Regression Discontinuity Design

1

Motivating example

- Many districts have summer school to help kids improve outcomes between grades
  - Enrichment, or
  - Assist those lagging
- Research question: does summer school improve outcomes
- Variables:
  - x=1 is summer school after grade g
  - y = test score in grade g+1

2

- Equation of interest
- $y_i = \beta_0 + x_i\beta_1 + \varepsilon_i$

- Problem: what do you anticipate is $cov(x_i, \varepsilon_i)$?

3

LUSDINE

- To be promoted to the next grade, students need to demonstrate proficiency in math and reading
  - Determined by test scores
- If the test scores are too low – mandatory summer school
- After summer school, re-take tests at the end of summer, if pass, then promoted

4

Situation

- Let Z be test score – Z is scaled such that
  - $Z \geq 0$ not enrolled in summer school
  - $Z<0$ enrolled in summer school
- Consider two kids
  - #1: $Z=\varepsilon$
  - #2: $Z=-\varepsilon$
  - Where ε is small

5

Intuitive understanding

- Participants in SS are very different
- However, at the margin, those just at Z=0 are virtually identical
- One with $z=-\varepsilon$ is assigned to summer school, but $z= \varepsilon$ is not
- Therefore, we should see two things

6

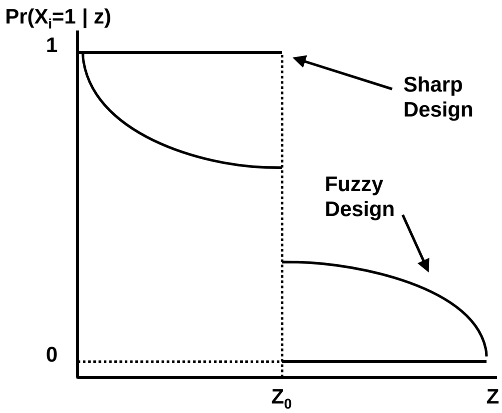
- There should be a noticeable jump in SS enrollment at $z<0$.
- If SS has an impact on test scores, we should see a jump in test scores at $z<0$ as well.

7

Variable Definitions

- y_i = outcome of interest
- x_i =1 if NOT in summer school, =1 if in
- $D_i = I(z_i \geq 0)$ -- I is indicator function that equals 1 when true, =0 otherwise
- z_i = running variable that determines eligibility for summer school. z is re-scaled so that $z_i=0$ for the lowest value where $D_i=1$
- w_i are other covariates

8



9

Key assumption of RDD models

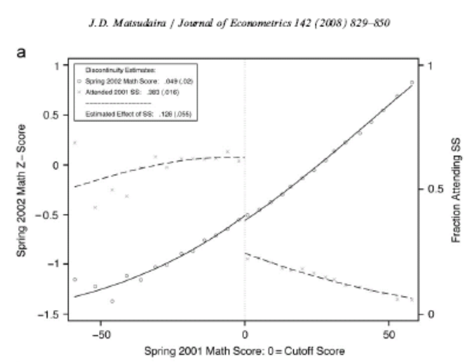
- People right above and below Z_0 are functionally identical
- Random variation puts someone above Z_0 and someone below
- However, this small different generates big differences in treatment (x)
- Therefore any difference in Y right at Z_0 is due to x

10

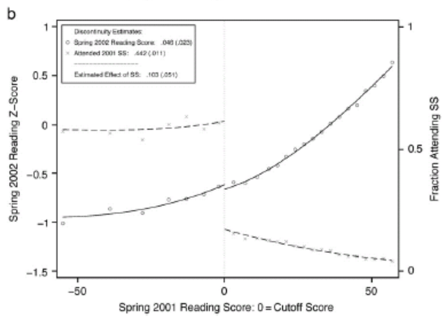
Limitation

- Treatment is identified for people at the z_i=0
- Therefore, model identifies the effect for people at that point
- Does not say whether outcomes change when the critical value is moved

11



12



13

Table 1

	Grade 3		
		Attended SS	
	Total	Yes	No
Outcomes			
2002 math score	641.8	620.4	648.5
	(.142)	(.241)	(.16)
	[36.57]		
2002 reading score	649.7	621.6	658.6
	(.176)	(.241)	(.204)
	[46.40]		
Summer school attendance			
Attended summer school 2001	.24	1	0
	(.002)	(0)	(0)
Days attended	4.373	18.208	0

14

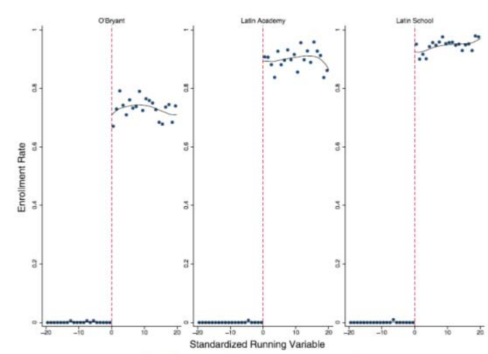
Table 2

	Effect of being mandated		Effect of SS attendance	
	Attendance (1st Stage)	Math (Reduced form)	Math (TSLS)	Reading (TSLS)
Strong 1st stage discontinuity				
Grade 3	.383	.049	.128	.087
	(.016)	(.02)	(.055)	(.065)
Grade 5	.385	.093	.241	.083
	(.006)	(.015)	(.039)	(.055)
Grade 6	.320	.061	.19	n.a.
	(.011)	(.014)	(.047)	(–)

Math:
Grade 3: 0.049/0.383=0.128
Grade 5: 0.093/0.385=0.241
Grade 6: 0.061/0.320=0.190

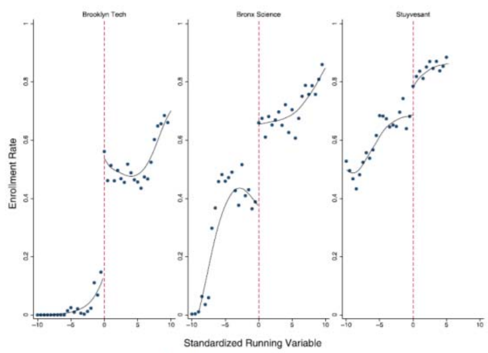
15

Example: Selective High Schools



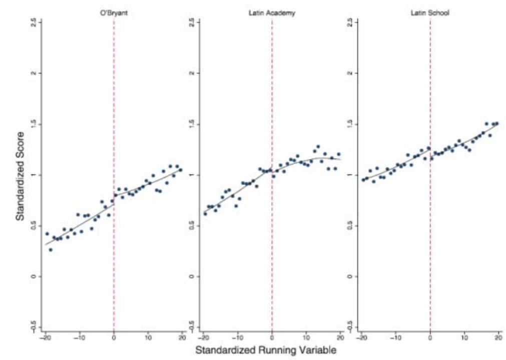
(b) Enrollment at each Boston exam school

16



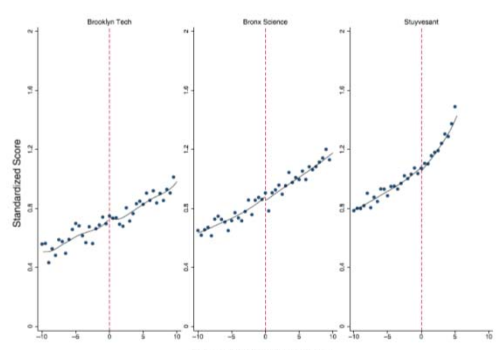
(b) Enrollment at any NYC exam school

17



(a) 10th grade math at Boston exam schools for 7th and 9th grade applicants

18



(a) Regents Advanced Math at NYC exam schools

19

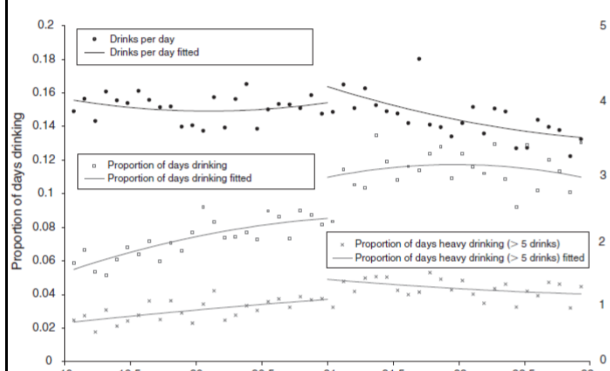


FIGURE 2. AGE PROFILE OF DRINKING INTENSITY

20

Table 1: Participation

	(1)
12 or more drinks in lifetime	
Over 21	0.0418
	(0.0242
Observations	16,107
R^2	0.02
Prob > Chi-Squared	
12 or more drinks in one year	
Over 21	0.0796
	(0.0254
Observations	16,107
R^2	0.02
Prob > Chi-Squared	
Any heavy drinking in last year	
Over 21	0.0761
	(0.0248
Observations	16,107
R^2	0.01
Prob > Chi-Squared	
Covariates	N
Weights	N
Quadratic terms	Y
Cubic terms	N
LLR	N

Table 2: Intensity

	(1)
Proportion of days drinking	
Over 21	0.0245
	(0.0086)
Observations	16,107
R^2	0.02
Prob > Chi-Squared	
Proportion of days heavy drinking	
Over 21	0.0120
	(0.0061)
Observations	15,825
R^2	0.00
Prob > Chi-Squared	
Drinks per day on days drinking	
Over 21	0.2387
	(0.2810)
Observations	9,906
R^2	0.00
Prob > Chi-Squared	
Covariates	N
Weights	N
Quadratic terms	Y
Cubic terms	N
LLR	N

21

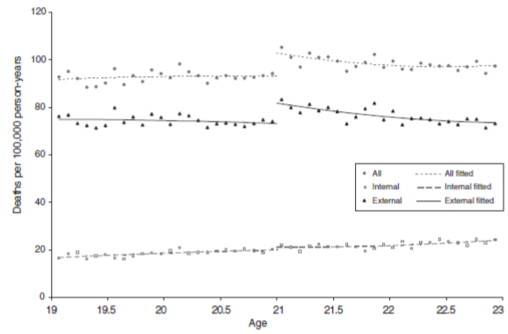


Figure 3. Age Profile for Death Rates

Notes: Deaths from the National Vital Statistics Records. Includes all deaths that occurred in the United States between 1997–2003. The population denominators are derived from the census. See online Appendix C for a list of causes of death.

--

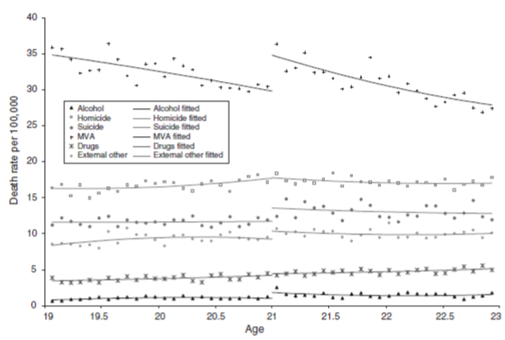


Figure 4. Age Profiles for Death Rates by External Cause

Notes: See notes to Figure 3. The categories are mutally exclusive. The order of precedence is homicide, suicide, MVA, deaths with a mention of alcohol, and deaths with a mention of drugs. The ICD-9 and ICD-10 Codes are in Appendix C.

23