Characterizing Cross-Site Variation in Local Average Treatment Effects in Multisite RDD contexts with an Application to Massachusetts High School Exit Exams

Sophie Litschwartz & Luke Miratrix

Background/Context:

Regression discontinuity designs (RDD) are frequently used in education research. Still, one common concern with RDD studies is how generalizable results are to populations far from the cut point. Like with RCT's one potential way to get at the generalizability of RDD results is to estimate treatment effect variation within multi-site RDD's (Bloom, 2012). In addition, RDD is a quasi-experimental design and so is often done in circumstances that are not tightly controlled by the researcher which we would expect to lead to larger cross-site variance (Weiss et al., 2017).Yet there is currently no established method for estimating such cross-site treatment effect variation.

One case where estimating RDD cross-site treatment is relevant is in Massachusetts, where the state is currently in the process of reviewing policies for their high school exit exam (known locally as Competency Determination policies). To graduate, students must score at least the minimum score to be designated "Needs Improvement", on the 10th grade MCAS exams in ELA, Math, and Science. Massachusetts also requires students classified as "Needs Improvement" but not "Proficient" in Math or ELA to complete an Education Proficiency Plan (EPP) to demonstrate proficiency in their non-proficient subject(s) by the time they graduate. Individual high schools across Massachusetts have considerable latitude in how they implement EPP's for their students. High schools can require students to demonstrate proficiency by taking a special proficiency exam, passing courses in the relevant area(s) in their junior and senior year, or a combination of the two. High schools have particularly large amounts of discretion with regards to the math EPP because state-wide all high schools require four years of ELA for all students but many high schools only have a baseline requirement of either two or three years of math. One question Massachusetts seeks to understand, as it reviews its Competency Determination policies, is whether students who are bound by the math EPP are completing more math courses in their senior year as the policy intends. In addition, and given how the EPP policy is implemented, Massachusetts would also like to know how the much the impact of the EPP on math completion varies by high school.

Research Design:

In this paper, we develop a method for estimating cross-site variation in the context of regression discontinuity design. We combine the fixed intercept/random coefficients (FIRC) model (Bloom et al., 2017) typically used in the context of multi-side RCT's with a standard local linear RDD analysis. We estimate a model with an indicator for each school, pooled coefficients on the running variable and the treatment by running variable interaction, and a partially pooled coefficient on the treatment variable. The regression model is therefore estimated as:

Level One -Student:

$$Y_{ij} = \alpha_j + \beta_{1j} T_{ij} + \beta_2 Score_{ij} + \beta_3 T_{ij} * Score_{ij} + \epsilon_{ij}$$

$$\epsilon_{ij} \overset{iid}{\sim} N[0, \sigma_y^2]$$

Level Two - School:

$$\beta_{1j} = \delta + e_{1j}$$

$$e_{1j} \overset{iid}{\sim} N[0, \sigma^2_{\beta_1}]$$

Y_{ij}: Indicator for whether student i in school j completed a math course two years after taking the MCAS

T_{ij}: Treatment indicator, 1 if student i in school j scores below 240

$Score_{ij}$: Raw 10th grade math MCAS score for student i in school j

δ: Local average treatment effect at the cutpoint

We subset the data to all units such that $|Score_{ij} - cut| < h_{opt}$, where h_{opt} is selected using cross-validation, which allows for a more flexible model and a discrete running variable (Imbens and Lemiuex, 2008).

This model allows us to answer not just the effect of the math EPP on completion of a math course in a student's senior year, but how much that effect varies by site. A model that simply runs a separate local linear regression by site and takes the variance of the treatment effect estimates will overestimate the treatment effect variance because measurement error in the treatment effect estimation will be confounded with true variance. The RDD FIRC model proposed in this study accounts for measurement error in the variance estimate and therefore produces a more accurate estimate of cross-site impact variance. Other models are also possible (e.g. different local trends could be fit for each site either as a fixed effect or random effect) we will conduct sensitivity checks to compare how the RDD FIRC model performs against these other models.

Population/Participants/Subjects:

Data for this study come from the Massachusetts Department of Elementary and Secondary Education (MA DESE). The total sample for this study is 66,492 10th grade students in 510 schools who took the 10th grade math MCAS exam for the first time in 2009, with a median of 90 10th graders per school. The optimal bandwidth was estimated as 10 raw score points on either side of the cut-score. Limiting the sample to just students within the optimal bandwidth reduces the sample to 23,277 students in 411 schools with a median of 13 treated and 20 not treated students in the bandwidth per school.

Findings/Results:

First we show that, in simulations, the RDD FIRC model has good coverage properties when there are at least 100 sites (Table One). In the Massachusetts example, we find that students on the margin required to complete a math EPP were 6.29 (standard error = 1.00) percentage points more likely to complete a math class two years after taking the MCAS exam than students not required to complete an EPP. This is compared to a baseline where 76% of students in the bandwidth complete a math course in their senior year. As expected, there is significant variation in the treatment effect. The standard deviation of the treatment effect is 8.69 percentage points. This means that while overall there was a positive effect of the math EPP on the probability of completing a math course, we estimate that in more than a third of schools the math EPP actually had a negative effect on the probability of completing a math course in a student's senior year. The next steps for this research is to explore how the RDD FIRC model compares to Empirical Bayes estimates of the treatment effect variation and whether site-level characteristics (e.g. proficiency level of the students, amount of math required to graduate) predict cross-site treatment effect variation.

Table 1: RDD FIRC Model Coverage

Total Schools	Average School Size	Coverage Rate $LATE$	Coverage Rate σ^2_{LATE}
10	10	0.75	0.64
10	50	0.89	0.82
10	130	0.89	0.84
10	150	0.89	0.85
10	200	0.89	0.85
100	10	0.94	0.90
100	50	0.95	0.94
100	130	0.94	0.92
100	150	0.95	0.95
100	200	0.94	0.95
300	10	0.94	0.93
300	50	0.95	0.94
300	130	0.95	0.94
300	150	0.95	0.95
300	200	0.94	0.96
505	10	0.94	0.93
505	50	0.94	0.95
505	130	0.94	0.95
505	150	0.94	0.95
505	200	0.94	0.95

References

Bloom, Howard S. "Modern regression discontinuity analysis." Journal of Research on Educational Effectiveness 5.1 (2012): 43-82.

Bloom, Howard S., et al. "Using multisite experiments to study cross-site variation in treatment effects: A hybrid approach with fixed intercepts and a random treatment coefficient." Journal of Research on Educational Effectiveness 10.4 (2017): 817-842.

Imbens, Guido W., and Thomas Lemieux. "Regression discontinuity designs: A guide to practice." Journal of econometrics 142.2 (2008): 615-635.

Weiss, Michael J., et al. "How much do the effects of education and training programs vary across sites? Evidence from past multisite randomized trials." Journal of Research on Educational Effectiveness 10.4 (2017): 843-876.