend this line of research, comparing RCT and RD estimates of the effect of a scholarship on Cambodian schoolchildren.

Purpose / Objective / Research Question / Focus of Study:

Description of the focus of the research.

Using a study which is amenable to analysis both as an RCT and an RD, we compare the RCT estimate of ATE to the RD estimate of LATE. We also explore how the RD estimate of average treatment effect depends on econometric specifications, including for parametric and non-parametric estimators, and for various bandwidths. Finally, we use Monte Carlo simulation based on observed data to explore the performance of RD estimators with a forcing variable which is standardized within schools by choosing a different cutoff in each school.

Setting:

Description of the research location.
(May not be applicable for Methods submissions)

The intervention was carried out in 204 schools in three provinces in Cambodia.

Population / Participants / Subjects:

Description of the participants in the study: who, how many, key features, or characteristics.
(May not be applicable for Methods submissions)

The intervention targeted 4th-graders in schools in Cambodia. After the first year, 4th-graders (who were 3rd-graders when the program began) in both treatment and control schools were eligible to receive scholarships.

Intervention / Program / Practice:

Description of the intervention, program, or practice, including details of administration and duration.
(May not be applicable for Methods submissions)

Schools were randomly assigned to treatment or control. Treatment schools were randomly assigned to receive either needs-based or merit-based scholarships. In schools assigned to needs-based scholarships, students scoring at or above the school median on a poverty index, where high scores indicated poorer students, received scholarships. In schools assigned to merit-based scholarships, students scoring at or above the school median on a test of math and language skills received scholarships. Thus, we created a forcing variable for all students in the treatment schools equal to the difference between the student's merit or poverty score and the school median. All scholarships were contingent on students maintaining a certain level of attendance and grades. Scholarships were around $20 per year, representing around 3% of the yearly per-capita income of the country.

Significance / Novelty of study:
Description of what is missing in previous work and the contribution the study makes.

This study has two novel features. First, it allows us to compare RCT estimates of ATE to RD estimates of LATE. In principle, the comparison will allow to determine the extent of locality of the RD estimator. Additionally, the RCT design allows us to estimate the LATE using a counterfactual population of students in control schools who would have otherwise been near the cutoff point on the forcing variable. Taking this as the true LATE, we estimate which econometric specifications of the RD design (e.g. parametric or non-parametric RD, which bandwidth) most effectively estimates the LATE.

Statistical, Measurement, or Econometric Model:
Description of the proposed new methods or novel applications of existing methods.

We compare estimates of ATE from an RCT to estimates of LATE from an RD. We also use a counterfactual of students in control schools who would have been near the cutoff point had they been in treatment schools to obtain a "true" value of the LATE, then use this to evaluate the performance of different econometric specifications of the RD. Finally, using Monte Carlo methods, we simulate possible treatment effects for students in control schools who would have been assigned to treatment (i.e. with high baseline merit or poverty scores) and see how well RD estimators are able to recover the true treatment effects.

Usefulness / Applicability of Method:
Demonstration of the usefulness of the proposed methods using hypothetical or real data.

RD is a common technique for estimating causal effects. Using RD, however, requires the analyst to make a set of decisions as to how the RD will be estimated. Some common choices involve selecting a non-parametric or a parametric RD; determining the degree of the polynomial used to estimate a parametric RD; and selecting the bandwidth to use to estimate the LATE. Our study provides some guidance as to the optimal specification by using an RCT to estimate the "true" LATE. Additionally, although RD can only estimate the LATE of an intervention near the cutoff point of a forcing variable, the ATE of the intervention is often more important to policymakers. This study shows how closely the LATE approximates the ATE in this data, contributing to the literature on the relationships between these two quantities. Finally, we demonstrate potential complications associated with applying RD methods to real data.

Research Design:
Description of the research design.
(May not be applicable for Methods submissions)

Data Collection and Analysis:
Description of the methods for collecting and analyzing data.
(May not be applicable for Methods submissions)

Two years after carrying out the intervention, a subset of students from both control and treatment schools were interviewed at home by trained interviewers. The interviewers determined how many years the children had stayed in school after the program began. They also administered math tests and tests of working memory. These were the outcomes of interest.

Findings / Results:
Description of the main findings with specific details.
(May not be applicable for Methods submissions)

We find that RCT estimates of the ATE and RD estimates of the LATE are very close for the merit-based treatment, although the RD estimates tend to be less precise than their RCT counterparts (see Table 1). For the needs-based treatment, the situation is slightly different: non-parametric the RD estimate of the LATE of the intervention on math scores is very similar to the RCT estimate of the effect near the cut-off point, but the parametric RD estimate is too high. At the same time the non-parametric RD estimate of the effect of needs-based treatment on grades level completed is much lower than the RCT estimate, but the parametric estimate is very close to the true value.
At the same time, we find that non-parametric estimates of treatment effect are relatively robust to econometric specification (see Table 2). Similar results are obtained using the IK (Imbens and Kalyanaraman, 2012) or the CCT (Calonico, Cattaneo & Titiunik, 2012) bandwidth, and using a triangular or rectangular kernel. However, estimates can be highly sensitive to unusual points or too-narrow bandwidths. For parametric RD, the results are quite sensitive to the functional form of the RD. In particular, the addition of an interaction to allow for different slopes above and below the cutoff point results in dramatically different estimates of treatment effect. Unlike non-parametric estimates, parametric estimates are also extremely sensitive to the choice of bandwidth.
Finally, we find evidence that RD can provide badly biased estimates of treatment effects with forcing variables which are nested within larger units.

Conclusions:
Description of conclusions, recommendations, and limitations based on findings.
Our findings can be used to inform decisions about which estimators to use in doing regression discontinuity, as well as providing evidence of how well RD estimators approximate RCT estimators.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

Aiken, L. S., West, S. G., Schwalm, D. E., Carroll, J. L. & Hsiung, 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(2)*, 207-244.

Black, D., Galdo, J. & Smith, J. A. (2005). Evaluating the regression discontinuity design using experimental data. Unpublished manuscript .

Calonico, S., Cattaneo, M. D. & Titiunik, R. (2012), Robust data-driven inference in the regression-discontinuity design. *Stata Journal 55*(2), 1-29.

Cook, T. D., Shadish, W. R. & Wong, V. C. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management 27*(4), 724{750.

Cook, T. D. & Wong, V. C. (2008). Empirical tests of the validity of the regression discontinuity Design. *Annales d'Economie et de Statistique* pp. 127-150.

Gleason, P. M., Resch, A. M., & Berk, J. A. (2012). Replicating Experimental Impact Estimates Using a Regression Discontinuity Approach.

Green, D. P., Leong, T. Y., Kern, H. L., Gerber, A. S., & Larimer, C. W. (2009). Testing the accuracy of regression discontinuity analysis using experimental benchmarks. *Political Analysis*, *17*(4), 400-417.

Imbens, G. & Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity Estimator. *The Review of Economic Studies 79(3)*, 933{959.

Shadish, W., Galindo, R., Wong, V., Steiner, P., & Cook, T. (2011). A randomized experiment comparing random and cut-off based assignment. *Psychological Methods 16*(2), 179-191.

Wing, C., Cook, T. D., & Society for Research on Educational Effectiveness. (2011). How Can Comparison Groups Strengthen Regression Discontinuity Designs. *Society For Research On Educational Effectiveness*

Appendix B. Tables and Figures

Not included in page count.

Table 1: Basic Comparison of RCT and RD results

	Merit-based scholarship				Poverty-based scholarship			
	RCT		RD		RCT		RD	
	ATE	ATE, narrow	Non-parametric	Parametric	ATE	ATE, narrow	Non-parametric	Parametric
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Grades completed								
β	0.22*	0.34*	0.25	0.40*	0.36**	0.30*	0.03	0.24*
	(0.11)	(0.14)	(0.25)	(0.16)	(0.11)	(0.15)	(0.14)	(0.12)
N	897	436	744	455	831	446	800	459
Math score								
β	0.15	0.16	0.20	0.33***	0.00	-0.02	0.00	0.15*
	(0.10)	(0.12)	(0.20)	(0.08)	(0.08)	(0.12)	(0.12)	(0.08)
N	940	453	780	471	883	472	853	484

Note: ATE narrow uses close to 50% of the sample to estimate the effects; Non-parametric: triangular kernel, I&K optimal bandwidth; Parametric: linear control of the forcing variable, bandwidth close to 15%

Table 2: Non-parametric RD results

	Merit-based scholarship				Poverty-based scholarship			
	IK bandwidth		CCT bandwidth		IK bandwidth		CCT bandwidth	
	Triangular (1)	Uniform (2)	Triangular (3)	Uniform (4)	Triangular (5)	Uniform (6)	Triangular (7)	Uniform (8)
Grades completed								
β	0.25	0.30	-0.41	-0.41	0.03	0.04	0.06	0.04
	(0.25)	(0.26)	(0.32)	(0.31)	(0.14)	(0.15)	(0.17)	(0.17)
N	694	628	467	467	800	800	567	487
Math score								
β	0.20	0.18	0.27	0.26*	0.00	0.02	0.00	-0.03
	(0.20)	(0.22)	(0.21)	(0.11)	(0.12)	(0.13)	(0.15)	(0.16)
N	727	659	491	368	765	693	566	468