corinna.birner@stud-mail.uni-wuerzburg.de
max.mueller@stud-mail.uni-wuerzburg.de

Understanding Metrics based on Mastering Metrics

Corinna Birner & Max Müller

University of Würzburg

July 2, 2020

Chapter 4: Regression Discontinuity Designs

Overview

Today we will continue our journey on the path from cause to effect. Therefore, we will discuss regression discontinuity (RD) designs as one possible way of finding causal relations in our data. Our topics will be the following:

Minimum Legal Drinking Age

Cut off difference

Whats a RDD?

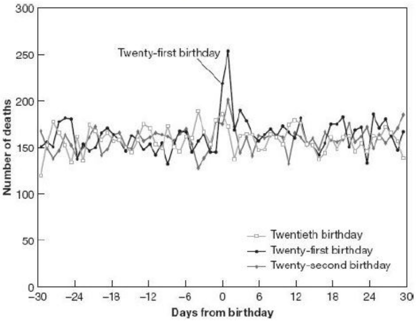
- We use rules, even arbitrary ones, to exploit certain circumstances
- E.g. The State of California limits elementary school class size to 32 students; 33 is one too many.
- 32 in that case would be our threshold.
- Regression discontinuity (RD) designs exploit discontinuities in policy assignment.
- Assume that units on different sides of the discontinuity are similar. Their treatment status differs only because of the institutional setup, and therefore differences in outcomes can be attributed to the different treatment status.

Minimum Legal Drinking Age

- Lets take a look at one of the most important thresholds for young americans: If they are over 21, they can drink legally
- This experiment emerges from the fact that a small change in age (measured in months or even days) generates a big change in legal access.
- The difference a day makes can be seen in Figure 4.1, which plots the relationship between birthdays and funerals.
- So we could assume that the individuals only differ in their legal drinking access one day before and one day after turning 21.

Minimum Legal Drinking Age

FIGURE 4.1
Birthdays and funerals



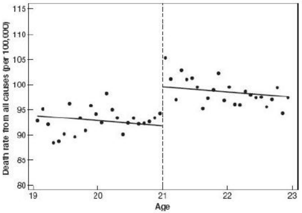
- Shows the number of deaths among Americans aged 20–22 between 1997 and 2003.
- Deaths here are plotted by day, relative to birthdays, which are labeled as day 0.

Birthday Comparisons

- Mortality risk shoots up on and immediately following a twenty-first birthday
- Does not happen around the twentieth and twenty-second birthdays
- So there has to be something special around the 21st Birthday that influences death rates.

Cut-Off

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 minimum legal drinking age (MLDA) cutoff).

- This figure plots death rates (measured as deaths per 100,000 persons per year) by month of age (defined as 30- day intervals), centered around the twenty-first birthday.
- The X-axis extends 2 years in either direction, and each dot in the figure is the death rate in one monthly interval.

Cut-Off

- Few rates to the left of the age-21 cutoff are above 95.
- At ages over 21, however, death rates shift up, and few of those to the right of the age-21 cutoff are below 95.
- Extrapolating the trend line drawn to the left of the cutoff, we might have expected an age-21 death rate of about 92, while the trend line to the right of 21 starts markedly higher, at around 99.
- Jump in trend lines at age 21 illustrates the subject of this chapter, regression discontinuity designs.

The Treatment Variable

- RD is based on the idea that rigid rules create valuable experiments.
- Causal question here is the effect of legal access to alcohol on death rates.
- The treatment variable in this case can be written as:

$$D_a = \begin{cases} 1 \text{ } if \text{ } a \geq 21 \\ 0 \text{ } if \text{ } a < 21 \end{cases}$$

Features of RD Designs

- This representation highlights two signal features of RD designs:
  - Treatment status is a deterministic function of a, so that once we know a, we know D_a.
  - Treatment status is a discontinuous function of a, because no matter how close a gets to the cutoff, D_a remains unchanged until the cutoff is reached.
- The variable that determines treatment, age in this case, is called the running variable.
- 2 Kinds of RDD´s:
  - Sharp RD designs: Treatment switches cleanly off or on as the running variable passes a cutoff.
  - Fuzzy RD: The probability or intensity of treatment jumps at a cutoff.

Regression in RD

- In this case it's a sharp RD, because the MLDA is a sharp function of age
- Mortality clearly changes with the running variable, a, for reasons unrelated to the MLDA.
- A simple RD analysis of the MLDA estimates causal effects using a regression like:

$$\bar{M}_a = \alpha + \rho D_a + \gamma a + \epsilon_a$$

- $\bar{M}_a$ is the death rate in month a, includes the treatment dummy, D_a, and linear control for age in months.

The Regression Framework

- Fitted values from this regression produce the lines drawn in Figure 4.2.
- The negative slope, captured by γ, reflects smoothly declining death rates among young people as they mature.
- The parameter ρ captures the jump in deaths at age 21.
- This Regression generates an estimate of ρ equal to 7.7.
- When cast against average death rates of around 95, this estimate indicates a substantial increase in risk at the MLDA cutoff.

Control for other Things?

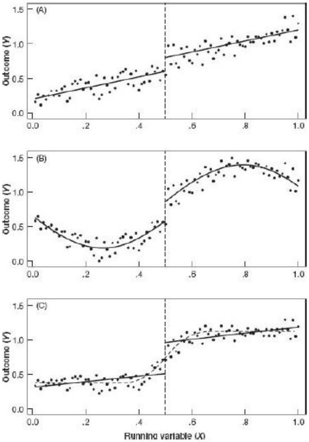
- Should we not control for other things?
- The OVB formula tells us that the difference between the estimate of ρ in this short regression and the results any longer regression might produce depend on the correlation between variables added to the long regression and D_a.
- But: D_a is determined solely by a.
- So we can be sure that no OVB afflicts this short regression.

Regression in RD

- Although treatment isn´t randomly assigned, we know where it comes from.
- Treatment is determined by the running variable.
- RD uses regression methods to estimate causal effects, but we don't need matching and regression strategies which are based on treatment-control comparisons conditional on covariate values.
- In this case, only the running variable, the changes in access to alcohol for young people, matters.

Problem of Nonlinearity

FIGURE 4.3
RD in action, three ways



- Panel A: relationship between the running variable (X) and the outcome (Y) is linear, clear jump at the cutoff
- Panel B similar, the relationship between average Y and X is nonlinear.
- Panel C shows nonlinearity mistaken for a discontinuity.

Public-Private Face-Off

- Two strategies reduce the likelihood of RD mistakes
- First: Nonlinear modeling strategy
- Nonlinearities in an RD framework are typically modeled using polynomial functions of the running variable.
- We modify our regression, if we think our model is non linear.

Modify the regression

- Figure 4.2: possibility of mild curvature in the relationship Between M_a and a, at least for the points to the right of the cutoff.
- Extension that captures this curvature uses quadratic instead of linear control for the running variable.
- The RD model with quadratic running variable control becomes:

$$\bar{M}_a = \alpha + \rho D_a + \gamma_1 a + \gamma_2 a^2 + \epsilon_a$$

- Where $\gamma_1 a + \gamma_2 a^2$ is a quadratic function of age

Modification

- A related modification allows for different running variable Coefficients to the left and right of the cutoff.
- To make the model with interactions easier to interpret, we center the running variable by subtracting the cutoff, a_0.
- Replacing a by $a - a_0$ (here, $a_0 = 21$), and adding an interaction term, $(a - a_0)D_a$, the RD model becomes:

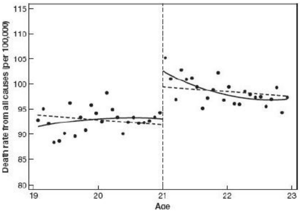
$$\bar{M}_a = \alpha + \rho D_a + \gamma(a - a_0) + \delta[(a - a_0)D_a] + \epsilon_a$$

Model with interaction terms

- An implication of the model with interaction terms is that away from the a_0 cutoff, the MLDA treatment effect is given by $\rho + \delta(a - a0)$.
- Estimates away from the cutoff constitute a bold extrapolation, there is no world in which drinking is illegal with 24 or legal with 12.
- It seems reasonable to say that those just under 21 provide a good counterfactual comparison for those just over 21.
- So estimates of the parameter ρ (the causal effect right at the cutoff) are most reliable.

Comparison of functions

FIGURE 4.4
Quadratic control in an RD design



Notes: This figure plots death rates from all causes against age in months. Dashed lines in the figure show fitted values from a regression of death rates on an over-21 dummy and age in months. The solid lines plot fitted values from a regression of mortality on an over-21 dummy

- Curvature, mildly concave to the left of age 21 and convex thereafter.
- This model generates a larger estimate of the MLDA effect at the cutoff than does a linear model, 9.5 deaths per 100,000.

Comparison

- Comparison: more elaborate model seems to give a better fit than the simple model: Death rates jump sharply at age 21, but then recover somewhat in the first few months after a twenty-first birthday.
- Which version is better?: simple RD model seems flexible enough to capture effects right at the cutoff.
- But: The fancier version fits the spike in death rates near twenty-first birthdays, while also capturing the subsequent partial recovery in death rates.
- How convincing is the argument that the jump in Figure 4.4 is indeed due to drinking?
- few people die from alcohol poisoning alone, more deaths from alcohol-related diseases
- But: alcohol is closely tied to motor vehicle accidents (MVA), the number-one killer of young people.

Estimates

TABLE 4.1
Sharp RD estimates of MLDA effects on mortality

Dependent variable	Ages 19–22 (1)	Ages 19–22 (2)	Ages 20–21 (3)	Ages 20–21 (4)
All deaths	7.66	9.55	9.75	9.61
	(1.51)	(1.83)	(2.06)	(2.29)
Motor vehicle accidents	4.53	4.66	4.76	5.89
	(.72)	(1.09)	(1.08)	(1.33)
Suicide	1.79	1.81	1.72	1.30
	(.50)	(.78)	(.73)	(1.14)
Homicide	.10	.20	.16	−.45
	(.45)	(.50)	(.59)	(.93)
Other external causes	.84	1.80	1.41	1.63
	(.42)	(.56)	(.59)	(.75)
All internal causes	.39	1.07	1.69	1.25
	(.54)	(.80)	(.74)	(1.01)
Alcohol-related causes	.44	.80	.74	1.03
	(.21)	(.32)	(.33)	(.41)
Controls	age	age, age^2, interacted with over-21	age	age, age^2, interacted with over-21
Sample size	48	48	24	24

Notes: This table reports coefficients on an over-21 dummy from regressions of month-of-age-specific death rates by cause on an over-21 dummy and linear or interacted quadratic age

- second row reveals strong effects of legal drinking on MVA fatalities, effects large enough to account for most of the excess deaths related to the MLDA.
- Effects from direct alcohol poisoning appear to be modest

Alcohol causes death?

- Table 4.1 supports the MLDA story, showing clear effects for causes most likely attributable to alcohol but little evidence of an increase due to internal causes.
- Not much of a jump in deaths due to internal causes, while the standard errors suggest that the small jump in internal deaths is likely due to chance.
- straightforward regression estimation is also called parametric RD.

Close to the boundary

- Second RD strategy: Nonparametric RD for the small set of points close to the boundary, nonlinear trends need not concern us at all.
- Suggests an approach that compares averages in a narrow window just to the left and just to the right of the cutoff.
- A drawback here is that if the window is very narrow, there are few observations left, meaning the resulting estimates are likely to be too imprecise to be useful.

Nonparametric RD

- Nonparametric RD amounts to estimating the former regression in a narrow window around the cutoff.

  $\mathrm{M}_a = \alpha + \rho D_a + \gamma a + \epsilon_a$;

  in a sample such that $\mathrm{a}_0 - b \leq a \leq a_0 + b$
- The parameter b describes the width of the window and is called a bandwidth.
- Table 4.1 can be seen as nonparametric RD with a bandwidth equal to 2 years of age for the estimates reported in columns (1) and (2) and a bandwidth half as large (that is, including only ages 20–21 instead of 19–22 for the estimates shown in columns (3) and (4).

Parametric vs. Non-parametric approach

- Parametric tests: assume underlying statistical distributions (funtional form) in the data - Therefore, several conditions of validity must be met so that the result of a parametric test is reliable.
- Non-parametric:Do not rely on any distribution and can always be applied.
- We need to be aware of the assumptions associated with a parametric procedure.
- E.g. in parametric tests, we assume Effect homogeneity, which means that a added control variable has the same effect for every part of the distribution. (Treated individual)

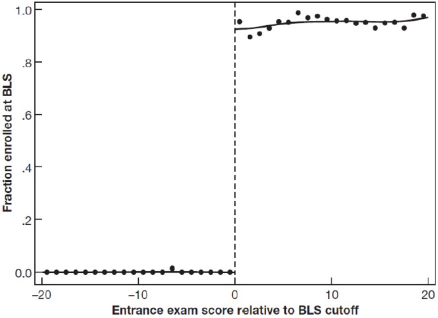
Peer effects in exam schools

- In bigger cities in the US, the public school systems include a few very selective exam schools
- in Boston for example the John D. O'Bryant School, Boston Latin Academy, Boston Latin School
- the exam schools choose students based on competitive admission tests
- What is the effect of going to such an exam school?
- Problem: Peer Effect: intelligent students benefit from studying with intelligent students

Exam Schools - application for an RDD?

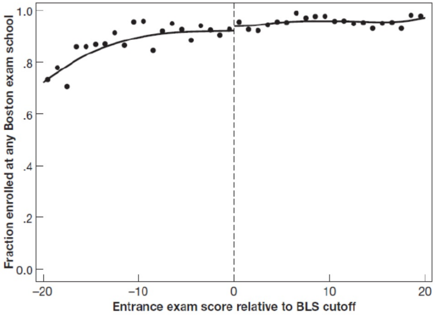
- the discrete nature of exam school admission policies creates a natural experiment
- among applicants with scores close to admission cutoffs, whether an applicant falls to the left ot the right of the cutoff might be as good as randomly assigned
- with the sharp cut-off, it could be a good application for an RD
- But: some admitted student choose to go elsewhere while many of the rejected at one exam school end up at another
- When discontiniuities change treatment probablities or average characteristics instead of flicking a simple on-off switch, the RD design is said to be fuzzy

The exam school treatment



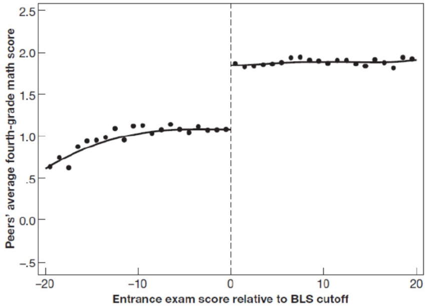
- BLS is the most prestigious exam school
- sample for the figures are applicants for BLS with score close to the entrance cutoff
- where do applicants who miss the cutoff go?

The exam school treatment



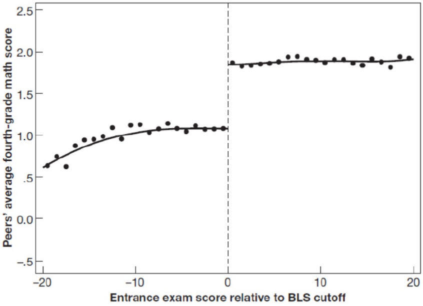
- Figure shows enrollment at any Boston exam school
- most students that miss BLS cut-off end up at another exam school
- Can we therefore only compare two schools?

The exam school treatment



- one controversial question in education research are peer effects
- to what extend does the quality of fellow students contribute the own success?

The exam school treatment



- exam schools induce an experiment in peer quality
- applicants who qualify for an exam school attend school with much higher achieving peers
- figure shows peer quality around the BLS cut-off

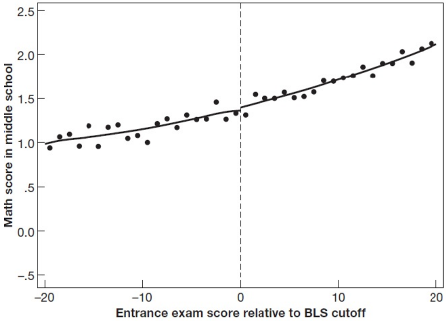
Fuzzy RD

- in a fuzzy RD research design, applicants who cross a threshold are exposed to a more intense treatment or the probability of being treated changes
- whereas in a sharp RD research design treatment switches cleanly on or off at the cut-off
- What is the effect of average schoolmate ability X_i on achievement Y_i of student i?

Peer Effect

- A regression for Boston exam schools would be:
- $$Y_i = \theta_0 + \theta_1 \bar{X}_{(i)} + \theta_2 X_i + u_i$$
- the estimated coefficient on peer quality is around 0.25
- this means that a one standard deviation increase in the ability of middle school peers is associated with a 0.25σ gain in middle school achievement
- at the cut-off in our exam school example, peer quality jumps substantially

The Peer Effect



- figure plots seventh and eight grade math scores against exam school admission score for applicants close to cut-off
- applicants close to cut-off are exposed to much stronger peer group but this exposure generates no parallel jump in applicant's middle school achievement

Peer Quality

- in order to figure out how big the jump of middle school grades is, we estimate:
- $$Y_i = \alpha_0 + \rho D_i + \beta_o R_i + \epsilon_{0i}$$
- D_i is a dummy indicating applicants who qualify and R_i is the running variable of test scores
- produces an estimate of $\rho = -.02$ (with standard error .10)
- the equation is the reduced form for a 2SLS setup where the endogenous variable is the average peer quality $\bar{X}_{(i)}$

Peer Quality - RD with IV

- the first-stage equation would be:

$$\bar{X}_{(i)} = \alpha_1 + \phi D_i + \beta_1 R_i + \epsilon_{1i}$$

- the second-stage equation can be written as:

$$Y_i = \alpha_2 + \lambda \hat{X}_{(i)} + \beta_2 R_i + \epsilon_{2i}$$

- IV-assumption: exam-school qualification has no direct effect on test scores but influences achievement, if at all, through peer quality

exam schools - Results

- the 2SLS estimate of λ is -.023 (with standard error of .132), doesn't differ significantly from zero
- it seems there is no effect of peer quality on school achievement
- but what about our exclusion restriction?

exams schools, peer quality and exclusion restriction

- exam schools might differ from public schools in other ways: teacher quality, college placement courses, etc.
- these differences should all have a positive effect on achievement
- the OVB associated with 2SLS estimates of exam school peer effect is positive
- Conclusion: if anything, the estimated peer effect is too large

In a nutshell

- the RD design exploits abrupt changes in treatment status that arise when treatment is determined by a cut-off
- we need to know the relationship between the running variable and potential outcomes in the absence of treatment
- finding a perfect control strategy is very hard but we can gain confidence that our strategy is good when RD estimates remain similar as we change details of the RD model
- a sharp RD is when treatment itself switches on or off at a cut-off whereas in fuzzy designs the probability or intensity of treatment jumps
- in fuzzy designs a dummy for clearing the cut-off becomes an instrument and the fuzzy design is analyzed by 2SLS

Thank you very much for listening