AFRE 835: Introductory Econometrics

Chapter 18: Regression Discontinuity Designs

Spring 2017

Regression Discontinuity (RD)

- The *treatment effects* literature seeks to measure the causal impact of a program or policy

  . . . e.g., a job training program, an enducation program, a electricity rate structure, etc.
- The fundamental problem is that exposure to the treatment is often an endogenous variable.

  . . . individual choose to participate in a job training program or attend summer school or choose a time-of-use electric rate option.
- Controlling for the selection process can be challenging in a non-experimental setting.
- **Regression Discontinuity (RD)** designs take advantage of the fact that, for some treatments, access to the treatment is a discontinuous function of one or more *forcing* variables.

Examples

- Thistlethwaite and Campbell (1960) studied the effect of student scholarship on career aspirations, where scholarships were awarded only above a specific test score threshold;
- Angrist and Lavy (1999) studied effect of class size on student test scores, using the "Maimonides Rule" requiring classes to be split when they reached a given threshold;
- Van der Klaauw (2003) studied effect of financial aid offers on college attendance, using rule that relates aid to student SAT scores and GPA;
- Hahn *et al.* (1999) studied impact of anti-discrimination law, using the fact that it only applied for firms with at least 15 employees;
- Matsudaira (2007) studied the effect of a remedial summer school program, mandatory for students with a test score below a given level;
- Card *et al.* (2004) studied effect of medical services, where its availability is restricted by age.

Regression Discontinuity, Illustration, van der Klaauw (2003)

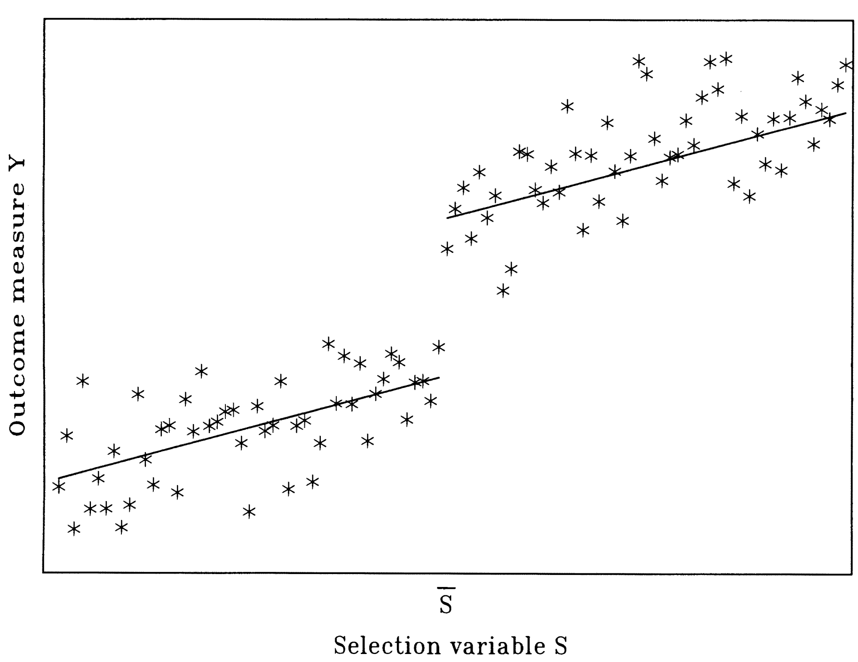
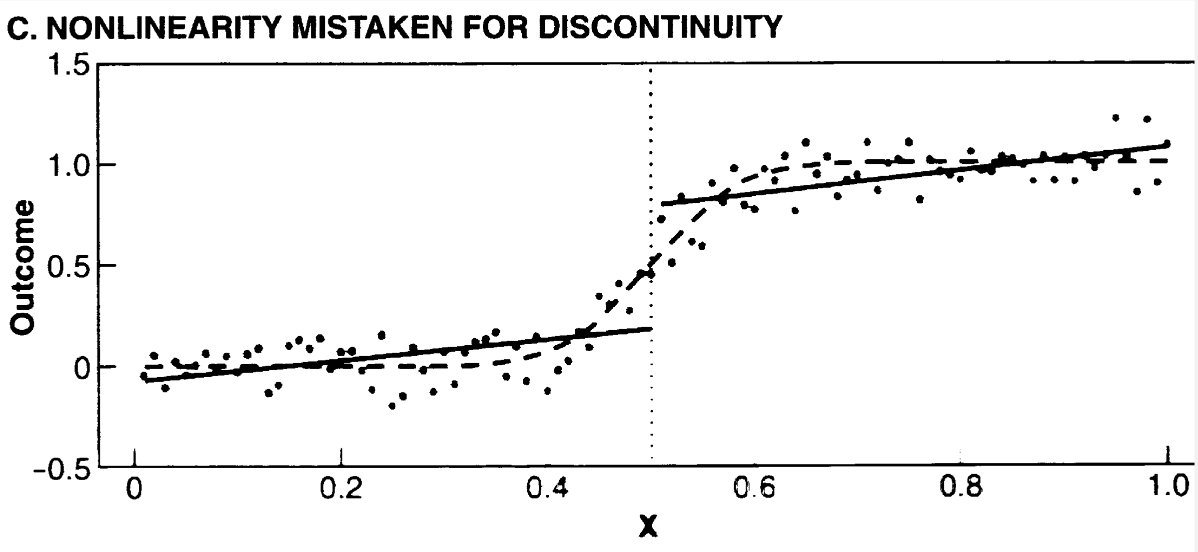


FIGURE 1

The Concern (AP p. 254)



A Sharp Regression Discontinuity Design

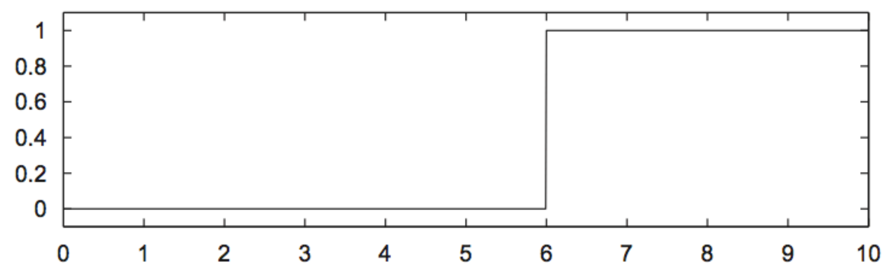


Fig. 1. Assignment probabilities (SRD).

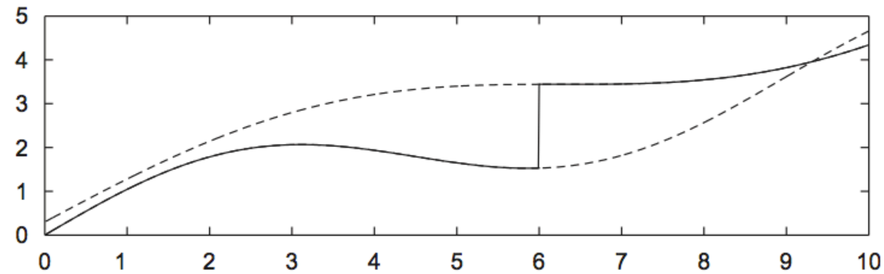


Fig. 2. Potential and observed outcome regression functions.

A Fuzzy Regression Discontinuity Designs

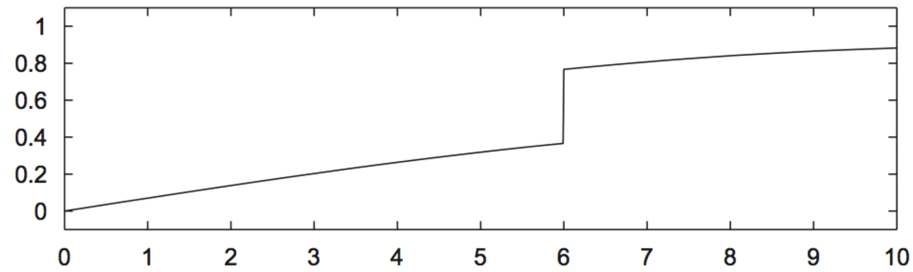


Fig. 3. Assignment probabilities (FRD).

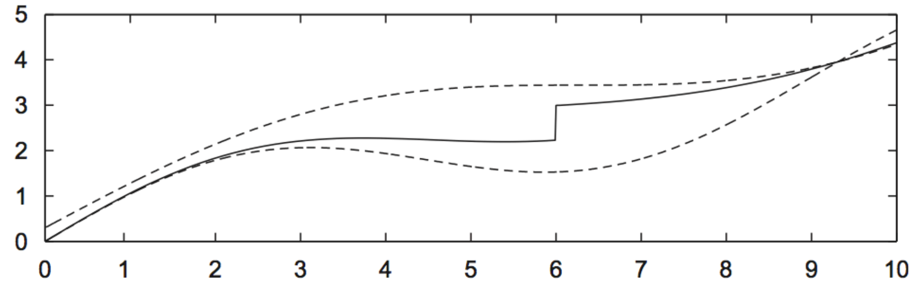


Fig. 4. Potential and observed outcome regression (FRD).

Regression Discontinuity Illustrations

1. Van der Klaauw (2003), "Estimating the effect of financial aid offers on college enrollment: a regression-discontinuity approach," *International Economic Review* **43**: 12491287.
2. Matsudaira. J. D. (2007) "Mandatory Summer School and Student Achievement," *Journal of Econometrics* 142:829-850.
3. Black, S. E. (1999), "Do Better Schools Matter? Parental Valuation of Elementary Education," *Quarterly Journal of Economics*, 114(2): 577-599.

Paper #1: Van der Klaaux (2003)

$S = \phi_0 \times$(first three digits of total SAT score) $+ \phi_1 \times$ GPA

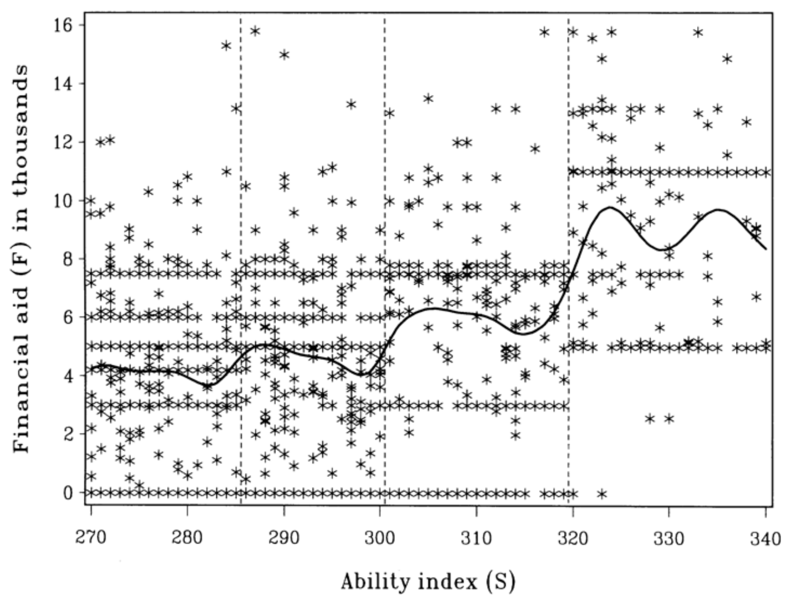


FIGURE 3

FINANCIAL AID OFFERS—FILERS. RAW DATA AND SPLINE SMOOTH (SOLID CURVE)

Paper #1: Van der Klaauw, (cont'd)

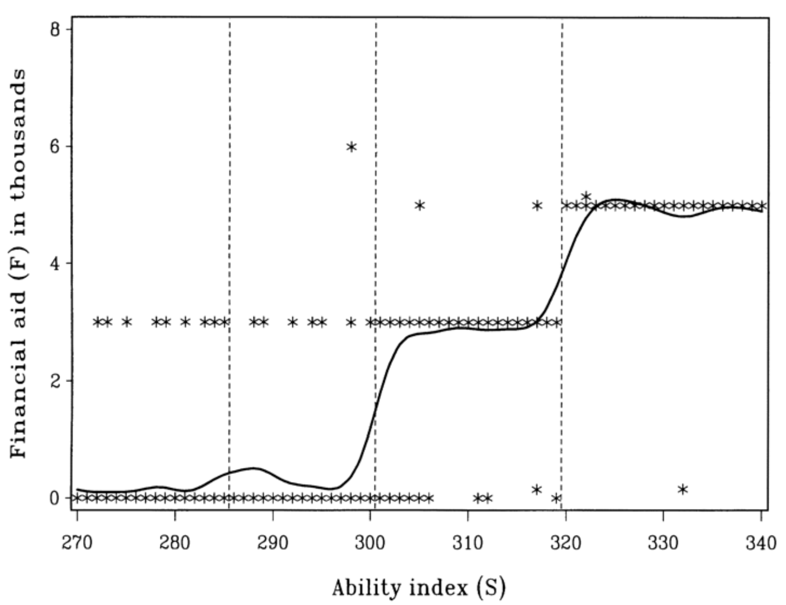


FIGURE 4

FINANCIAL AID OFFERS—NONFILERS. RAW DATA AND SPLINE SMOOTH (SOLID CURVE)

Paper #1: Van der Klaauw, (cont'd)

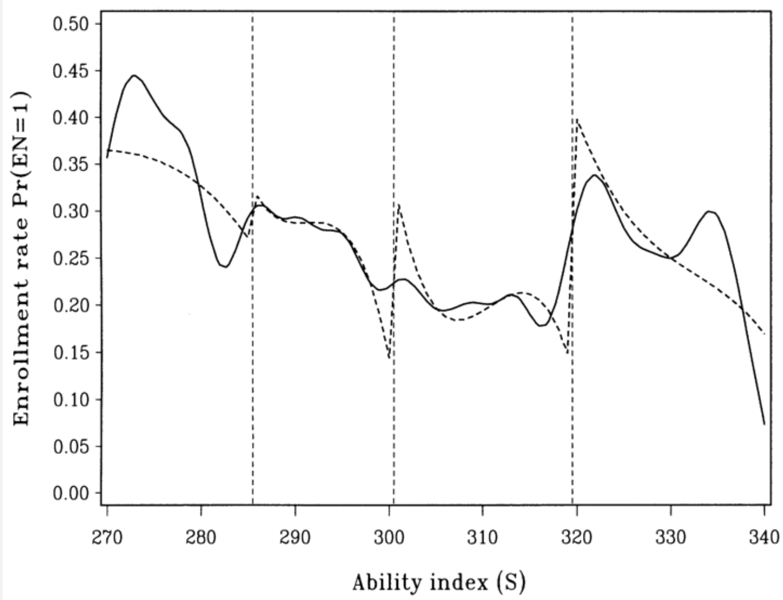


FIGURE 7

ENROLLMENT PROBABILITY—FILERS. PIECEWISE CUBIC REGRESSION (DASHED CURVE) AND NONPARAMETRIC SPLINE SMOOTH (SOLID CURVE)

Paper #1: Van der Klaauw, (cont'd)

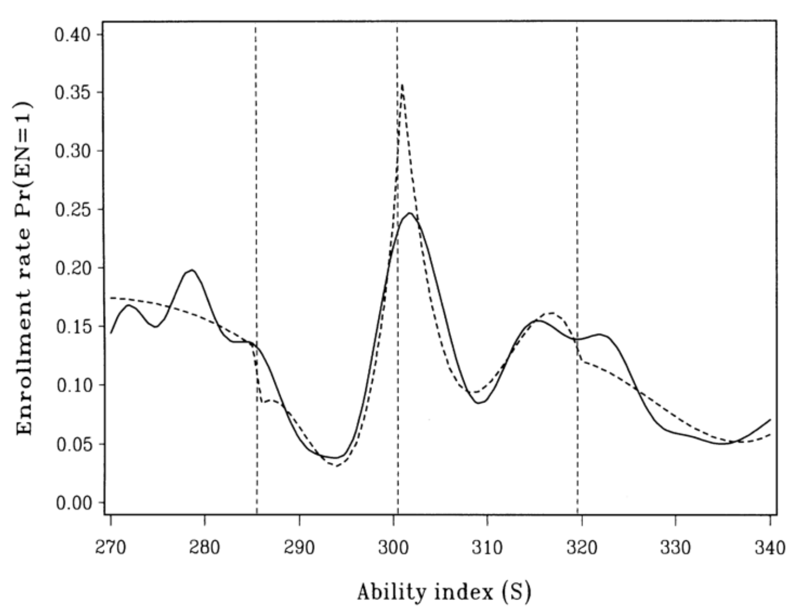


FIGURE 8

ENROLLMENT PROBABILITY—NONFILERS. PIECEWISE CUBIC REGRESSION (DASHED CURVE) AND NONPARAMETRIC SPLINE SMOOTH (SOLID CURVE)

Paper #1: Van der Klaauw, (cont'd)

TABLE 3

LOCAL WALD ESTIMATES OF FINANCIAL AID EFFECT

	Filers			Nonfilers		
	Estimate	Std Error	Obs	Estimate	Std Error	Obs
3-point intervals						
$\bar{S}_1$	0.010	(0.238)	171	0.524	(4.656)	77
$\bar{S}_2$	0.040	(0.041)	169	0.036	(4.895)	61
$\bar{S}_3$	0.067	(0.029)	107	−0.030	(0.052)	32
Pooled intervals	0.049	(0.021)	447	0.015	(0.038)	170
2-point intervals						
$\bar{S}_1$	0.052	(5.034)	109	0.076	(0.307)	53
$\bar{S}_2$	0.075	(0.049)	120	0.060	(0.052)	42
$\bar{S}_3$	0.107	(0.073)	64	−0.043	(0.045)	18
Pooled intervals	0.088	(0.028)	293	0.033	(0.040)	113

NOTES: The aid amount F is measured in thousands of 1991 dollars. Bootstrap standard errors are in parentheses. They were calculated with 10,000 bootstrap samples.

Paper #2: Mandatory Summer School, Matsudaira (2008)

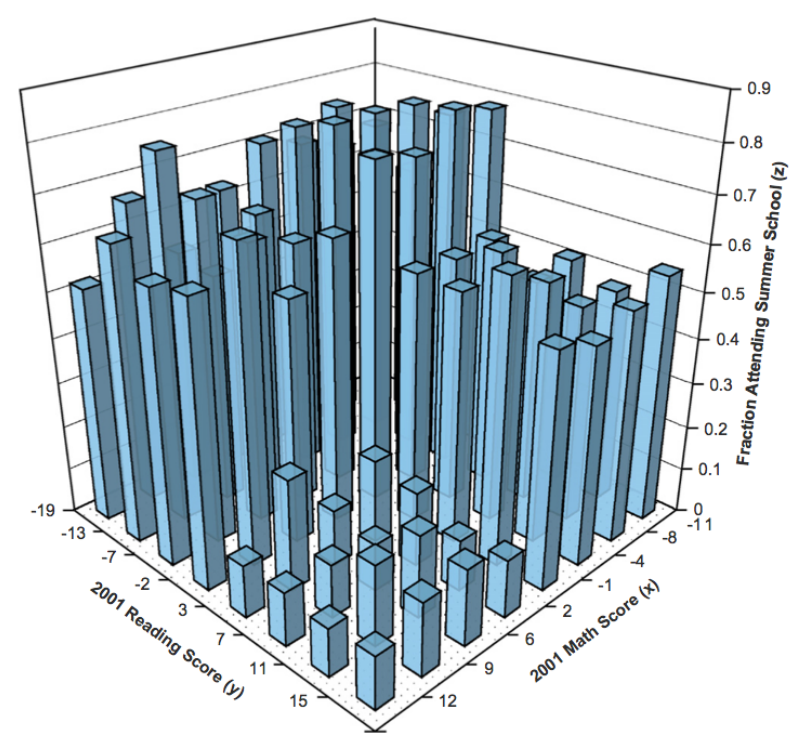
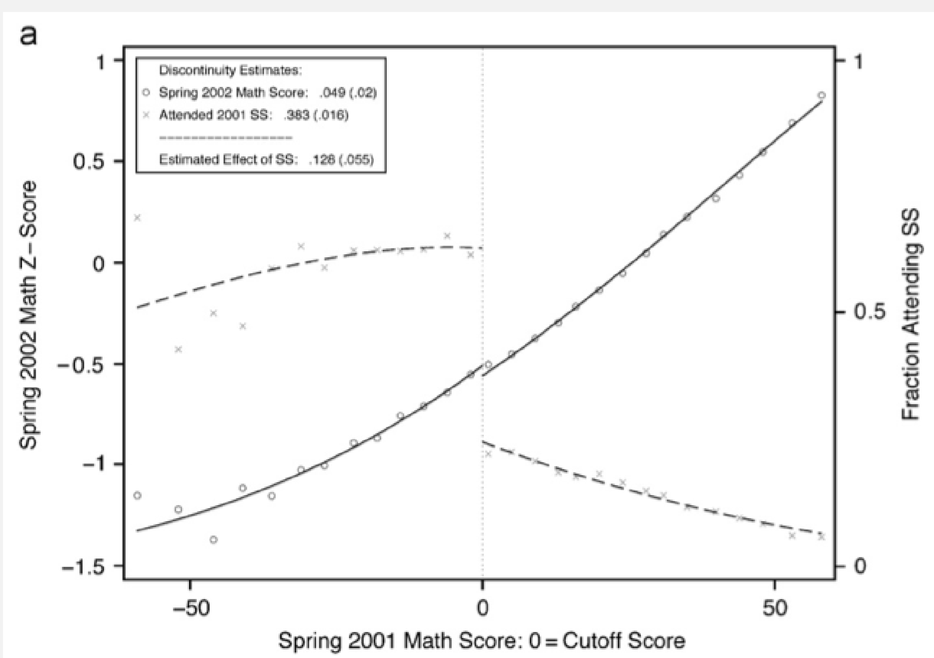


Fig. 1. Fraction attending summer school by 2001 math and reading scores: Grade 5.

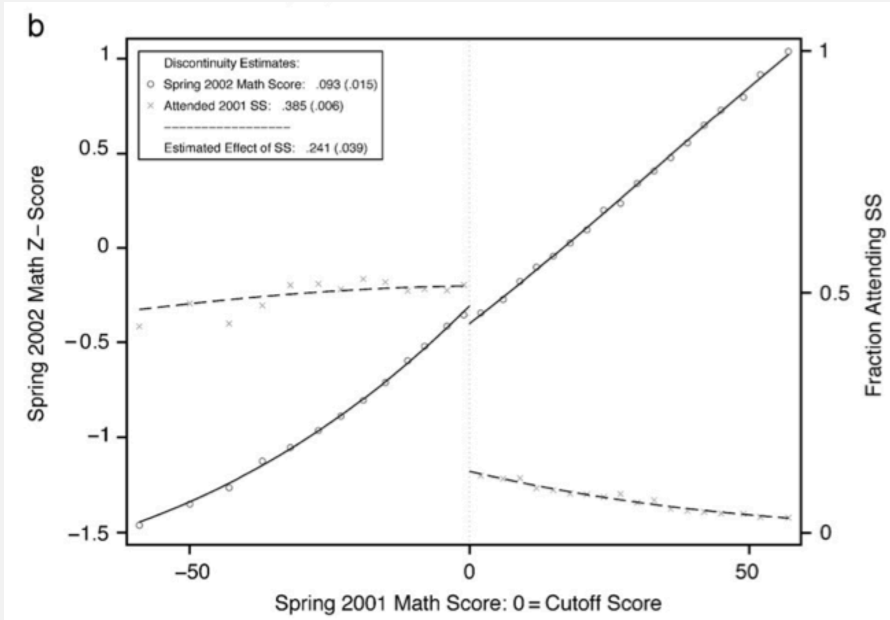
The Problem

- Matsudaira suggests considering that the treatment effect relationship of interest is
$$Y_i = \alpha + T_i\theta + v_i \tag{1}$$
where
  - Y_i denotes the spring 2002 test score outcome;
  - T_i denotes the indicator function =1 if the student attended summer school in 2001
  - v_i represents all other determinants of the spring 2002 test score.
- The problem is that T_i is likely correlated with v_i.
- The solution Matsudaira employs is to use D_i (an indicator for being *mandated* to attend summer school) as an instrument for summer school attendance.

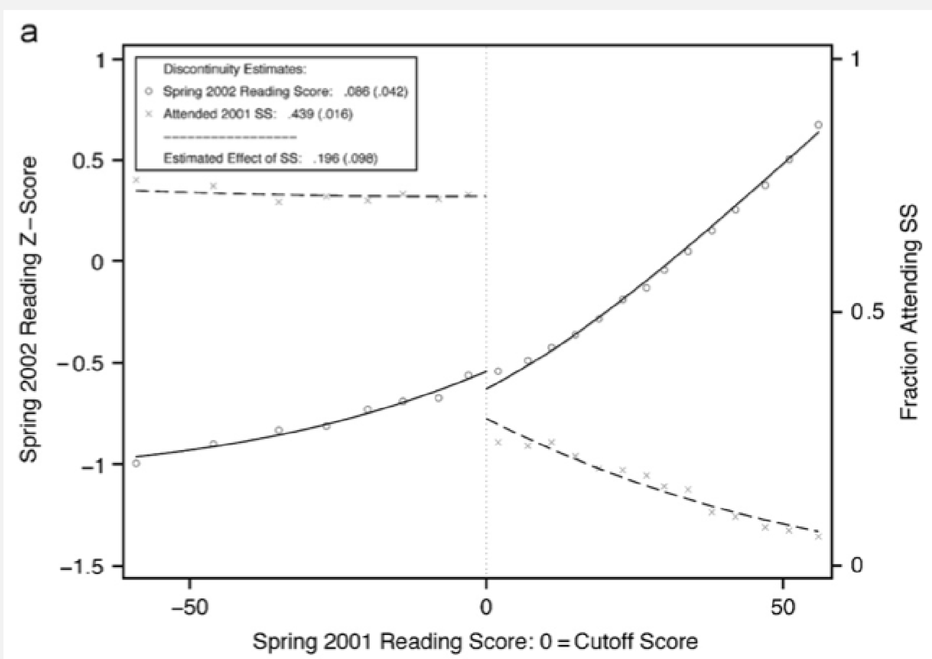
Grade 3 - Math



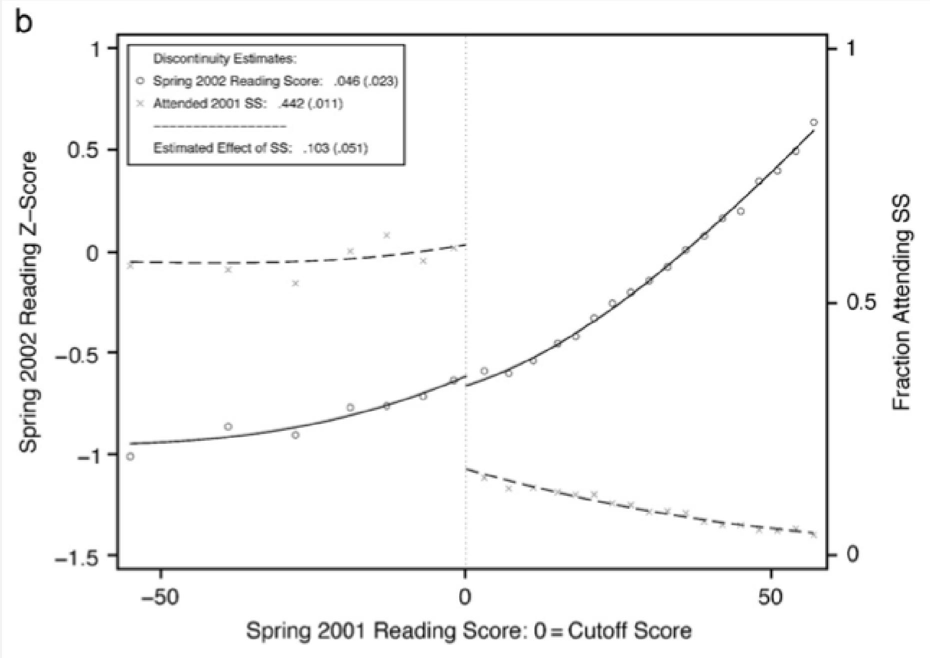
Grade 5 - Math



Grade 3 - Reading



Grade 5 - Reading



Math Driven Summer School Attendence

Table 2
The effect of being mandated to summer school on attendance and 2002 math scores and the effect of summer school attendance for ma
(students who passed the 2001 reading exam)

	Effect of being mandated		Effect of SS attendance		No. observatio
	Attendance (1st Stage)	Math (Reduced form)	Math (TSLS)	Reading (TSLS)	
Strong 1st stage discontinuity					
Grade 3	.383	.049	.128	.087	55,931
	(.016)	(.02)	(.055)	(.065)	
Grade 5	.385	.093	.241	.083	59,258
	(.006)	(.015)	(.039)	(.055)	
Grade 6	.320	.061	.19	n.a.	51,810
	(.011)	(.014)	(.047)	(–)	
Weak 1st stage discontinuity					
Grade 4	.000	.004	10.258	57.724	58,689
	(.015)	(.031)	(373)	(2111)	
Grade 7	.108	.051	.474	.303	48,199
	(.046)	(.023)	(.216)	(.215)	

Reading Driven Summer School Attendence

Table 3
The effect of being mandated to summer school on attendance and 2002 reading scores and the effect of summer school attendance f
reading (students who passed the 2001 math exam)

	Effect of being mandated		Effect of SS attendance		No. observatio
	Attendance (1st Stage)	Reading (Reduced form)	Reading (TSLS)	Math (TSLS)	
Strong 1st stage discontinuity					
Grade 3	.439	.086	.196	.201	55,385
	(.016)	(.042)	(.098)	(.068)	
Grade 4	.303	.052	.173	.109	60,052
	(.014)	(.024)	(.08)	(.062)	
Grade 5	.442	.046	.103	.095	47,484
	(.011)	(.023)	(.051)	(.049)	
Weak 1st stage discontinuity					
Grade 7	.127	−.013	−.103	−.364	36,243
	(.078)	(.034)	(.290)	(.423)	

Some Recommendations for SRD, Imbens and Lemieux (2008)

1. Graph the data
2. Estimate the treatment effect by running linear regressions on both sides of the cutoff point.
3. The robustness of the results should be assessed by employing various specification tests.
   - Looking at possible jumps in the value of other covariates at the cutoff point.
   - Testing for possible discontinuities in the conditional density of the forcing variable.
   - Looking whether the average outcome is discontinuous at other values of the forcing variable.
4. Additional recommendations regarding bandwidth, etc.

Spatial RDD

- The idea behind Regression Discontinuity Designs is typically thought of in terms of a forcing variable creating a discontinuity in the treatment outcome.
- In this case, the treatment outcome of interest can be graphed against the forcing variable to visualize the local average treatment effect.
- There is, however, a *spatial* branch of RDD's where the forcing variable is a spatially delineated boundary separating observations with and without a given *treatment* of interest.
- Examples include:
  - Black (1999), who uses school boundaries to examine how much parents value elementary education;
  - Chay and Greenstone (2005), who examine the value of air quality;
  - Greenstone and Gallagher (2008), who examine the value of Superfund (i.e., hazardous waste) cleanups.

Paper #3: Black(1999)

- Black is interested in assessing the value parents place on school quality;
- A traditional approach would be to look at how such values are capitalized into housing prices (i.e., hedonic price analysis);
- The problem is complicated by the fact that the treatment of interest (high quality schooling) is likely correlated with other (often unobservable) neighborhood attributes.
- In particular, better schools are likely located in better neighborhoods, with
  - better shopping;
  - lower crime rates;
  - cultural ;