Applied Microeconometrics I

Lecture 10: Regression discontinuity

Tuomas Pekkarinen

Aalto University

October 14, 2021

Lecture Slides

- Institutional rules often assign individuals to “treatments" which can be exploited for estimating causal effects
- The most typical case are threshold rules that are based on some ex-ante variable
  - Score in entry exams
  - Income for subsidy eligibility
  - Project quality score for public R&D subsidies
  - Age limit for alcohol consumption
- This ex-ante variable is called the running (forcing, assignment) variable.
- Selected threshold of the running variable assigns individuals into “treated" and “not treated"
- The idea in RDD design is to exploit the randomness of assignment around the threshold

Regression discontinuity design

- The main idea in the RDD is to compare the outcomes below (control) and above (treated) the threshold
- We assume that:
  - Treatment status is a deterministic function of the running variable
  - Treatment status is a discontinuous function of the running variable
- Sharp desgin: Treatment switched from 0 to 1 at the threshold
- Fuzzy design: The probability of treatment jumps at the threshold

Regression discontinuity design

- RDD works when:
  - Variation in treatment status is as good as randomly assigned around the threshold
  - There is no way to precisely manipulate the running variable
  - There are enough observations around the threshold

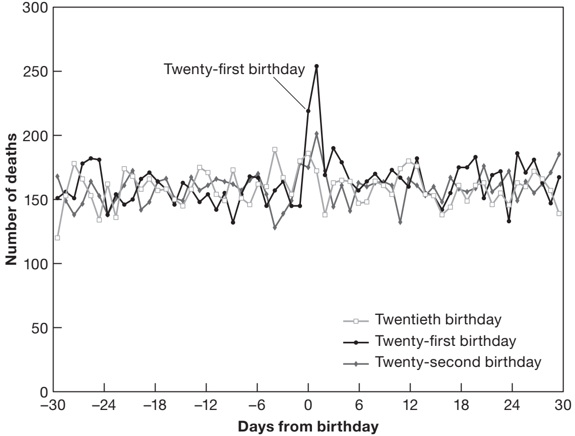
Example: Effect of the Minimum Legal Drinking Age (MLDA) on death rates

Carpenter and Dobkin (2009)

1. outcome variable y_i: death rate
2. treatment D_i: legal drinking status
3. running variable x_i: age
4. cutoff: MLDA transforms 21-year-olds from underage minors to legal alcohol consumers.

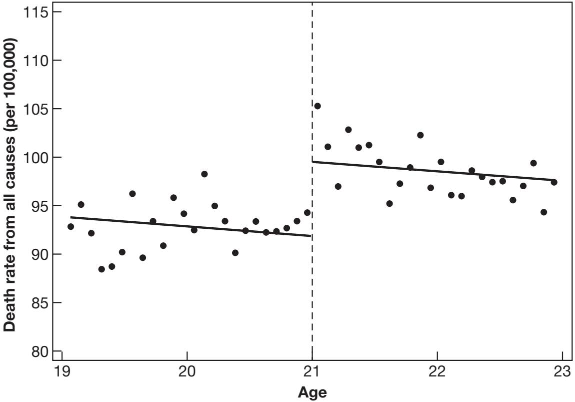
Example: Effect of the Minimum Legal Drinking Age (MLDA) on death rates

FIGURE 4.1
Birthdays and funerals



Example: Effect of the Minimum Legal Drinking Age (MLDA) on death rates

FIGURE 4.2
A sharp RD estimate of MLDA mortality effects



Notes: This figure plots death rates from all causes against age in months. The lines in the figure show fitted values from a regression of death rates on an over-21 dummy and age in months (the vertical dashed line indicates the

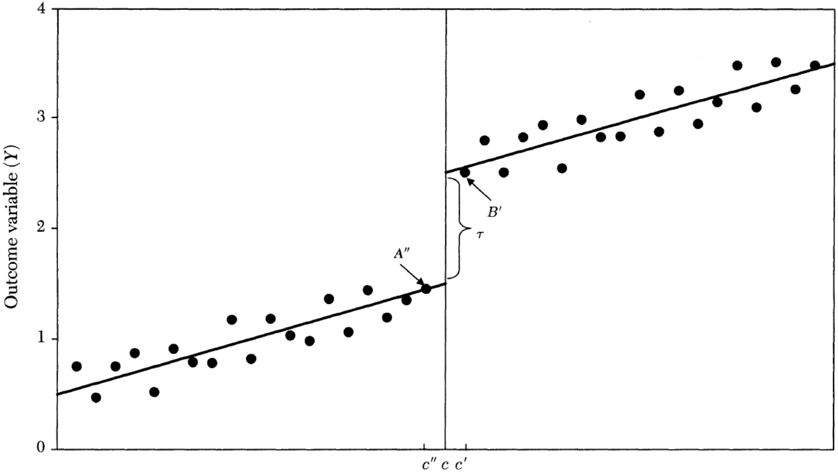
Sharp Regression Discontinuity Design

- Suppose that treatment status (D_i) is deterministic and discontinuous function of the running (assignment, forcing) variable (x_i):
  - $D_i = 1$ if $x_i > c$
  - $D_i = 0$ if $x_i < c$
- In this case, we have a sharp RDD
- All individuals to the right of the cut off are exposed to the treatment and all those to the left are denied the treatment

Sharp Regression Discontinuity Design: Linear case

- Suppose we can write the relationship between Y, D, and X as:
$$Y = \alpha + D\tau + X\beta + \epsilon$$
- We are assuming that the relationship between Y and X is linear
- Y is a discontinuous function of D generating a treatment effect τ

Simple linear RD set up



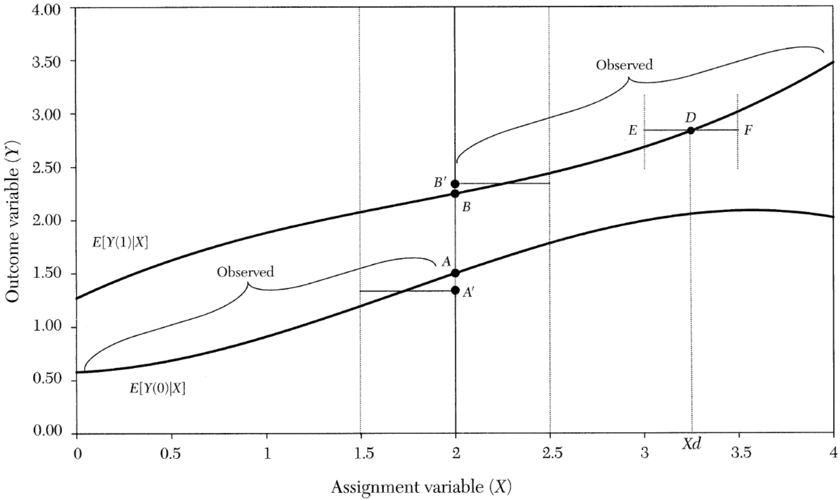
Sharp Regression Discontinuity Design: Linear case

- Y jumps at $X = c$
- We assume that all factors, other than D, affecting Y evolve smoothly with respect to X
- B' would be a reasonable guess for value of Y when $D = 1$
- A' would be a reasonable guess for value of Y when $D = 0$
- Then $B' - A'$ would be the impact of treatment on Y

Sharp Regression Discontinuity Design: Linear case

- Inherent tradeoff in RDD:
  - Estimates are more accurate, the closer we are to the thershold
  - The closer we are to the threshold, the less data we have
- We need to use data away from the threshold
- As a result we need to assume a functional form for the relationship between Y and X

Nonlinear RD set up



Sharp Regression Discontinuity Design: Specifying the functional form

- One way to estimate the treatment effect in an RD set up is to specify the functional form between Y and X
- We already saw the linear example
- But in general the relationship can be any $f(X_i)$:

$$Y_i = \alpha + \tau D_i + f(X_i) + \epsilon_i$$

- $f((X_i)$ can be, for example, ρ:th order polynomial:

$$f(X_i) = \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + ... + \beta_\rho X_i^\rho$$

- $f(X_i)$ can also be estimated separately at each side of the cutoff point
- Relies on the assumption that $f(X_i)$ is an adequate description of the relationship between Y and X
- The further away from the threshold we are, the bolder this assumption is

Sharp Regression Discontinuity Design: Estimation within a bandwidth

- In the previous graph:

$$B - A = \lim_{\epsilon \to 0} E[Y_i | X_i = c + \epsilon] - \lim_{\epsilon \to 0} E[Y_i | X_i = c - \epsilon]$$

- which at the limit is equal to:

$$E[Y_i(D_i = 1) - Y_i(D_i = 0) | X_i = c]$$

- This is the treatment effect at the thershold c
- Around the threshold we can use the outcomes below the threshold as a valid counterfactuals for outcomes above the threshold

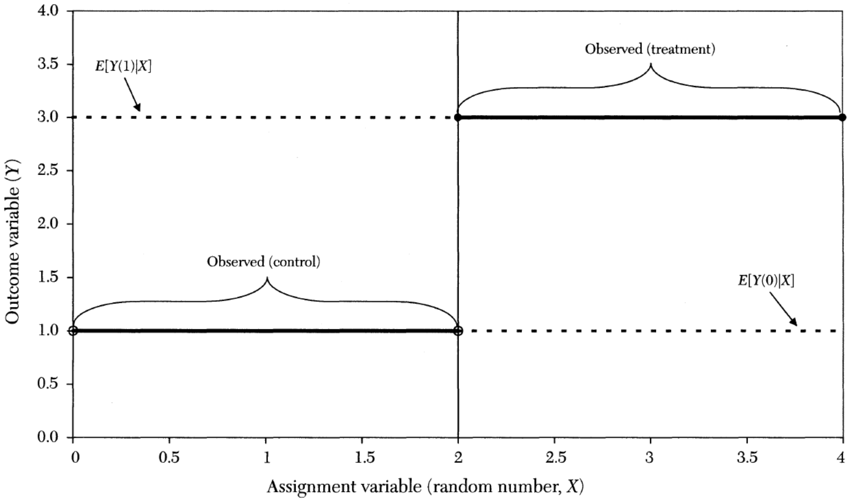
Sharp Regression Discontinuity Design: Estimation within a bandwidth

- How should we estimate $E[Y_i|X_i = c + \epsilon]$ and $E[Y_i|X_i = c - \epsilon]$
- Non-parametric methods: Local linear regressions within a given bandwidth (window) of width h around the threshold
- How to choose h?
- Tradeoff between precision and bias
- Literature on optimal bandwidths

RD design as a local RCT

- The relationship between RDD and RCT
- In RCT the assignment variable X is completely random and therefore independent of Y_{0i}, Y_{1i}
- The average treatment effect can be computed as a differences in mean value of Y on the right and left hand side of the threshold
- RDD as an RCT where individuals have incomplete control over X
- Then treatment is as good as randomly assigned only around the cutoff point

RCT as RDD



4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Outcome variable (Y)
E[Y(1)|X]
Observed (treatment)
Observed (control)
E[Y(0)|X]
0
0.5
1
1.5
2
2.5
3
3.5
4
Assignment variable (random number, X)

Validity of RDD

- RDD relies on the assumption that individuals are not able to influence the assignment variable precisely
- There are ways to test this assumption:
  - Baseline characteristics should have the same distribution just above and below the threshold
  - Density of the running variable, X, should be continuous at the threshold (McCrary test)

Sharp design example: Causal effect of incumbency, Lee (2008

- Does a democratic candidate for a seat in the U.S. house of representatives have an advantage if his party won the seat in the previous election?
- Exploits the fact the previous election winner is determined by rule $D_i = 1$ if $x_i \geq c$ where c the threshold for winning (50 % in a two party state)
- Because D_i is a deterministic function of x_i there should be no confounding factors other than x_i

Probability of winning the election

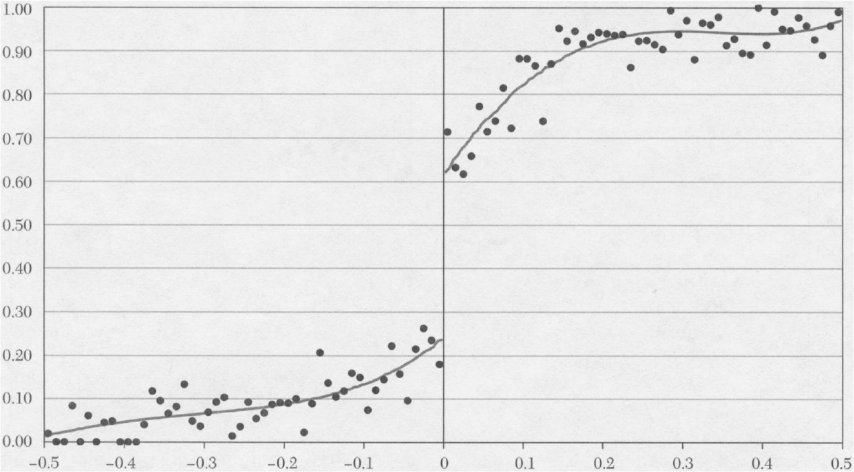


Figure 10. Winning the Next Election, Bandwidth of 0.01 (100 bins)

Estimates with different bandwidths and functional forms

TABLE 3
RD Estimates of the Effect of Winning the Previous Election on Probability of Winning the Next Election

Bandwidth:	1.00	0.50	0.25	0.15	0.10	0.05	0.04	0.03	0.02	0.01
Polynomial of order:										
Zero	0.814	0.777	0.687	0.604	0.550	0.479	0.428	0.423	0.459	0.533
	(0.007)	(0.009)	(0.013)	(0.018)	(0.023)	(0.035)	(0.040)	(0.047)	(0.058)	(0.082)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.011]	[0.201]	[0.852]	[0.640]	[0.479]	
One	0.689	0.566	0.457	0.409	0.378	0.378	0.472	0.524	0.567	0.453
	(0.011)	(0.016)	(0.026)	(0.036)	(0.047)	(0.073)	(0.083)	(0.099)	(0.116)	(0.157)
	[0.000]	[0.000]	[0.126]	[0.269]	[0.336]	[0.155]	[0.400]	[0.243]	[0.125]	
Two	0.526	0.440	0.375	0.391	0.450	0.607	0.586	0.589	0.440	0.225
	(0.016)	(0.023)	(0.039)	(0.055)	(0.072)	(0.110)	(0.124)	(0.144)	(0.177)	(0.246)
	[0.075]	[0.145]	[0.253]	[0.192]	[0.245]	[0.485]	[0.367]	[0.191]	[0.134]	
Three	0.452	0.370	0.408	0.435	0.472	0.566	0.547	0.412	0.266	0.172
	(0.021)	(0.031)	(0.052)	(0.075)	(0.096)	(0.143)	(0.166)	(0.198)	(0.247)	(0.349)
	[0.818]	[0.277]	[0.295]	[0.115]	[0.138]	[0.536]	[0.401]	[0.234]	[0.304]	
Four	0.385	0.375	0.424	0.529	0.604	0.453	0.331	0.134	0.050	0.168
	(0.026)	(0.039)	(0.066)	(0.093)	(0.119)	(0.183)	(0.214)	(0.254)	(0.316)	(0.351)
	[0.965]	[0.200]	[0.200]	[0.173]	[0.292]	[0.593]	[0.507]	[0.150]	[0.244]	
Optimal order of the polynomial	4	3	2	1	1	2	0	0	0	1
Observations	6,558	4,900	2,763	1,765	1,209	610	483	355	231	106

Sharp design example: Causal effect of incumbency,Lee (2008

- Result suggest that incumbency raises the re-election probability by 40%
- Checks for validity
  - Bunching in the distribution of x near the cutoff c?
  - Discontinuities in pretreatment covariates

Fuzzy RDD

- In sharp RDD treatment jumps from 0 to 1 at the threshold
- In fuzzy RDD the probability of treatment jumps at the threshols

$$Pr(D_i = 1|x_i) = \begin{cases} g_1(x_i) & \text{if } x_i \geq c \\ g_0(x_i) & \text{if } x_i < c \end{cases}$$

so that $g_1(x_i) \neq g_0(x_i)$

Fuzzy RDD

- A treatment effect can be recovered by dividing the jump in the relationship between Y and X at the threshold (the reduced form) by the jump in the the probability of treatment at the threshold (the first stage):

$$\tau = \frac{\lim_{\epsilon \to 0} E[Y_i | X_i = c + \epsilon] - \lim_{\epsilon \to 0} E[Y_i | X_i = c - \epsilon]}{\lim_{\epsilon \to 0} E[D_i | X_i = c + \epsilon] - \lim_{\epsilon \to 0} E[D_i | X_i = c - \epsilon]}$$

- Note the analogy to the Wald estimate in the IV strategy
- The threshold as an instrument that creates exogenous variation in the probability of treatment
- We identify the effect for the individuals at the threshold

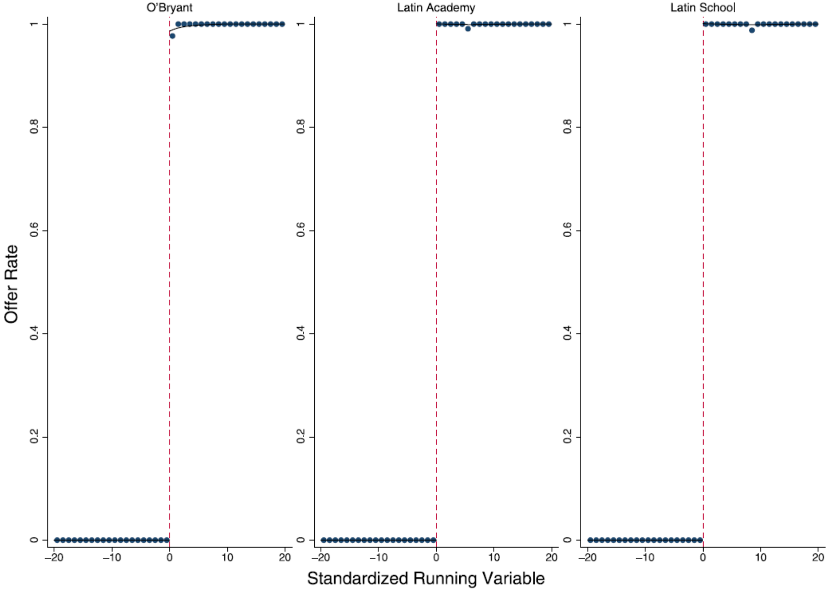
Example of "Fuzzy Design": Abdulkadiroglu, Angrist, and Pathak, Econometrica 2014

- What is the effect of attending an elite high school on student achivement?
- Focus on competitive elite schools in Boston and New York
- These schools select their students based on admissions tests
- Admission threshold creates a discontinuity in the probability of being admitted
- Autors use these entry thresholds to estimate the effect of attending an elite school on test scores
- Parallels to situation in Helsinki high schools

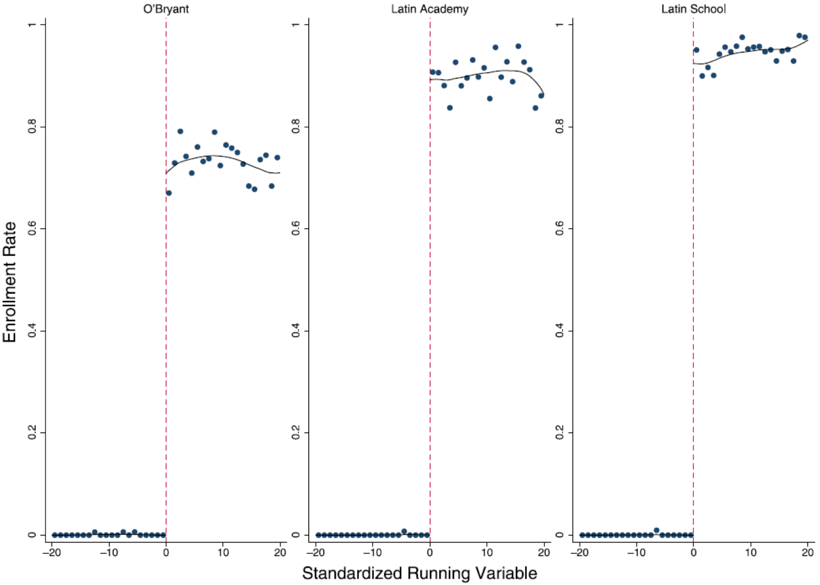
Example of "Fuzzy Design": Abdulkadiroglu, Angrist, and Pathak, Econometrica 2014

- We would expect the probability of receiving an offer from a school to jump from 0 to 1 at the entry threshold
- However, the probability of enrollment may not jump from 0 to 1
    - Some applicants receive multiple offers and only choose to enroll in the preferred school
    - Rejected slots will be filled from the waiting list below the threshold
- There's clear ranking between schools
    - Ones who are admitted to the best school are very likely to enroll
    - Ones who are below the threshold of the worst elite school should not be able to enroll in any of the elite schools

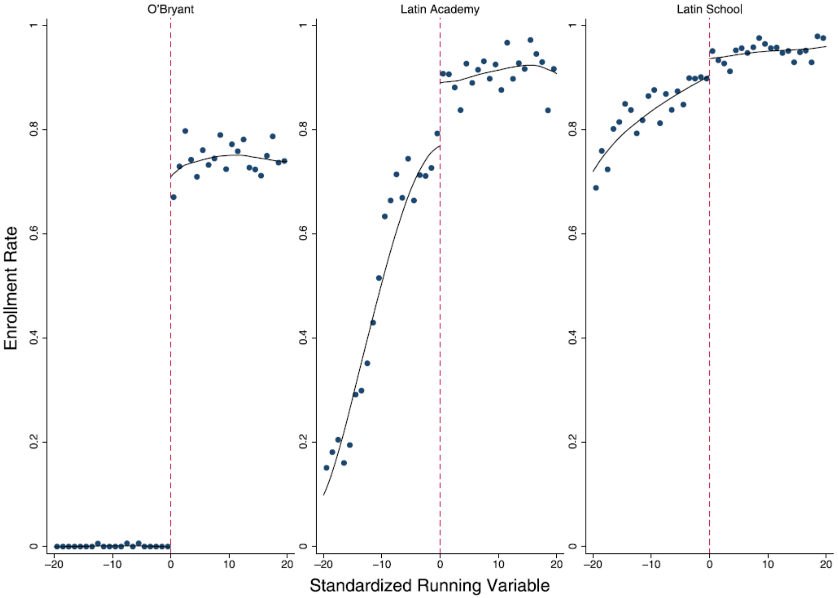
Offers at each Boston elite school



Enrollment at each Boston elite school



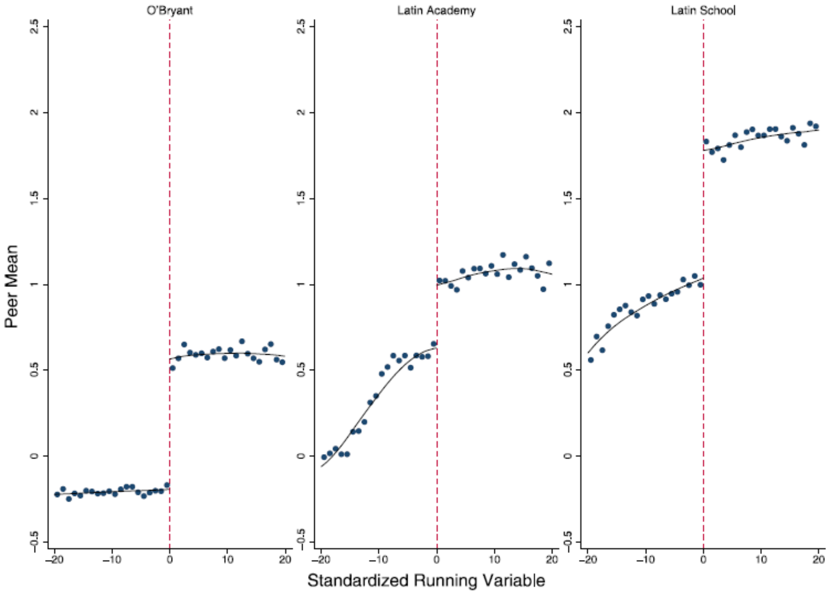
Enrollment at any Boston elite school



Example of "Fuzzy Design": Abdulkadiroglu, Angrist, and Pathak, Econometrica 2014

- Most rejected applicants are admitted to some other elite school
- Does the school quality really vary at all at these thresholds?
- One way to examine this is to check how the quality of fellow students jumps at the threshold
- Peer quality = the average test score of one's peers in the same school

Peer quality at the elite school thresholds



Example of "Fuzzy Design": Abdulkadiroglu, Angrist, and Pathak, Econometrica 2014

- Suppose we are intrested in the effect of peer quality on student achievement
- Denote student's end of high school test score with Y and her pre high school test score with X
- One could try to estimate the effect of peers' average pre high school test scores, $\bar{X}$, with the following regression:

$$Y_i = \theta_0 + \theta_1 \bar{X}_i + \theta_2 X_i + u_i$$

- What could go wrong here?

Example of "Fuzzy Design": Abdulkadiroglu, Angrist, and Pathak, Econometrica 2014

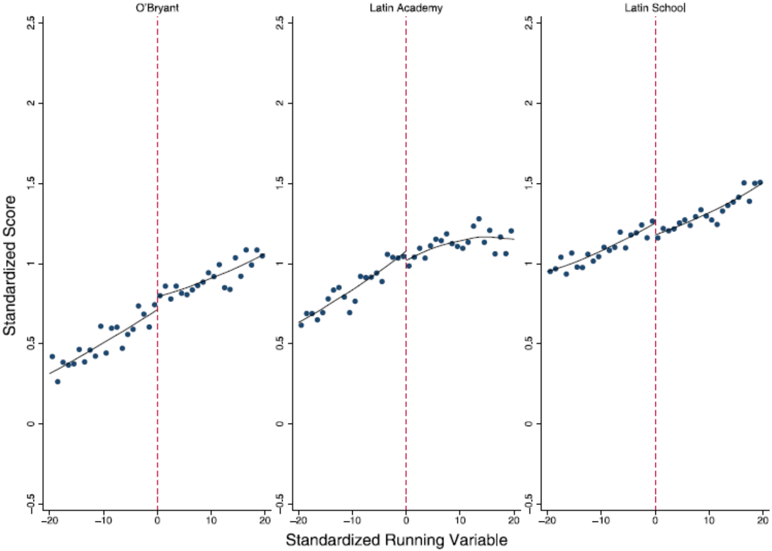
- Entry thresholds create "as good as random" variation in the entry probability
- We can write the reduced form as:

$$Y_i = \alpha_0 + \rho D_i + \beta_0 R_i + e_{0i}$$

  where $D_i = 1$ for accepted applicants and R_i is the running variable
- The first stage can be written as:

$$\bar{X}_i = \alpha_1 + \phi D_i + \beta_1 R_i + e_{1i}$$

Reduced form: 10th grade math test scores



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

Example of "Fuzzy Design": Abdulkadiroglu, Angrist, and Pathak, Econometrica 2014

- There is hardly any visible reduced form
- Given this, it is not surprising that 2SLS estimates are approximately zero for all outcomes
- Elite schools do not seem to have any effect on achievement
- What does the locality of RDD imply for the intepretation of these estimates?

2SLS: Boston and New York combined

TABLE IX

2SLS ESTIMATES FOR BOSTON AND NEW YORK[a]

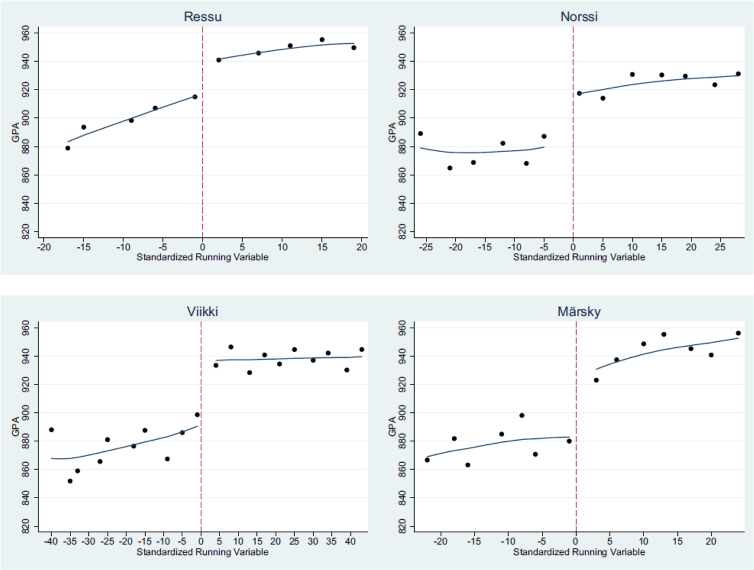
	Math					English				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	2SLS Estimates (Models With Cohort Interactions)									
Peer mean	−0.038		0.064	−0.035		0.006		0.044	−0.047	
	(0.032)		(0.080)	(0.044)		(0.030)		(0.064)	(0.051)	
Proportion nonwhite		0.145	0.421		0.160		−0.014	0.141		0.063
		(0.110)	(0.279)		(0.137)		(0.102)	(0.218)		(0.134)
Years in exam school				−0.003	0.006				0.045	0.027
				(0.036)	(0.030)				(0.034)	(0.025)
	First-Stage F-Statistics (Models With Cohort Interactions)									
Peer mean	65.8		9.1	50.0		39.8		5.7	22.8	
Proportion nonwhite		65.8	17.6		60.0		52.3	12.4		41.2
Years in exam school				12.0	16.2				10.6	15.8
N	31,911	33,313	31,911	31,911	33,313	31,222	32,185	31,222	31,222	32,185

(*Continues*)

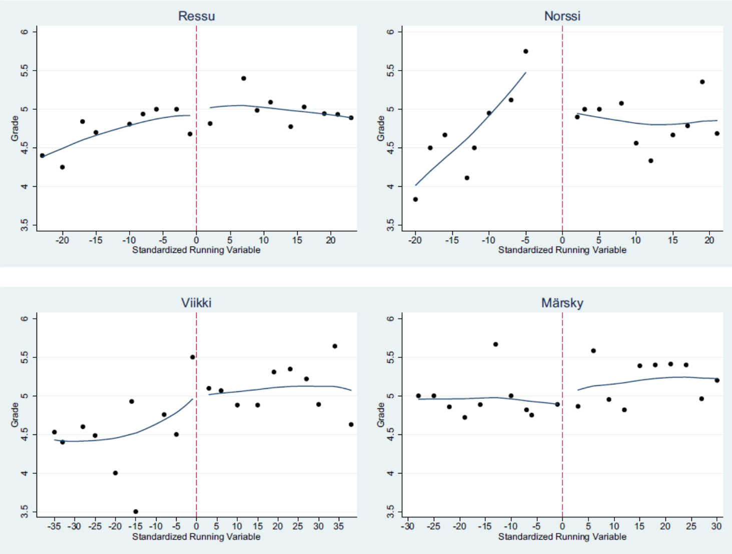
Are elite schools in Helsinki any better?

- Lassi Tervonen's master thesis from University of Helsinki is a replication of Abdulkadiroglu et al with data from Helsinki region
- There are more or less clear elite schools in Helsinki
- Entry thresholds based on comprehensive school GPA
- Just as in Boston the peer quality jumps at the threshold
- Reduced form and 2SLS effects are zero

Peer quality at the elite school thresholds in Helsinki



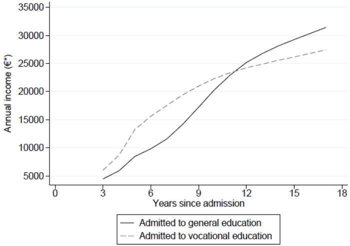
Reduced form: Mother tongue matriculation exam grade

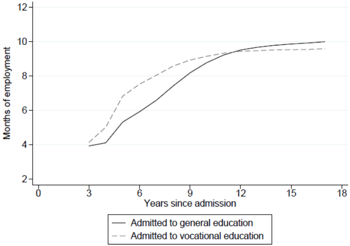


Silliman and Virtanen: Labor market returns to vocational secondary education

- In many European education system the critical choice concerns the type of secondary education: academic or vocational
- Trade-off
  - Academic education provides general skills and prepares for further education
  - Vocational education provides specific skills and prepares directly for the labor market
- Typically vocational education graduates earn more in the early stage of the career and less later on

Annual earnings and employment of Finnish vocational and academic track graduates





Silliman and Virtanen: Labor market returns to vocational secondary education

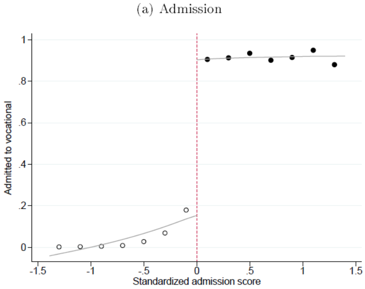
- Mean differences between types of graduates may be driven by selection
  - Academic aptitude
  - Preferences
- Would students who are marginally admitted to academic secondary education benefit from studying in the vociational track instead?

Silliman and Virtanen: Labor market returns to vocational secondary education

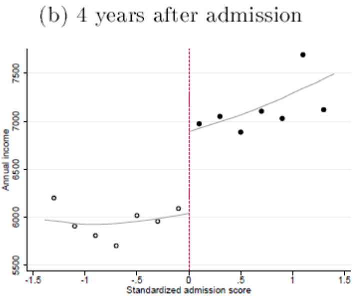
- Students selected based on their compulsory school GPA: c_{ik}
- Over-subscribed programs have an admission cutoff: τ_k
- Focus on students who apply to both academic and vocational programs
- Distance to the cutoff k for student i is: $a_{ik} = c_{ik} - \tau k$
- Use cut-offs from the applicants' first-ranked preference:

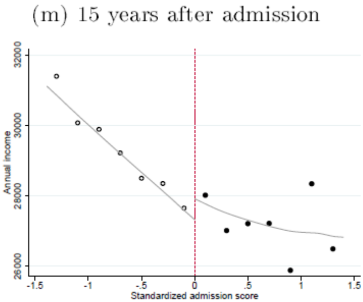
$$r_{ik} = \begin{cases} a_{ik} & \text{if Vocational } \succ \text{ Academic} \\ -1a_{ik} & \text{if Academic } \succ \text{ Vocational} \end{cases}$$

Admission and enrollment around the cutoffs

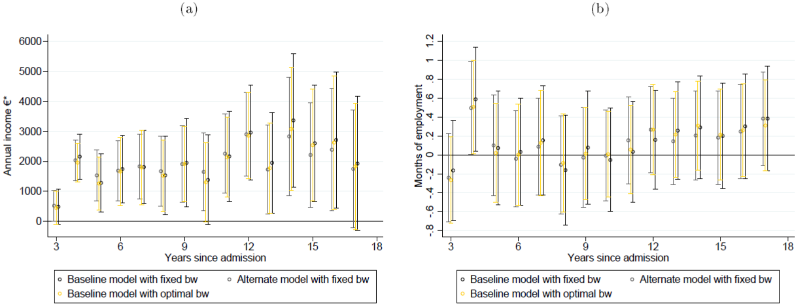


Earnings around the cutoffs 4 and 15 years after admission





Year-by-year RDD estimates of the effect of enrollment into vocational education



Silliman and Virtanen: Labor market returns to vocational secondary education

- Vocational education increases earnings until age 33
- No sign of trending off
- No effects on employment
- Vocational seems to be beneficial for applicants at the margin
- Selection based on comparative advantage

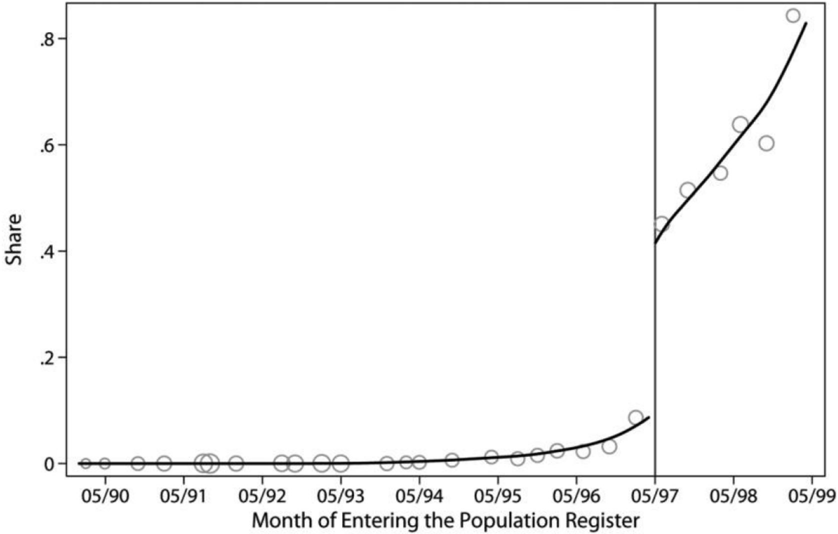
Example: Integration plans for immigrantsSarvimäki and Hämäläinen, 2016

- Labour market integration of immigrants is a hot topic in many countries
- Active labour market policies targeted at immigrants
- Sarvimäki and Hämäläinen study the effect of immigrant integration plans in Finland
- Mandatory for recently arrived immigrants who are unemployed or collect welfare benefits

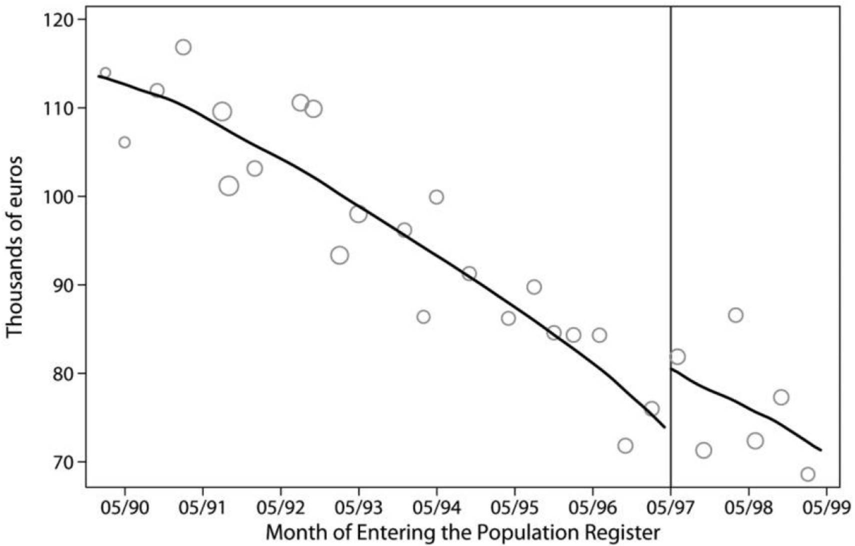
Example: Integration plans for immigrants Sarvimäki and Hämäläinen, 2016

- Integration plans were implemented on May 1 1999
- Applied to those immigrant who arrived after May 1 1997
- Immigrants who had arrived earlier were exempted
- RDD: Use May 1 1997 cutoff to identify the effect of integration plans on earnings and benefit uptake

First stage: Integration plans by month of arrival



Reduced form: Earnings by month of arrival



Example: Integration plans for immigrants Sarvimäki and Hämäläinen, 2016

- Use only immigrants who arrived within h days of the cutoff for estimation
- Use optimal bandwidth algorithms to choose h: 42 months for earnings, 40 months for plans

Example: Integration plans for immigrants Sarvimäki and Hämäläinen, 2016

- Reduced form: OLS estimation of the following regression:

  $$y_i = \alpha + \beta \mathbb{1}[r_i \geq r_0] + \delta_0 (r_i - r_0) + \delta_1 \mathbb{1}[r_i \geq r_0](r_i - r_0) + X_i \eta + \epsilon_i$$

  where y_i is the outcome for immigrant i, $\mathbb{1}$ is an indicator function, r_i is date of arrival, r_0 is May 1 1997, and X_i are observable controls

- First stage: OLS estimation of the following regression:

  $$D_i = \mu + \gamma \mathbb{1}[r_i \geq r_0] + \lambda_0 (r_i - r_0) + \lambda_1 \mathbb{1}[r_i \geq r_0](r_i - r_0) + X_i \pi + \varepsilon_i$$

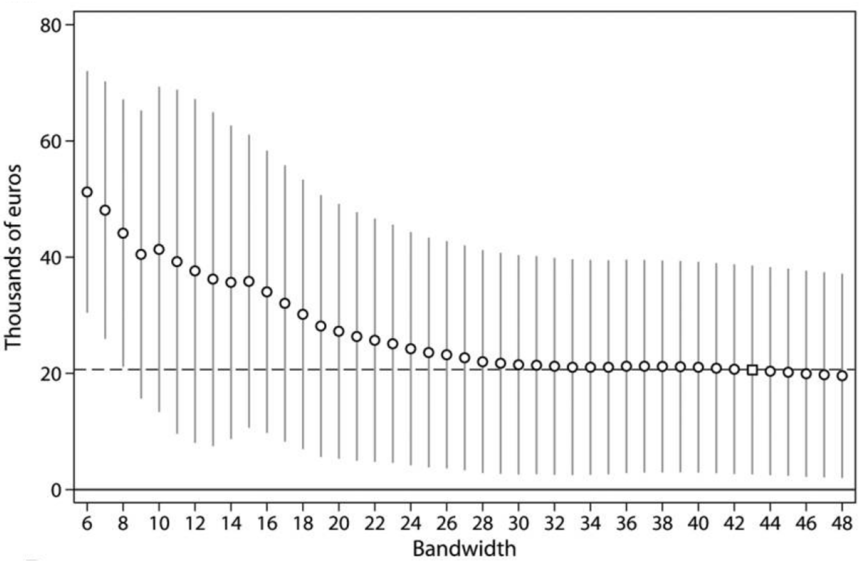
  where D_i is indicator for immigrant i getting an integration plan

- The local average treatment effect of the integration plan is $\hat{\tau} = \frac{\hat{\beta}}{\hat{\gamma}}$

Impact of the integration plans on earnings and benefits

	Earnings		Benefits	
	(1)	(2)	(3)	(4)
Reduced form	7,286	7,238	−2,785	−2,684
	(4,094)	(3,091)	(1,758)	(1,281)
First-stage	.35	.35	.35	.35
	(.02)	(.02)	(.02)	(.02)
Local average treatment effect (LATE)	20,916	20,702	−8,017	−7,698
	(11,891)	(9,107)	(5,103)	(3,681)
Compliers' expected outcomes in the absence of the treatment	44,445	44,420	61,249	60,810
	(9,962)	(8,900)	(4,314)	(3,049)
LATE relative to the baseline	.47	.47	−.13	−.13
Additional covariates	No	Yes	No	Yes
Bandwidth (months)	42	42	40	40
First-stage *F*-statistic for the excluded instrument	322.0	390.1	318.1	384.5
Observations	16,615	16,615	16,173	16,173

Sensitivity w.r.t bandwidth



Example: Integration plans for immigrants Sarvimäki and Hämäläinen, 2016

- Integration plans increased earnings and reduced benefits take-up
- However, they had no effect on total amount of training received by the immigrants
- The authors interpret that the effect is coming through changes in the content of training

What did we do last time?

- RDD: exploit randomness of treatment assignment around a threshold
  - Y_i, outcome
  - X_i, running variable
  - D_i, treatment which is a deterministic and discontinuous function of X_i
- RDD as a RCT with incomplete influence of the assignment of treatment

What did we do last time?

- Sharp RDD
  - $D_i = 1$ if $X_i \geq c$
  - $D_i = 0$ if $X_i < c$
- Estimation
  - Assume: $Y_i = \alpha + \tau D_i + f(X_i) + v_i$
  - Estimate:

$$\lim_{\epsilon \to 0} E[Y_i | X_i = c + \epsilon] - \lim_{\epsilon \to 0} E[Y_i | X_i = c - \epsilon]$$

  - Choose bandwidth h
  - Limit data to $X \in [c - h, c + h]$
  - Non-parametric estimation within these data
- Test that baseline characteristics are balance around the threshold
- Test that the density of X is continuous at the threshold

What did we do last time?

- Fuzzy RD

$$Pr(D_i = 1|x_i) = \begin{cases} g_1(x_i) & \text{if } x_i \geq c \\ g_0(x_i) & \text{if } x_i < c \end{cases}$$

  so that $g_1(x_i) \neq g_0(x_i)$

- IV analogy: Divide the jump in the relationship between Y and X at the threshold (the reduced form) by the jump in the the probability of treatment at the threshold (the first stage):

$$\tau = \frac{\lim_{\epsilon \to 0} E[Y_i|X_i = c + \epsilon] - \lim_{\epsilon \to 0} E[Y_i|X_i = c - \epsilon]}{\lim_{\epsilon \to 0} E[D_i|X_i = c + \epsilon] - \lim_{\epsilon \to 0} E[D_i|X_i = c - \epsilon]}$$

What did we do last time?

- Abdulkadiroglu et al
  - Admission test threshold to gain access to Boston elite high schools
  - Distcontinuity in the probability of enrolling (the first stage)
  - No jump in high school achivement (reduced form)
  - Jump in the peer quality
- Can we use the RD setting to estimate the effect of peer quality on student achievement?

What did we do last time?

- Problematic exclusion restrtiction: Admission to elite school only affects student performance through peer quality
- But other inputs will change at the threshold as well
- Denote achievement of student i with y_i, peer quality with a_i, and all other relevant school inputs with w_i and assume that:

$$y_i = \beta a_i + \gamma w_i + \eta_i$$

where η_i is the error term and $Cov(a, \eta) \neq 0$ and $Cov(w, \eta) \neq 0$

- Suppose we instrument a with z knowing that the exclusion restriction does not necessarily hold
- We assume that $Cov(z, \eta) = 0$ and $Cov(z, a) \neq 0$. However, we also have that $Cov(z, w) \neq 0$
- We have that:

$$Cov(y, z) = \beta Cov(a, Z) + \gamma Cov(w, z)$$

- so that

$$\frac{Cov(y, z)}{Cov(a, z)} = \beta + \gamma \frac{Cov(w, z)}{Cov(a, z)} = \beta + \gamma \rho$$

where ρ is the 2SLS estimate of the effect of w on a using z as instrument

What did we do last time?

- 2SLS version of the omitted variable bias
- Can we put a sign on this bias?
  - We would expect inputs to affect achievement positively:$\gamma > 0$
  - We would expect the other inputs to be affected positively by a: $\rho > 0$
- Bias is likely to be positve
- 2SLS effects are close to zero
- No evidence on peer quality effects