Applied Microeconometrics I

Lecture 11: Regression discontinuity design (continued)

Stefano Lombardi

Aalto University

October 10, 2023

Lecture Slides

What did we do last time?

- Institutional rules that create experiments
  - goal: identify causal effect of D on Y
  - threshold rules, based on **running variable** X
  - X assigns people to D based on whether $X \geq c$ or not
- Identification:
  - We want to relate the jump of Y at the threshold to that of D
  - this can be done only if no other variables jump at c
  - *Imperfect manipulation* ensures *continuity* of X at the threshold and *local randomization* as a consequence
  - If D is as good as randomly assigned at c, then we can compare average outcome below c with that above c
- Notes:
  - As long as X continuity holds, X (and therefore D) is allowed to be correlated with unobservables
  - (local) randomization in *observational data*

What did we do last time?

- Two types of RDD:
  1. **Sharp**: treatment probability jumps by exactly 1 at c
  2. **Fuzzy**: the jump is smaller than 1

- Example: effect of minimum legal drinking age on mortality
  - *sharp design* for eligibility to drink (the graph we saw)
  - *fuzzy design* for actual drinking behavior

- **Sharp RDD**:
  - linear vs. non-linear case
  - we *always* need to control for X
    - we need to extrapolate *at* c, and to do so we need $f(X_i)$
    - the estimate of τ depends on *assumed* $f(X_i)$
  - estimation: local methods (non-parametric) are more flexible/credible
    - we still rely on assumptions on $f(X_i)$
    - must choose h and how much weight to give to observations
    - optimal h for "precision vs. bias tradeoff"

What did we do last time?

- Model specification and estimation in **sharp RDD**
  - re-center X at c and estimate **pooled regression**:

$$Y_i = \alpha_l + \tau D_i + f(\tilde{X}_i) + u_i, \quad \tilde{X}_i \in [-h, h]$$

  with $f(\tilde{X}_i) = f_l(\tilde{X}_i) + D[f_r(\tilde{X}_i) - f_l(\tilde{X}_i)]$

  - for local linear regression: $f_l(\tilde{X}_i) = \beta_l \tilde{X}_i$ and $f_r(\tilde{X}_i) = \beta_r \tilde{X}_i$
  - $\tau = \alpha_r - \alpha_l$ is the treatment effect at c (jump in Y at c)
  - local linear regression (different at the two sides) with same weight to all observations (rectangular kernel) often used
    - this *literally* means fitting two linear regressions (or one pooled) with OLS by using re-centered data close to c.

- Partial tests for validity (on top of plotting Y vs. X):
  1. institutional rules defining X_i:
  2. local random assignment:
  3. continuity (McRary test):
  4. relationship between Y **and** X

Today: Regression Discontinuity Design (continued)

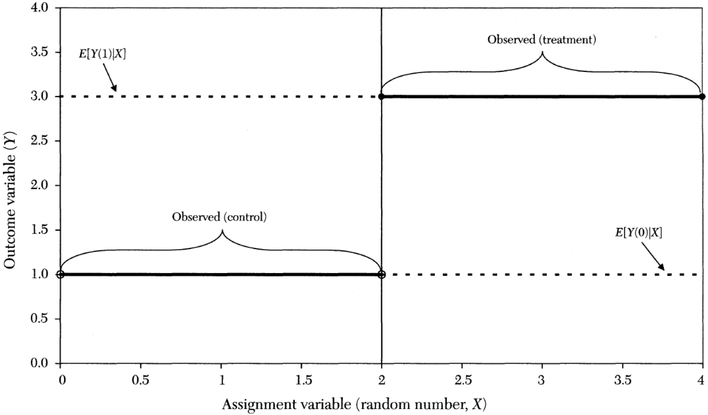
- Sharp RDD:
  - Sharp RDD as a local RCT
  - Example: effect of incumbency on re-election (Lee, 2008)
- Fuzzy RDD:
  - Fuzzy RDD as IV
  - Identification
  - Estimation
  - Interpretation of the results and validity checks
- Several examples

The relationship between sharp RDD and RCT

- RCT is "gold standard" for doing causal inference
- How do we randomize units in a standard RCT?
  - we *assign randomly generated numbers* to units
  - drawing from the uniform distribution gives same ex-ante probability of being treated to each i ("fair lottery")
- Suppose we run a **standard RCT** such that:
  - We draw random variable $X_i \sim U[0,4]$
  - everyone with $X_i \geq 2$ is treated, the others are not
- Then $D_i = 1[X_i \geq 2]$ and X_i is assignment/running variable in a sharp RDD framework (with $c = 2$)
- Due to randomization, X_i (and therefore D_i) is independent of $Y_i(0), Y_i(1)$
- In RCT, continuity is a direct consequence of randomization

The relationship between sharp RDD and RCT

RCT as a sharp RDD



- X_i is independent of $Y_i(0), Y_i(1) \implies E[Y_i(0)|X]$, $E[Y_i(1)|X]$ curves are flat
- Therefore, they are continuous at $X = c$.

The relationship between sharp RDD and RCT

Sharp RDD as a local RCT

- Since the curves are flat, in an RCT the ATET is $E[Y_i|X_i \geq c] - E[Y_i|X_i < c]$, using *all data above/below* c
  - controlling for X is irrelevant (since X independent of Y)
  - using only data around c would be sub-optimal!

- Suppose that we have a **different RCT**:
  - D is a training program for unemployed (provided at the employment office)
  - As before, the employment office draws X_i and assign i to the program if $X_i \geq 2$
  - But the employment office **contaminates randomization**:
    1. everyone receives a monetary compensation inversely proportional to X_i to compensate for "bad luck" when drawing X_i
    2. we (researchers) don't know the exact monetary compensation rule, we only know that it's continuous over X_i for all i.

The relationship between sharp RDD and RCT

Sharp RDD as a local RCT

- This will change job search incentives: people with high X_i will receive little money compared to those with small X_i; hence, they will have higher incentives to search hard.
- The potential outcomes will be affected: X_i has an effect on Y_i (the curves are not flat anymore as in **Figure 2**)
- This is a **"smoothly contaminated" RCT**:
  - using all observations below/above c leads to biased estimates: X affects potential outcomes!
  - however, individuals have no control over X and the monetary compensation is almost identical across c (X *continuity*)
  - then D_i is still as good as randomly assigned around c
- We can think of sharp RDD as a setting where X is endogenous, but randomization still holds at the threshold

The relationship between sharp RDD and RCT

- **Sharp RDD is a local RCT** where people have incomplete control over X
  - the average treatment effect is the difference in mean value of Y on the right and left hand side of the threshold (close to c)
  - we *must* control for X: that's how we extrapolate beyond c
  - this requires to make assumptions about the relationship between Y and X (which in general is not known)
- Abstracting from the example: sharp RDD is a local RCT, *even if we work with observational data*
- Other lessons:
  - After randomization, *never do anything* other than measuring Y in the treated and control group
  - If randomization is contaminated (e.g. implementation issues), we can at least try doing sharp RDD using data close to c

Sharp RDD: Effect of incumbency on re-election

Lee (2008)

- Aim: estimate **incumbent advantage** effect on re-election
- Does a democratic candidate for a seat in the U.S. house of representatives is more likely to be elected 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 $X_i \geq c$
- The threshold for winning c is 50% in a two party state
- D_i is a deterministic function of X_i, and at c there should be no (discontinuous jumps of) confounding factors: by definition X defines winner in past election
- Y_i is "winning the next election" (graph in the next slide)

Probability of winning the election

Lee (2008)

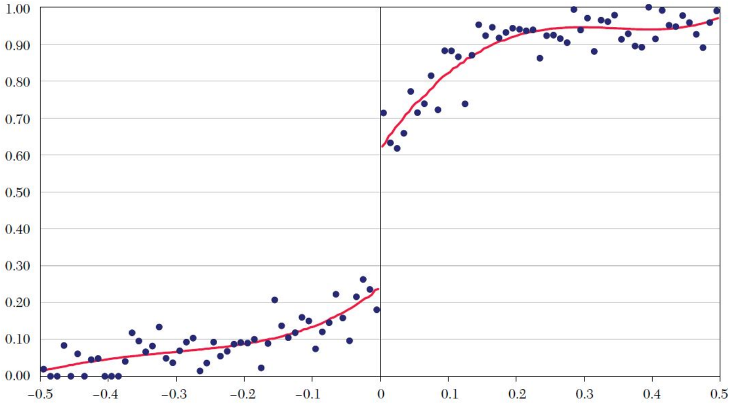


Figure 10. Winning the Next Election, Bandwidth of 0.01 (100 bins)

Estimation: different bandwidths and functional forms

Lee (2008)

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 RDD: Effect of incumbency on re-election

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 pre-treatment covariates?
  - Estimation results robust to the inclusion/exclusion of pre-determined 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 threshold less than 1:

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

  so that $g_1(X_i) \neq g_0(X_i)$, and $g_1(X_i) - g_0(X_i) < 1$

- "fuzziness" might be due to:
  1. additional rules assigning D_i (other than the threshold rule)
  2. the fact that the eligible people can choose to take D_i or not
- Note the parallel with **RCT with imperfect compliance**:
  - crossing c does *not* imply $D_i = 1$ for all
  - that is, there is imperfect take-up of the treatment

Fuzzy RDD

- Since $Pr(D_i = 1|X_i)$ jumps by less than 1 at c, the jump in the relationship between Y and X is no longer an ATET.
    - D *take-up* is chosen endogenously (before, $D_i = 1[X_i \geq c], \forall i$)
    - however, we can exploit the randomized *offer/eligibility* at c
- **IV setting**:
    - *Reduced form*: jump in Y at c
    - *First stage*: jump in the probability of treatment at c
- Scaling up RF by FS:

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

- Analogous to **Wald estimator** in IV (one binary endogenous variable and instrument)
- The **threshold is an instrument** that creates exogenous variation in the probability of treatment

Fuzzy RDD

- Therefore, we can apply what we know about IV!
- τ_F identification needs *more than local randomization*
  1. **First stage**: at c, there is a (large) jump in the relationship between D and X (probability of treatment take-up)
  2. **Exclusion restriction**: crossing c does not impact Y except through the treatment effect
  3. **Monotonicity**: crossing c cannot cause some i to take D and others to reject it
- τ_F is the effect for the **compliers** at c
  - Units who take D when $X \geq c$, but would not do it otherwise
  - τ_F is the average treatment effect on the sub-population affected by the instrument (LATE)
  - "local" LATE (at c)
- The difference between *sharp vs. fuzzy RDD* is parallel to that between *RCT with perfect vs. imperfect compliance*

Model specification and estimation

Fuzzy RDD

- Setting:
  - we focus around c, and recenter X at c so that $\tilde{X}_i \in [-h, h]$
  - D_i is *take-up* (which is endogenous and jumps less than 1 at c)
  - we instrument D_i with the *eligibility* $T_i = 1[\tilde{X}_i \geq 0]$
- The **main model** is:

$$Y_i = \alpha + \tau D_i + f(\tilde{X}_i) + u_i$$

- We need a **first stage regression**:

$$D_i = \gamma + \delta T_i + g(\tilde{X}_i) + v_i$$

- Plugging in the FS in the model we get the **reduced form**:

$$Y_i = \alpha_r + \tau_r T_i + f_r(\tilde{X}_i) + \varepsilon_i$$

- Note:
  - τ is equal to RF/FS (numerically identical to 2SLS)
  - for $f(.)$ and $g(.)$ we use same order of polynomial in $\tilde{X}$
  - as in the sharp case, need to choose h (bias-variance tradeoff)

Fuzzy RDD: Elite colleges and student achievement

Abdulkadiroglu, Angrist, Pathak (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

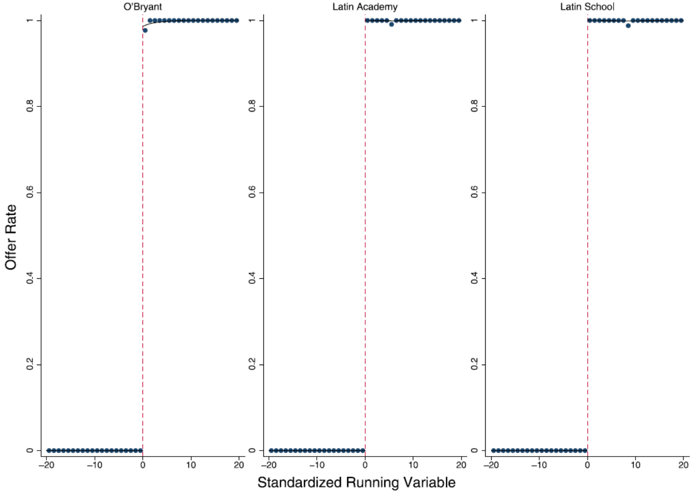
Fuzzy RDD: Elite colleges and student achievement

Abdulkadiroglu, Angrist, Pathak (2014)

- The probability of *receiving an offer* (i.e., *eligibility*) from a school jumps from 0 to 1 at the entry threshold
- However, 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
- The first case is a *sharp design*, the second is a *fuzzy design*.
  - Are we interested in the effect of the *offer* or of *enrollment*?
- In Boston, there's clear ranking between schools
  - Those admitted to the best school are very likely to enroll
  - Those 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

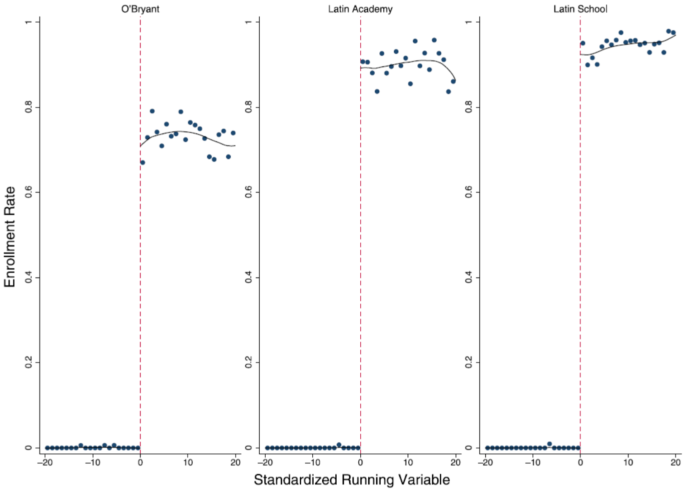
Abdulkadiroglu, Angrist, Pathak (2014)



(a) Offers at each Boston exam school

Enrollment at each Boston elite school

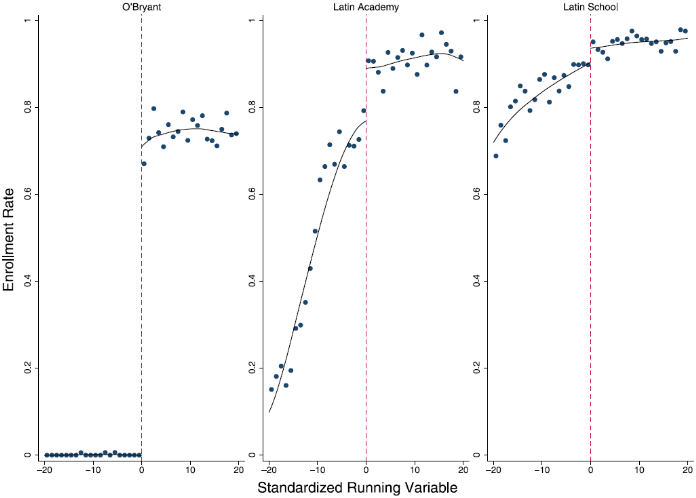
Abdulkadiroglu, Angrist, Pathak (2014)



(b) Enrollment at each Boston exam school

Enrollment at any Boston elite school

Abdulkadiroglu, Angrist, Pathak (2014)



(c) Enrollment at any Boston exam school

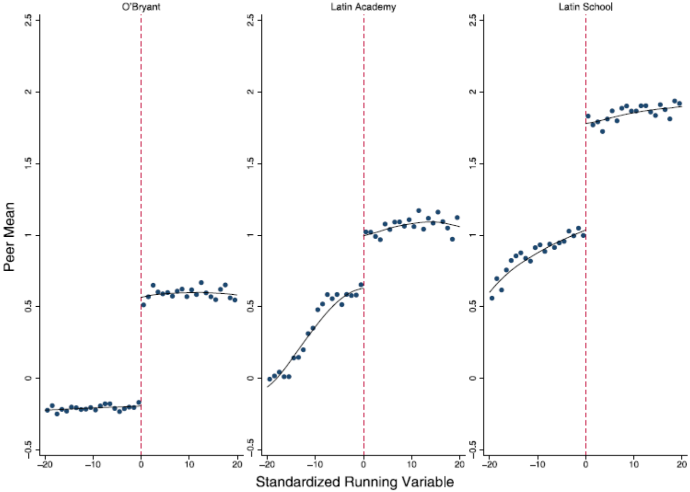
Peer quality at the elite schools

Abdulkadiroglu, Angrist, Pathak (2014)

- Most rejected applicants are admitted to some other elite school (apart from those rejected by O'Bryant)
- Does the **school quality** really vary at all at these thresholds?
- In other words, how to test for the effect of school quality? How do we even measure "school quality"?
- One way to examine this is to check how the quality of fellow students jumps at the threshold
- **Peer quality**: average pre-determined test score of one's peers in the same school (e.g., in middle school)

Peer quality at the elite schools

Abdulkadiroglu, Angrist, Pathak (2014)



(a) Baseline peer math score at Boston exam schools for 7th and 9th grade applicants

The causal effect of peer quality

Abdulkadiroglu, Angrist, Pathak (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 (re-centered) 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?

The causal effect of peer quality

Abdulkadiroglu, Angrist, Pathak (2014)

- Entry thresholds create "as good as random" variation in the entry probability
- As a consequence, applicants are exposed to different peers (we saw this in the last graph: the first stage)
- The **first stage** is:

$$\bar{X}_i = \alpha_1 + \phi T_i + \beta_1 X_i + e_{1i}$$

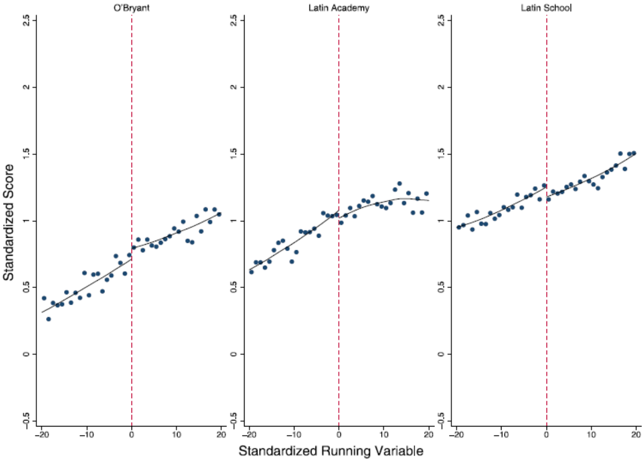
- The **reduced form** is:

$$Y_i = \alpha_0 + \rho T_i + \beta_0 X_i + e_{0i}$$

$T_i = 1$ if at c or above (offered to enroll; might or might not accept); running variable X_i is admission test score; $\bar{X}_i$ is the (endogenous) variable of interest, the average quality i's peers

Reduced form: 10th grade math test scores

Abdulkadiroglu, Angrist, Pathak (2014)



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

2SLS estimates

Abdulkadiroglu, Angrist, Pathak (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

Abdulkadiroglu, Angrist, Pathak (2014)

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*)

Elite colleges and student achievement

Abdulkadiroglu, Angrist, Pathak (2014)

- Does enrollment in elite schools affect performance?
  - Admission test threshold to enter Boston elite high schools
  - *Fuzzy RDD*: discontinuity in prob. of enrolling (first stage)
  - However, no jump in high school achievement (reduced form)
  - Hence, no **2SLS effect**
- Are the elite schools any different in quality from the others?
  - now peer quality is the endogenous variable (instrumented with admission cutoff)
  - clear first stage, no reduced form
  - Hence, no effect of peer quality on high school achievement
- Can we even use this RD setting to estimate the effect of peer quality on student achievement?
  - strong first stage
  - peer quality at c is as good as randomly assigned
  - what else?

The exclusion restriction

Abdulkadiroglu, Angrist, Pathak (2014)

- Admission to elite school only affects student performance through peer quality
- But other inputs will change at the threshold as well (school resources, teacher quality, curriculum, etc.)
- 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 $Cov(w, \eta) \neq 0$ (due to OVB)
- what if we instrument a with the entry threshold?

The exclusion restriction

Abdulkadiroglu, Angrist, Pathak (2014)

- Suppose we instrument a with z (admission threshold) but the exclusion restriction does not necessarily hold
- We assume that $Cov(z, \eta) = 0$ (exogeneity) and $Cov(z, a) \neq 0$ (first stage).
- However, we have $Cov(z, w) \neq 0$ (no exclusion restriction)
- Compute the covariance between the outcome and z:

$$Cov(y, z) = \beta Cov(a, z) + \gamma Cov(w, z) + 0$$

- so that by dividing everything by $Cov(a, z)$:

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

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

The exclusion restriction

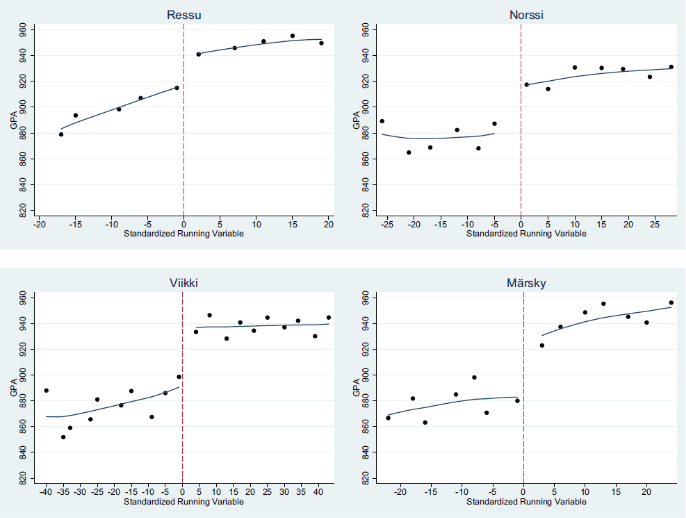
Abdulkadiroglu, Angrist, Pathak (2014)

- This is the 2SLS version of the omitted variable bias
- Can we put a **sign on this bias**?
  - We expect inputs to affect achievement positively: $\gamma > 0$
  - We expect the other inputs to be affected positively by a: $\rho > 0$
- Therefore, the bias is likely to be positive: we overestimate the effect of peer quality.
- Still, even if we likely overestimate the true effect of peer quality, the 2SLS effects are close to zero
- Note that we could come up with alternatives models/stories that would rationalize a negative bias

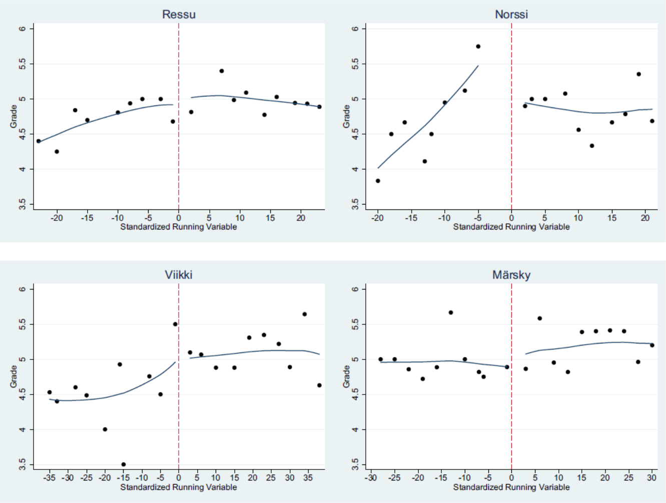
Are elite schools in Helsinki any better?

- Lassi Tervonen's MSc thesis from University of Helsinki is a replication of Abdulkadiroglu et al. (2014) with data from Helsinki region
- There are more or less clear elite schools in Helsinki
    - very high thresholds to get in
- Entry thresholds for secondary education are based on comprehensive school GPA
- Just as in Boston the peer quality jumps at the threshold (strong first stage)
- Outcome: standardized national matriculation exam score (for everyone who went to academic track of secondary school)
- RF and 2SLS are zero (no effect on marginal students)

First stage: Peer quality at elite schools in Helsinki



Reduced form: Mother tongue matriculation exam grade



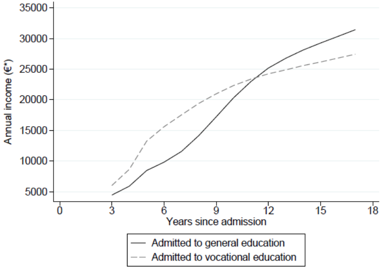
Labor market returns to vocational secondary education

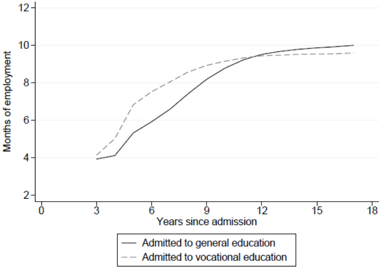
Silliman and Virtanen (2022)

- Policy-wise, how important is it to look at elite schools?
- *Content* of education is arguably more policy relevant
- In many European education system the critical choice concerns *type of secondary education*: academic or vocational
- Trade-off in the content:
  - 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

Labor market outcomes of admitted students, by track

Silliman and Virtanen (2022)





Labor market returns to vocational secondary education

Silliman and Virtanen (2022)

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

Labor market returns to vocational secondary education

Silliman and Virtanen (2022)

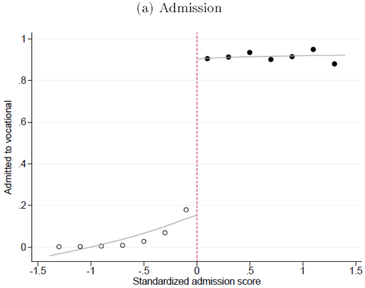
- As we saw, students (i) are selected to programs (k) based on their compulsory school GPA: c_{ik}
- Over-subscribed programs have an admission cutoff: τ_k
- Students provide a list of ranked preferences
- Focus on students who apply to both academic and vocational programs (first two preferences)
- Distance to the cutoff k for student i is: $a_{ik} = c_{ik} - \tau_k$
- Use cut-offs from the applicants' first-ranked preference (define r_k so that above the threshold is "vocational"):

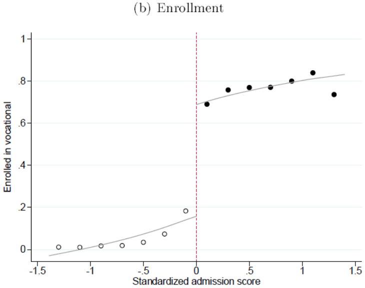
$$r_{ik} = \begin{cases} a_{ik} & \text{if Vocational } \succ \text{ Academic} \\ -a_{ik} & \text{if Academic } \succ \text{ Vocational} \end{cases}$$

r_{ik} is the running variable (centered at the cutoff)

Admission and enrollment around the cutoffs

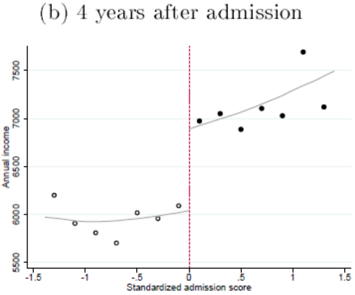
Silliman and Virtanen (2022)





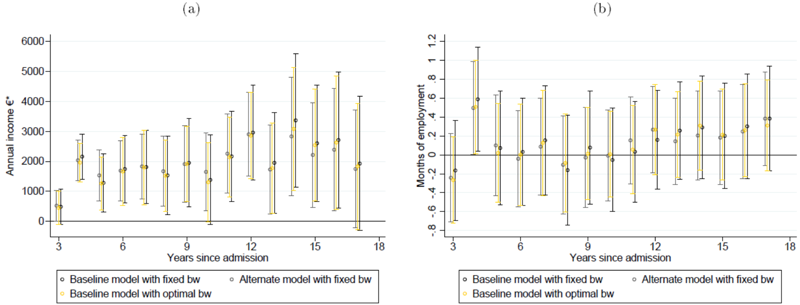
Earnings around cutoffs 4 and 15 years after admission

Silliman and Virtanen (2022)



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

Silliman and Virtanen (2022)



Labor market returns to vocational secondary education

Silliman and Virtanen (2022)

- Vocational education increases earnings until age 33
- No sign of trending off
- No effects on employment: results are not "mechanically" explained by the fact that vocational education students enter the labor market early and work more
- At this margin, vocational seems to be beneficial for applicants
- Selection based on comparative advantage: these marginal students are better off entering vocational education
- Very policy relevant: perhaps surprisingly, vocational education helps people at the margin, also in the long run

Integration plans for immigrants

Sarvimäki and Hämäläinen (2016)

- A non-education RDD example
- Labour market integration of immigrants is a hot topic in many countries (very policy relevant)
- 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
- Note: before the 1990's, net immigration in Finland was *negative* (no existing programs targeted to immigrants, just the regular ones in Finnish)

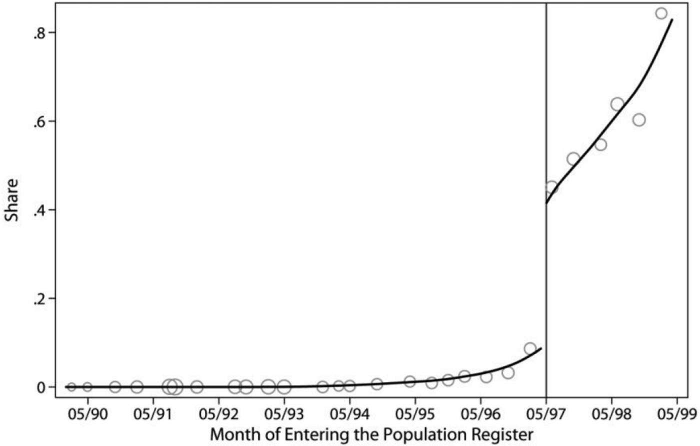
Integration plans for immigrants

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

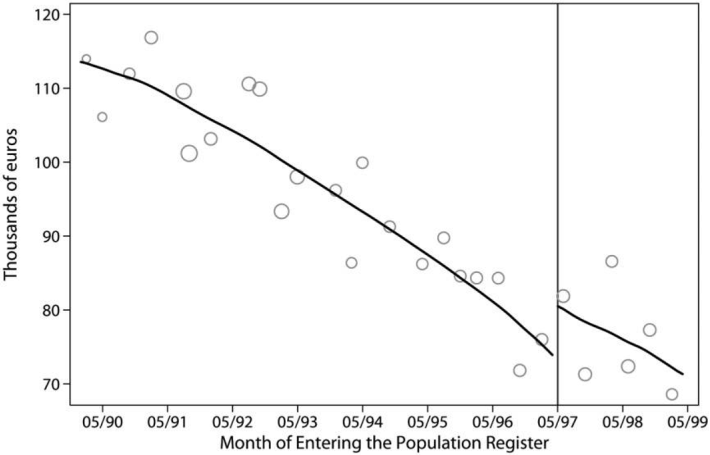
First stage: Integration plans by month of arrival

Sarvimäki and Hämäläinen (2016)



Reduced form: Earnings by month of arrival

Sarvimäki and Hämäläinen (2016)



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 for the two outcomes: 42 months for earnings, 40 months for benefits

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, r_i is date of arrival, r_0 is May 1 1997, and X_i are pre-determined 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 $D_i = 1$ if immigrant i gets an integration plan

- The **LATE** of the integration plan is $\hat{\tau} = \frac{\hat{\beta}}{\hat{\gamma}}$

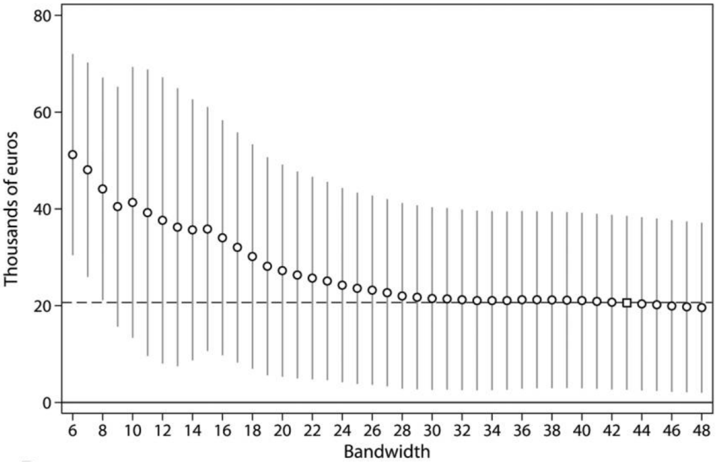
Impact of the integration plans on earnings and benefits

Sarvimäki and Hämäläinen (2016)

	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 to bandwidth selection

Sarvimäki and Hämäläinen (2016)



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
- This makes sense: before, there was no training targeted to immigrants

Sharp Regression Discontinuity Design

Nonlinear case

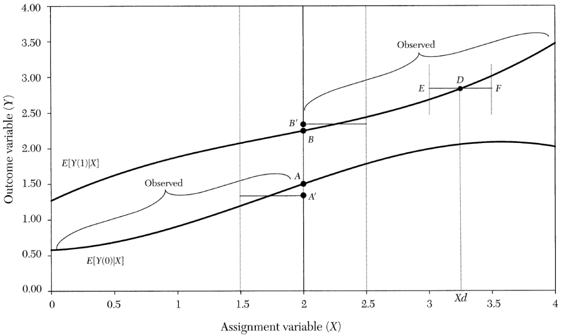


FIGURE 1 – Non-linear RDD

- $Y_i(0)$, $Y_i(1)$: potential outcomes. The i-level treatment effect is never observed.
- $E[Y_i(0)|X]$, $E[Y_i(1)|X]$: underlying relationships between the average outcomes and X (non-linear; different above/below c).
- We *never* observe $E[Y_i(1)|X]$ below $c = 2$ (the opposite holds for $E[Y_i(0)|X]$).