The Regression Discontinuity Design

Magne Mogstad

Introduction

Introduction

Incentives or participation in a service or program is sometimes assigned based on cutoff values, rather than on discretion of administrators or individuals

The Regression Discontinuity (RD) design

- compares individuals on different sides of the cutoff point
  - essentially a local experiment
  - discontinuity can be used as instrument
  - creates random variation in treatment (cond. on running variable)

Origin in the educational psychology literature (Thistlewaite and Campbell, 1960)

- increasingly popular in economics and other sciences
- caution: identified average causal effect is now specific to compliers at the cutoff!

What we will cover

How to apply and interpret the RD design:

- Sharp and fuzzy RD
- Set up in potential outcome framework
  - Assumption: Interpretaton and implications
  - Estimation and inference
- Know how to motivate and check your RD design

Several review papers on RD. It's not rocket science. But some important details, especially for estimation and inference.
See reading lists.

Notation

Definitions

- Outcome Y and binary treatment $D \in \{0, 1\}$
- A variable R that has a discontinuous relationship with D at $R = c$:

$$\mathbb{P}[D = 1 | R = r] \text{ has a discontinuity at } r = c$$

- R is called the **running, forcing** or **assignment** variable
- c is called the **threshold** or **cutoff** value

Examples

- Vote shares (R) to determine election outcomes (D)
- Age (R) to determine program eligibility (D)
- Distance to geographic boundaries (R) affects institutions (D)
- GPA (R) to determine summer school attendance (D)

Intuition

Argument

- We do not believe that either D or R is exogenous
- Suppose that Y_d varies continuously with R at $R = c$
  $\rightarrow$ Contrast with D: varies a lot (discontinuously) at $R = c$
- At $R = c$, potential outcomes vary little, treatment varies a lot
  $\implies$ Changes in Y at $R = c$ should be causally due to D

Limitations and caveats

- The maintained assumption: Y_d varies continuously at $R = c$
  $\rightarrow$ We will talk about the interpretation of this more later
- Without further assumption, only point identify effects for compliers with $R = c$
  $\rightarrow$ Limited external validity; growing area of interest, but few results so far

Sharp RD

Sharp Designs

Definition

- D changes deterministically from 0 to 1 when R crosses c:

$$D = \mathbb{1}[R \geq c]$$

- Example: vote margin - elected ($D = 1$) if and only if $R \geq 0$

Sharp Designs

Connection to selection on observables

- Sharp design implies selection on observables holds for R:
  $\rightarrow (Y_0, Y_1) \perp\!\!\!\perp D | R$ since D is deterministic given R
- But note that overlap necessarily does not hold:

$$\mathbb{P}[D = 1 | R = r] = \begin{cases} 1 & \text{if } r \geq c \\ 0 & \text{if } r < c \end{cases}$$

- The only place overlap almost holds is exactly at R = c
  - "almost" is why RD designs need a smoothness assumption on the potential outcomes as a function of R (around cutoff)

Is the discontinuity sharp or fuzzy?

Is it perfect compliance or not?

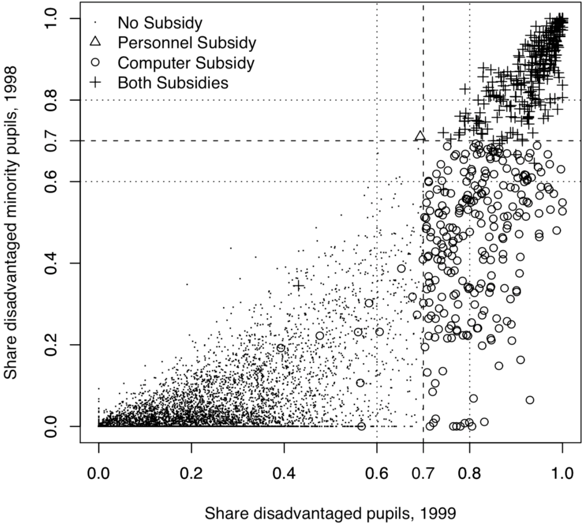
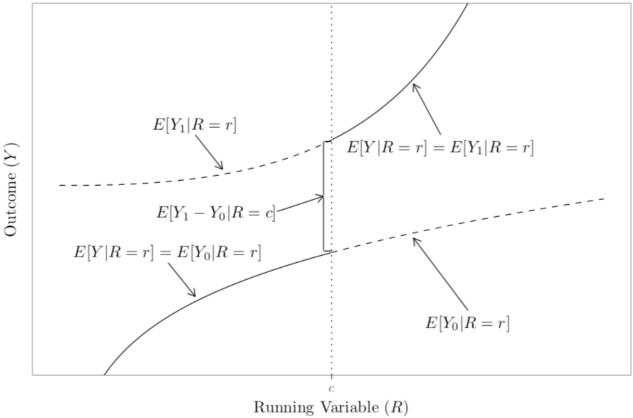


Illustration of a Sharp Design



- $\mathbb{E}[Y_d|R=r]$ nonparametrically identified only where the lines are solid
- With continuity, should be able to point identify $\mathbb{E}[Y_1-Y_0|R=c]$

Formal Identification Argument for Sharp Design

Assumptions

- $\mathbb{E}[Y_d|R=r]$ is continuous at $r=c$ for $d=0,1 \rightarrow$ That's it!
- (We only need this for $d=1$ if we have $R=c \implies D=1$)

Argument

- Take the limit of $\mathbb{E}[Y|R=r]$ as $R \downarrow c$ and $R \uparrow c$:

$$\lim_{r \downarrow c} \mathbb{E}[Y|R=r] \overbrace{=}^{\text{sharp design}} \lim_{r \downarrow c} \mathbb{E}[Y_1|R=r] \overbrace{=}^{\text{continuity}} \mathbb{E}[Y_1|R=c]$$

$$\lim_{r \uparrow c} \mathbb{E}[Y|R=r] = \lim_{r \uparrow c} \mathbb{E}[Y_0|R=r] = \mathbb{E}[Y_0|R=c]$$

- Subtracting the second term from the first shows identification:

$$\underbrace{\mathbb{E}[Y_1 - Y_0|R=c]}_{\text{ATE at the cutoff}} = \underbrace{\lim_{r \downarrow c} \mathbb{E}[Y|R=r] - \lim_{r \uparrow c} \mathbb{E}[Y|R=r]}_{\text{features of the data}}$$

Fuzzy RD

Fuzzy Designs

Definition

- In a **fuzzy design**, $\mathbb{P}[D = 1|R = r]$ is discontinuous at c
- More general than sharp design: $\mathbb{P}[D = 1|R = r] = \mathbb{1}[R \geq c]$
- Now $\mathbb{1}[R \geq c]$ is a strong (but not exact) predictor of D
- Think of $Z \equiv \mathbb{1}[R \geq c]$ as an incentive for D

Fuzzy designs are IV designs

- $(Y_0, Y_1) \perp\!\!\!\perp Z|R$ since Z is a function of R - so Z is exogenous
- Z is also relevant, local to $R = c$:

$$\lim_{r \downarrow c} \mathbb{P}[D = 1| \overbrace{Z = 1}^{\text{(redundant)}}, R = r] \neq \lim_{r \uparrow c} \mathbb{P}[D = 1| \overbrace{Z = 0}^{\text{(redundant)}}, R = r]$$

- In contrast, SD was like having a perfect control variable
- Complication: no variation in D or Z away from $R = c$

Example: Class size and student achievement

Angrist and Lavy (1999, Quarterly Journal of Economics):

- Aim: estimate the effects of class size on student achievement in Isreal
- Concern: disadvantaged students are overrepresented in small classes
  - thus, bias in OLS estimates
- Solution: Public schools in Israel generally use the following rule:
  - A class should have no more than 40 students

Example: Fuzzy RD as IV

The class size rule produces a systematic but discontinuous relationship

- between the number of students in a school and the average number of students in a class

Angrist and Lavy:

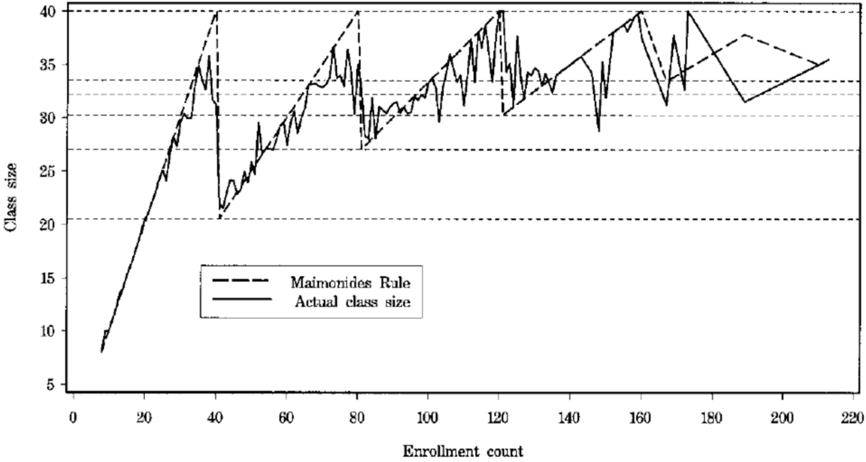
- Begin by computing the class size predicted by the school size:
  - pclassize = s_size/(int((s_size-1)/40)+1)
- then use predicted class size as an instrument for observed class size:

$$IV = \frac{cov(testscore_i, pclassize_i)}{cov(classize_i, pclassize_i)} =$$

$$\frac{reduced\ form}{first\ stage} = \frac{cov(testscore_i, pclassize_i)/var(pclassize_i)}{cov(classize_i, pclassize_i)/var(pclassize_i)}$$

Fuzzy Regression Discontinuity

Example — Class size (Angrist & Lavy, 1999)



Identification in Fuzzy Designs

Notation

- $Z = \mathbb{1}[R \leq c]$ and potential outcomes D_0, D_1 $\rightarrow$ Think of Z as whatever "incentive" for D gets applied when $R \geq c$
- Assume monotonicity (no defiers), so $D_1 \geq D_0$
- Types $T = \{at, nt, cp\}$, with $[T = cp] = [D_1 = 1, D_0 = 0]$
- Assume $\mathbb{E}[Y_d|R = r, T = t]$ and $\mathbb{P}[T = t|R = r]$ continuous at $r = c$

Result

- The limiting Wald estimand as $R \rightarrow c$ is a LATE:
  $\frac{\lim_{r\downarrow c} \mathbb{E}[Y|R=r] - \lim_{r\uparrow c} \mathbb{E}[Y|R=r]}{\lim_{r\downarrow c} \mathbb{E}[D|R=r] - \lim_{r\uparrow c} \mathbb{E}[D|R=r]} = \mathbb{E}[Y_1 - Y_0|R = c, T = cp]$
- Derivation is like LATE, but limits add some subtlety (more on this in TA session)
- Very "local" parameter - not only $R = c$, but also $T = cp$

Example: Results from class size discontinuities

	Dependent variable	
	Class size	Test scores
First stage	.776	
	(.053)	
Reduced form		-.085
		(.002)
IV estimate		-.109
		(.033)
OLS estimate		.135
		(.027)

Note: All estimates include control for family background

Interpretation and extrapolation

Fuzzy RD is IV with discontinuity as instrument
How should the corresponding LATE be interpreted?

It is informative of the average causal effects for individuals whose treatment status is shifted if we marginally change the cutoff

Whether that is policy relevant or not depends on whether the cutoff is something that one can (hypothetically) change

Motivates two types of extrapolations

- At $R = c$: effects for subpopulations other than compliers
  - Same as in usual IV
    (see e.g. Mogstad, Torgovitsky, Santos, 2018, Econometrica)
- Groups with values of the running variable away from the threshold c
  - Difficult problem, in part because R is not exogenous

Interpretation and extrapolation

Some attempts:

- Dong and Lewbel (2014): Estimate slopes of the two conditional mean functions to extrapolate away from the threshold
- Angrist and Rokkanen (2012): It is possible to extrapolate if conditional on exogenous covariates the regression functions are flat
- Bertanha & Imbens (2014): If at the threshold
  1. $E[Y_1|Type = a] = E[Y_1|Type = c]$
  2. $E[Y_0|Type = n] = E[Y_0|Type = c]$

  then it is more plausible that one can extrapolate the average effect for compliers to other subpopulations (?)

Taking RD designs to data

Taking RD designs to the data

For two reasons, one cannot only use data at the cut-off:

- No individuals with $R \geq c$
  - for whom we observe Y_0
- Usually too few individuals with $R \approx c$
  - for whom we observe Y_1

Thus, we need to use data away from cut-off to produce

- estimates of $E(Y_1 \mid R = c)$ and $E(Y_0 \mid R = c)$

To this end, we use parametric and/or non-parametric regressions

- to flexibly estimate $E(Y \mid R)$ seperately for $R \geq c$ and $R < c$

Presentation

Graphical Analyses

We can perform some visual checks and plot

- regression function $E[D|R = r]$
- regression function $E[Y|R = r]$
- regression functions $E[Z|R = r]$ for covariates that do not enter the assignment rule Z
    - more on this later
- density $f_R(r)$
    - more on this later

In all cases use estimators that do not smooth across the cutoff value.

Estimation for the Sharp case

Linear regression

The basic assumption is that $E[Y_D|R = r]$ are continuous in R
One can estimate the average effect directly in a single regression,

$$Y = \alpha + \tau D + f(R) + \epsilon$$

Problem: $f(\cdot)$ is unknown
A natural start is to estimate this with a linear spline

$$Y = \alpha + \tau D + \beta(R - c) + \gamma(R - c)D + \epsilon$$

for $c - h \leq R \leq c + h$
Covariates can be added (why?)

Estimation for the FRD Case
2SLS

The above is readily extended to the fuzzy RD case since

$$Y = \alpha_{yl} + \tau D + \beta_{yl}(R - c) \cdot \mathbf{1}[R < c] + \beta_{yr}(R - c) \cdot \mathbf{1}[R \geq c] + \epsilon$$

can be estimated with 2SLS using

- D as the endogenous regressor,
- the indicator $\mathbf{1}[R \geq c]$ as the excluded instrument

This

1. can be estimated on local neighborhood of c (f.e. $r \in [c - h, c + h]$)
2. has the advantage that other covariates can be added in straightforward manner
3. gives standard errors

Estimation for the FRD Case

2SLS

Sometimes and RD involves many (K) discontinuities
This gives K different (local) effects, but sometimes we can construct a sensible pooled estimate

- define continuous endogenous variable w
- use rules to predict w
  - directly: $pred(w)$ and use as instrument
  - indirectly: use discontinuities as instrument (Wald estimator)

Example: class size and predicted class size
$(enroll/[int((enroll-1)/40)-1])$
But combining discontinuities is like using many instruments and estimand is now a weighted average of LATEs.

Bandwidth Choice

Although RD is identified locally, estimation uses data away from the discontinuity

- Investigate the sensitivity of the inferences to bandwidth choice

Obviously, bandwidth choices affect both estimates and standard errors

But if the results are critically dependent on a particular bandwidth choice, they are clearly less credible than if they are robust to such variation in bandwidths.

Sensitivity to Bandwidth Choice

Example — Table 6, Angrist & Lavy (1999)

	5th grade						4th grade					
	Reading comprehension			Math			Reading comprehension			Math		
	+/− 5 Sample	+/− 3 Sample		+/− 5 Sample	+/− 3 Sample		+/− 5 Sample	+/− 3 Sample		+/− 5 Sample	+/− 3 Sample	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Regressors												
Class size	−.687	−.588	−.451	−.596	−.395	−.270	−.175	−.234	−.380	.018	−.118	−.247
	(.197)	(.198)	(.236)	(.254)	(.254)	(.281)	(.130)	(.157)	(.205)	(.162)	(.202)	(.234)
Percent dis-	−.464	−.452		−.433	−.416		−.350	−.372		−.291	−.323	
advantaged	(.039)	(.045)		(.050)	(.058)		(.034)	(.043)		(.043)	(.055)	
Segment 1	−5.09	−4.54	−10.7	−7.54	−6.94	−12.6	−1.62	−2.67	−6.94	−1.89	−3.57	−7.31
(enrollment 36–45)	(2.40)	(2.59	(3.19)	(3.07)	(3.34)	(3.80)	(1.77)	(2.23)	(2.90)	(2.21)	(2.87)	(3.31)
Segment 2	−1.64	−2.18	−2.96	−1.57	−2.17	−2.89	−1.52	−2.16	−3.83	−1.15	−2.50	−3.96
(enrollment 76–85)	(1.41)	(1.64)	(2.00)	(1.83)	(2.14)	(2.41)	(1.24)	(1.59)	(2.10)	(1.56)	(2.07)	(2.39)
Root MSE	7.46	7.24	8.67	9.41	9.14	10.2	6.72	6.70	8.30	8.25	8.53	9.52
N	471	302		471	302		415	265		415	265	

The table reports results from a sample of classes in schools with enrollment close to points of discontinuity. The unit of observation is the average score in the class. Standard errors are reported in parentheses. Standard errors were corrected for within-school correlation between classes. All estimates use 1[$f_{sc} < 32$] and interactions with dummies for enrollment segments as instruments for class size. Since there are three segments, there are three instruments. The models include dummies for the first two segments to control for segment main effects.

Non-parametric estimation

Nonparametric regression

- We need to estimate $\lim_{r \downarrow c} \mathbb{E}[Y|R=r]$ and $\lim_{r \uparrow c} \mathbb{E}[Y|R=r]$
- Local nonparametric methods - only use data near $R = c$
  $\rightarrow$ **Discontinuity sample** $\{i : c - h \leq R_i \leq c + h\}$ with **bandwidth** $h > 0$

Boundary bias

- Kernel regression estimators are called **local constant**
  $\rightarrow$ Regress Y_i on a constant (take the mean) for i near $R_i = c$
- **Local linear**: Regress Y_i on a constant and R_i near $R_i = c$
- This turns out to have much lower boundary bias
  $\rightarrow$ Linearity helps with adapting to curvature at the boundary

Non-parametric estimation

Sharp RD – Local Linear Regression

We are interested in the value of a regression function at the boundary of the support. Standard kernel regression

$$\widehat{\mu_l(c)} = \sum_{i:c-h<R_i<c}^{N} Y_i / \sum_{i:c-h<R_i<c}^{N} 1$$

does not work well for that case (slower convergence rates) Better rates are obtained by using local linear regression. First

$$\min_{\alpha_l,\beta_l} \sum_{i|c-h<R_i<c}^{N} (Y_i - \alpha_l - \beta_l(R_i - c))^2$$

The value of left hand limit $\mu_l(c)$ is then estimated as

$$\widehat{\mu_l(c)} = \hat{\alpha}_l + \hat{\beta}_l(c - c) = \hat{\alpha}_l$$

Similarly for right hand side

Non-parametric estimation

Fuzzy RD – Local Linear regression

Do local linear regression for both the outcome and the treatment indicator, on both sides,

$$\left(\hat{\alpha}_{yl}, \hat{\beta}_{yl}\right) = \arg\min_{\alpha_{yl},\beta_{yl}} \sum_{i:c-h\leq R_i<c} (Y_i - \alpha_{yl} - \beta_{yl}(R_i - c))^2$$

$$\left(\hat{\alpha}_{tl}, \hat{\beta}_{tl}\right) = \arg\min_{\alpha_{yl},\beta_{yl}} \sum_{i:c-h\leq R_i<c} (D_i - \alpha_{yl} - \beta_{yl}(R_i - c))^2$$

and the same on the right-side of the discontinuity $(\hat{\alpha}_{rl}, \hat{\beta}_{rl})$ and $(\hat{\alpha}_{rt}, \hat{\beta}_{rt})$
Then the FRD estimator is

$$\hat{\tau}_{FRD} = \frac{\hat{\tau}_y}{\tau_t} = \frac{\hat{\alpha}_{yr} - \hat{\alpha}_{yl}}{\hat{\alpha}_{tr} - \hat{\alpha}_{tl}}$$

Non-parametric estimation

Bandwidth

There are several options to select bandwidth

- Cross validation (Ludwig & Miller, 2007)
- Direct plug-in rules (Imbens & Kalyanaraman, 2012; Calonico, Cattaneo, and Titiunik, 2014)
  - IK: Find the bandwidth that minimize a first-order approximation of the MSE of the estimated treatment effect
  - CCT: add bias correction and derive a new MSE optimal bandwidth

CCT (and IT) provide a set of Stata (and R?) commands (rdrobust, rdbwselect, rdplot) to implement estimation in the RD design

Specification Checks

What does the RD assumption imply?

If variation in treatment near the cut-off is as good as random:

1. then pre-determined characteristics (X) should have the same distribution
   - just above and just below the cut-off
2. then the density of the assignment variable should not
   - change discontinuously around the cut-off

Thus, the distribution of X and the density of R

- are used to informally examine the validity of the RD design
  - graphically and formally

Typical specification checks

- Discontinuities in covariates
- Discontinuity in the distribution of the running variable
- Discontinuities in outcomes at values of R other than c
- Specification checks and sensitivity to bandwidth choice

Let's look at what is done in some economic applications of sharp and fuzzy RD

Applications

Practical advice when doing RD

1. Motivate validity of design
   - why individuals cannot manipulate assignment variable
2. Test validity of design, graphically and formally
   - Outcomes, covariates, and density
     - above and below cut-off
3. Show robustness of RD estimates with respect to
   - specification of f_r and f_l
   - choice of bandwidth
4. If fuzzy RD, think of it as IV

How do I find discontinuities?

There is no "recipe" that guarantees success

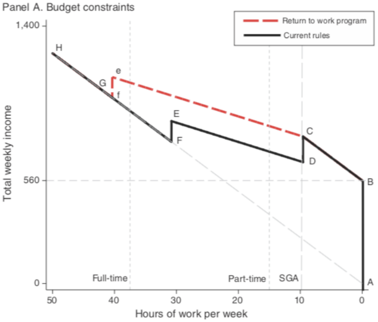
But often necessary ingredients: Detailed knowledge of

- i) the *economic mechanisms*, and
- ii) *institutions* determining treatment

May help to look at existings studies, keeping the goal in mind:

- Arbitrary rules, affecting similar individuals differently
- No manipulation of assigment variable

Example: Køstol and Mogstad (2014)



- Return-to-work reform for disability insurance (DI) in Norway
- Enacted on January 2005 increased incentives to work on DI
- Retroactive: applied to those on DI before January 1, 2004
  $\rightarrow$ Rules for those on DI after January 1, 2004 were unchanged
- Impact on labor market outcomes (participation and earnings)?

Data and Design

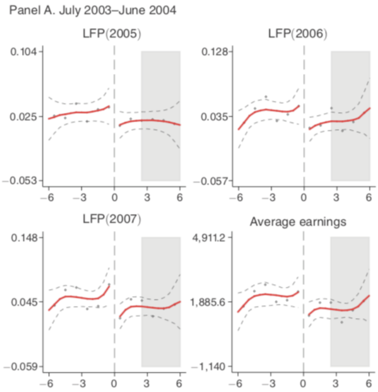
Data

- Sample includes individuals aged 18 - 49 at date of DI award
- Outcomes: participation (earn $>$ a given amount) and earnings
- Roughly 6,500 individuals awarded DI in 2003
  $\rightarrow$ Discontinuity samples below will have around 900

Design and implementation

- R is date that DI was awarded, cutoff is January 1, 2004
  $\rightarrow$ Award date is only observed monthly
- First specification uses a bandwidth of 1 month ("first difference")
- Preferred specification: 2 months - allows estimation of a slope
- Separate regression

Graphical Display of RDDs



- Bin scatters - visual evidence of treatment effect
- $D = \mathbb{1}[R < c]$ in this paper - left-hand side is the reform
- Grey area is another confounding reform - this region will not be used

Tabular Display of RDD Results

TABLE 1—PROGRAM IMPACT ON LABOR FORCE PARTICIPATION AND EARNINGS

					Comparison means [SD]	
Outcome variables:	FD	FD w/c	RD	RD w/c	Jan. and Feb. 04	Rejected applicants
LFP(2005)	0.022	0.033**	0.028	0.038	0.018	0.26
	(0.015)	(0.016)	(0.024)	(0.025)	[0.134]	[0.440]
LFP(2006)	0.031*	0.033*	0.039	0.042*	0.02	0.316
	(0.017)	(0.018)	(0.027)	(0.026)	[0.142]	[0.466]
LFP(2007)	0.054***	0.053**	0.087***	0.085***	0.034	0.316
	(0.02)	(0.022)	(0.031)	(0.031)	[0.182]	[0.466]
Average earnings	1,126**	1,247**	1,630**	1,644**	1,551	13,223
(2005–2007)	(508)	(535)	(799)	(781)	[5,033]	[21,314]

- 8.5% increase in participation of baseline of 3.4% for control
- Average earnings estimates represents > 100% increase
- Effects for RD (2 months) > FD (first difference; 1 month)
- "w/c" is with covariates - seem to not matter here, typical of RD

Threats to Identification

Manipulation

- Suppose that the reform had not been retroactively applied
  $\rightarrow$ i.e. suppose it had been announced ahead of time
- Then individuals might **manipulate** their DI claim/award
- So individuals sort to either side of January 1, 2004
  $\implies \mathbb{E}[Y_d|R = r]$ will not be smooth at c

Confounding treatments

- Suppose, in addition to $D = \mathbb{1}[R \geq c]$ we have $T = \mathbb{1}[R \geq c]$
  $\rightarrow$ e.g. R is age, a geographic boundary, time
- If T affects Y, it will confound the treatment effect of D

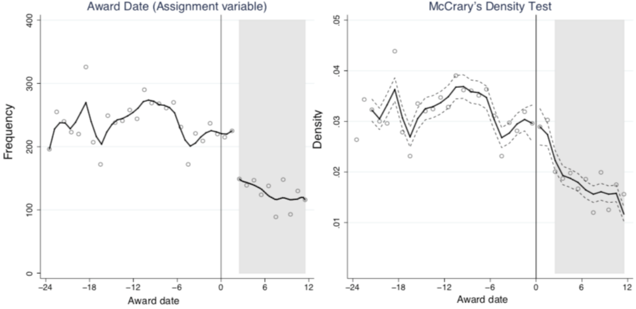
The McCrary (2008) Density Test

Implementation

- Manipulation suggests an unusually large density on one side of $R = c$
- McCrary (2008) proposed a test of this - widely used in RD papers
  $\rightarrow$ Basically doing an RD with outcome $\bar{R}_j$ and running variable M_j

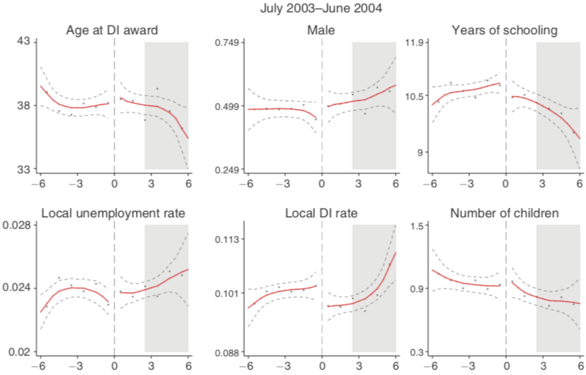
Example of a Density Test

Figure A.2: Distribution of the Assignment Variable and the McCrary Test



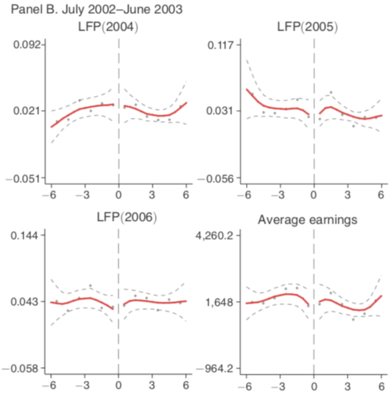
- Density test from Køstol and Mogstad (2014) - do not reject
- Three months to the right (shaded area) indicate a different policy change
- For this policy change one would likely reject based on the density test

Graphical Display of Covariate Test



- Various covariates - actual paper has even more of these
- Graphical evidence of no effect - reassures against sorting/manipulation

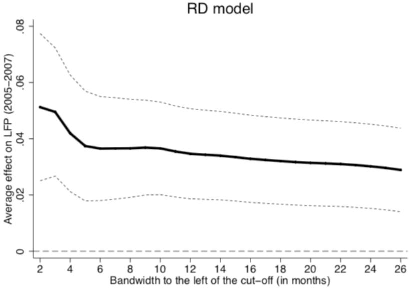
Graphical Example of Placebo Tests



- Køstol and Mogstad (2014) - same analysis a year before (no reform)
- Graphical evidence of no effect - reassures against seasonality concerns

Bandwidth Sensitivity

Figure A.3: INCREASING THE BANDWIDTH



- Sensible low-tech solution is to just try several bandwidths
- Above is Køstol and Mogstad (2014) sensitivity to left side bandwidth
- Right side bandwidth fixed due to a confounding program change
- IK/CCT approaches are reasonable starting points (IK here is 2 months)

Example of fuzzy RD: Kirkeboen et al. (2016)

Recall the discussion of IV with multiple unordered treatments
Kirkeboen analyzed that in the context of field of study choices
We now return to this paper, showing how they use discontinuities to get instrument for each choice

Example of fuzzy RD: Kirkeboen et al. (2016)

In OECD countries, most young people enroll in post-secondary education

Virtually all these students had to pick a field of study or college major

- Most OECD countries: Students enroll in specific field upon entry
- United States: Students choose major later in college

Choice of field is potentially as important as decision to enroll in college

- Earnings differentials by field rival college wage premiums (e.g. Altonji et al., 2012)

Yet, there is little evidence on the payoffs to different types of post-secondary education, including field and institution of study

Empirical challenges

This paper examines the payoff to different types of post-secondary education, including field and institution of study

Key empirical challenges:

1. Selection bias: One instrument per alternative
2. Getting the counterfactual right:
    - Several alternatives and no natural ordering
    - Researchers usually only observe the actual choice

Altonji et al. (2012) review the literature, concluding

> *"there is a long way to go on the road to credible measures of the payoffs to fields of study"*

Institutional details: Post-secondary system

The Norwegian post-secondary education sector

- 8 universities and 25 university colleges,
- all publicly funded and regulated

A typical post-secondary degree lasts 3-5 years
Field offering:

- Universities: offer a broad selection of fields
- University colleges: more vocationally oriented degrees in engineering, nursing, business administration, and teaching.

No tuition fees
All students are eligible for generous financial support (part loan/part grant)

Institutional details: Admission process

Every year in the late fall, the Ministry of Education and Research decides on the supply of slots in each field at every institution

- For many fields, demand exceeds supply

The admission process is centralized:

- Students submit up to 15 choices (field $\times$ institution)

Fields for which there is excess demand are filled based on an application score based on high school GPA

Applicants are ranked by their application score

- places are assigned according to sequential dictatorship

Institutional details: Admission process

This admission process generates a setup where

- applicants scoring above a certain threshold will receive an offer,
- similar applicants but with a marginally smaller application score will not receive an offer

These application thresholds are not perfectly foreseeable

- depend on that year's capacity
- depend on total applicant pool
- vary from year to year

Applicants are effectively randomized into different fields and institutions around these application thresholds

Our analysis below exploits these discontinuities in an instrumental variable setup

Identification: Example for fields

Course Ranking	Inst.	Field	Cutoff
1st best	A	1	57
2nd best	B	1	52
3rd best	A	2	48
4th best	A	3	45
	GPA = 49		
Local Course Ranking	Inst.	Field	Offer
Preferred	A	2	Yes
Next-best	A	3	No
	GPA = 47		
Local Course Ranking	Inst.	Field	Offer
Preferred	A	2	No
Next-best	A	3	Yes

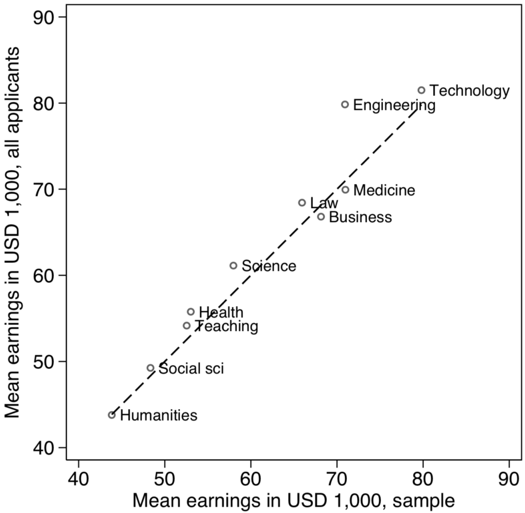
Identification: Example for institutions

Course Ranking	Inst.	Field	Cutoff
1st best	B	1	52
2nd best	A	2	48
3rd best	B	2	46
4th best	B	3	43
	GPA = 49		
Local Course Ranking	Inst.	Field	Offer
Preferred	A	2	Yes
Next-best	B	2	No
	GPA = 47		
Local Course Ranking	Inst.	Field	Offer
Preferred	A	2	No
Next-best	B	2	Yes

Summary statistics – applications

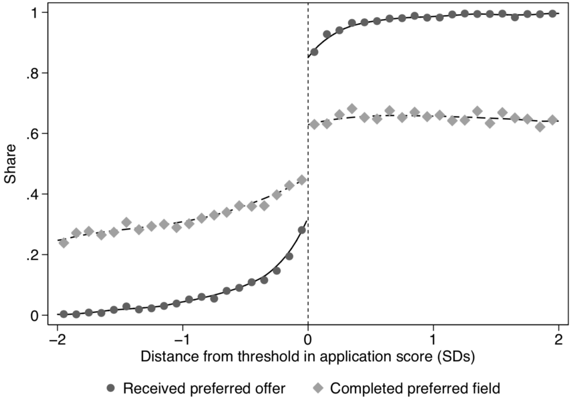
	Sample		All Applicants	
	Mean	SD	Mean	SD
# Fields ranked	3.0	(1.1)	2.2	(1.2)
# Inst. ranked	3.7	(2.4)	3.2	(2.5)
Rank of offered field	2.5	(2.0)	1.8	(1.6)
Offered rank=1	0.40		0.58	
Offered rank=2	0.25		0.15	
Offered rank=3	0.13		0.07	
No offer	0.01		0.11	
Observations	50,083		218,824	

Earnings by field: Sample vs. all applicants



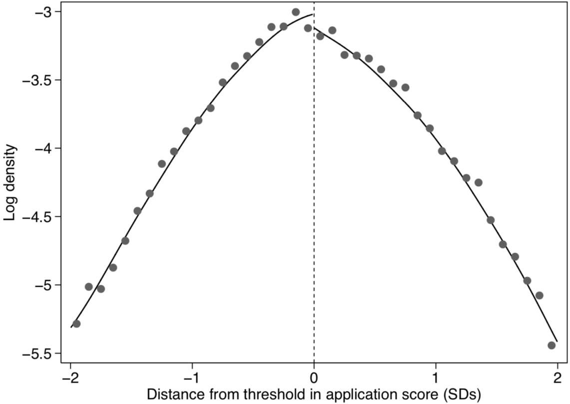
Admission cutoff and field of study

Offer and graduation rates



Admission cutoff and sorting

Bunching



Admission cutoff and sorting

Balancing

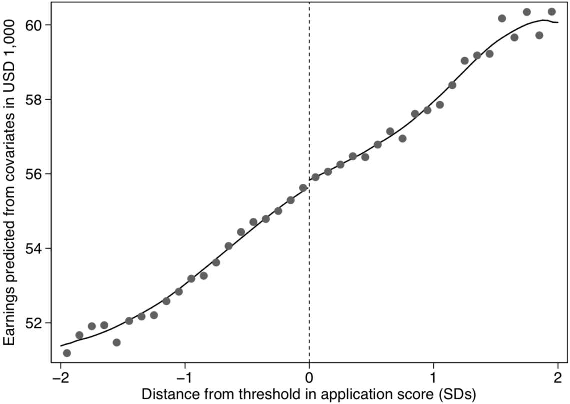
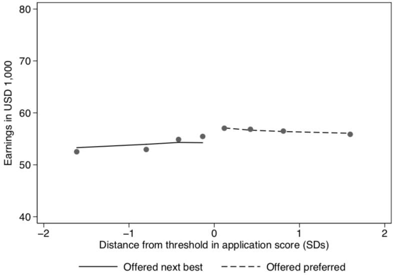


Illustration: Reduced forms

Above/below cut-off: Any field (left graph) & Specific preferred field (right graph)



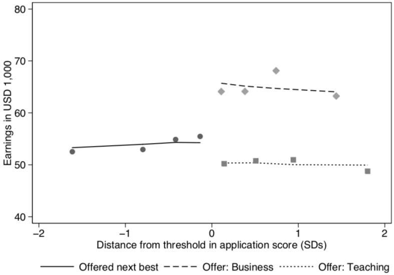
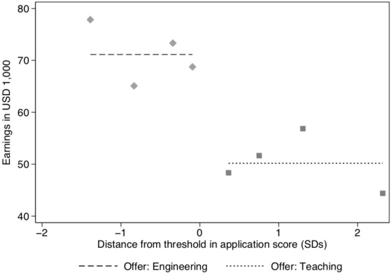
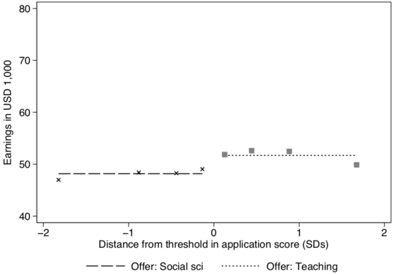


Illustration of approach: Reduced forms (con't)

Above/below cut-off by preferred and next-best field





IV model: Estimation

For every next best field k, we specify the system of equations:

$$Y = \sum_{j \neq k} \beta_{jk} D_j + X' \gamma_k + \lambda_{jk} + \epsilon$$

$$D_j = \sum_{j \neq k} \pi_{jk} Z_j + X' \psi_{jk} + \eta_{jk} + u, \quad \forall j \neq k$$

$\Longrightarrow$ A full matrix of the payoffs to field j compared to k for those who prefer j and have k as next-best field
X includes controls for background characteristics and the running variable (each side of cutoff)

Follow methods of Imbens and Rubin (1997) and Abadie (2002) to compute potential earnings levels

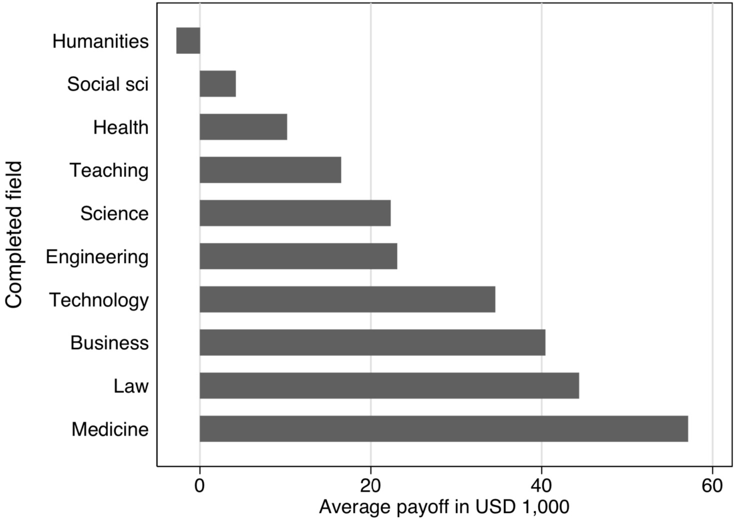
Payoffs by completed field and next best alternative

	Next best alternative (k):								
	Hum	Soc Sc	Teach	Health	Science	Eng	Tech	Bus	Law
Humanities		21*	-5	-23*	5	-39**	7	-42**	-156
		(11)	(10)	(12)	(12)	(15)	(48)	(11)	(437)
Social Sc	19**		10	-11	55**	-55**	-110	-28**	-76
	(7)		(12)	(13)	(21)	(21)	(103)	(11)	(86)
Teaching	22**	31**		2	23**	-34**	-35	-21**	23
	(5)	(8)		(7)	(9)	(13)	(37)	(7)	(128)
Health	19**	31**	8**		29**	-28**	-43**	-17**	-55
	(6)	(8)	(3)		(8)	(10)	(21)	(4)	(98)
Science	54**	70**	39**	30**		-2	17	-5	148
	(18)	(22)	(14)	(12)		(15)	(18)	(11)	(276)
Engineering	60	-6	75**	0	52**		-46	-13	-58
	(51)	(58)	(38)	(16)	(21)		(44)	(24)	(167)
Technology	42**	59**	22*	32**	68**	-6		7	-53
	(11)	(10)	(12)	(10)	(10)	(12)		(9)	(148)
Business	48**	62**	31**	30**	58**	-3	29*		4
	(11)	(12)	(9)	(11)	(10)	(13)	(16)		(83)
Law	46**	56**	37**	21*	40**	-28	-16	-1	
	(7)	(8)	(12)	(11)	(10)	(18)	(18)	(9)	
Medicine	83**	79**	63**	46**	81**	21	40**	23**	15
	(10)	(11)	(9)	(7)	(10)	(21)	(12)	(9)	(84)
Average y^k	24	40	44	49	28	89	73	71	104
Observations	8,391	11,030	10,987	3,269	6,422	3,085	1,245	4,403	1,251

▸ Reduced form

Average payoffs: 2SLS estimates

Average payoffs and counterfactual earnings by chosen field



Interpreting payoffs

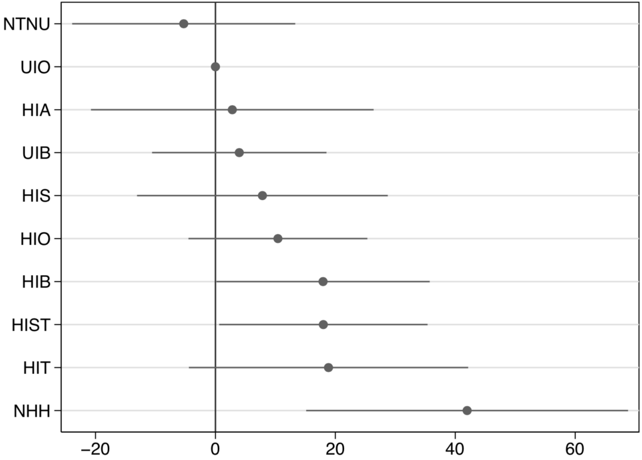
We interpret the payoffs as measures of earnings gains from completing one field of study as compared to another

- May not only arise from occupational specificity of human capital

Alternative explanations we explore and reject:

- Payoffs to field of study reflect differences in institutional quality
- Payoffs to field of study reflect differences in quality of peer groups
- Payoffs to field of study reflect labor market experience at the time we measure earnings

Estimated payoffs to largest institutions



Role of next-best alternatives: Alternative assumptions

Part (iii) of Proposition 2 motivated and guided our empirical analysis:

- Using information on next-best alternatives to identify payoffs

Proposition 2 also pointed out alternatives:

- Assuming constant effects
- Imposing strong restrictions on preferences

We test and reject both these alternatives

- Biases are economically substantial

Reduced form payoffs

	Next best alternative (k):								
	Hum	Soc Sc	Teach	Health	Science	Eng	Tech	Bus	Law
Humanities		-1.5	-1.5	-4.9*	-5.4*	-13.6**	3.3	-13.1**	-33.5
		(1.6)	(1.7)	(2.9)	(2.7)	(5.7)	(12.3)	(3.7)	(29.6)
Social Sc	0.9		1.7	-1.3	4.5	-14.2*	-5.4	-3.8	-14.6**
	(1.2)		(2.2)	(2.3)	(3.0)	(8.3)	(8.3)	(2.8)	(5.4)
Teaching	2.1**	3.0**		2.2	-0.1	-8.8**	-4.4	-2.9	1.7
	(0.9)	(1.1)		(1.4)	(1.8)	(3.6)	(6.4)	(1.9)	(10.3)
Health	-0.1	0.4	0.4		-0.8	-6.7**	-9.4**	-4.5**	-4.0
	(1.0)	(0.7)	(0.5)		(1.0)	(1.9)	(4.1)	(1.1)	(4.6)
Science	6.5**	2.9	2.9	4.6*		-4.2	7.6*	-4.0**	13.4
	(3.0)	(2.1)	(1.9)	(2.5)		(3.8)	(4.4)	(1.9)	(13.9)
Engineering	7.7	6.9	17.2**	1.1	3.8		-3.6	-1.3	-0.9
	(6.8)	(8.0)	(5.8)	(6.3)	(2.8)		(5.9)	(5.7)	(16.1)
Technology	9.1**	13.9**	6.9	11.2**	8.0**	0.9		4.7	1.1
	(3.4)	(3.0)	(4.6)	(3.6)	(1.7)	(1.9)		(4.3)	(7.9)
Business	4.4**	7.3**	1.6	1.5	4.0**	-3.5	7.2		12.7*
	(1.6)	(1.6)	(1.7)	(2.8)	(1.6)	(3.1)	(5.6)		(6.6)
Law	9.8**	6.8**	11.0**	3.9	4.9*	-10.0	-5.0	-0.6	
	(1.9)	(1.2)	(5.5)	(3.0)	(2.9)	(8.1)	(6.2)	(3.0)	
Medicine	32.8**	21.7**	27.8**	19.9**	22.0**	11.4	17.4**	12.5**	21.2**
	(5.2)	(3.6)	(6.3)	(2.8)	(2.8)	(10.4)	(3.9)	(5.7)	(8.7)

Back