Regression Discontinuity Designs

Xiaoyang Ye
xiaoyang.ye@princeton.edu

Example: Angrist & Lavy (1999)

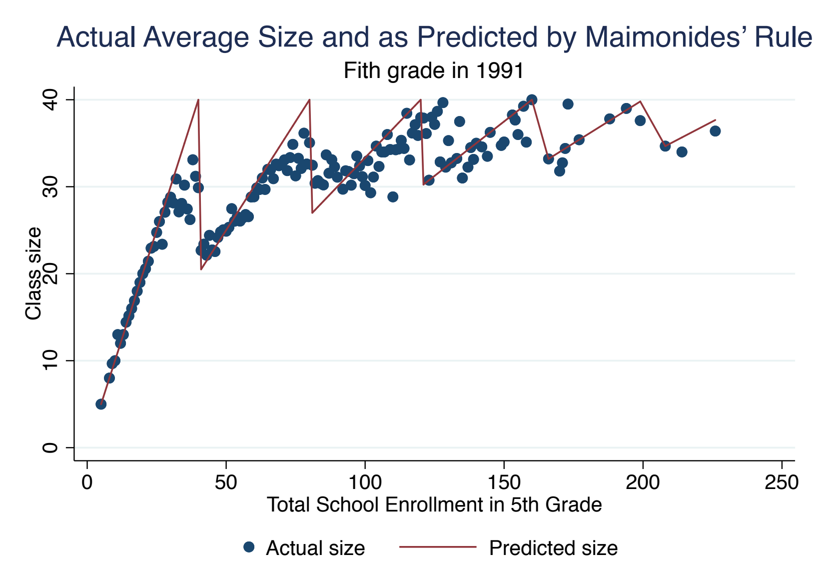
Effects of class size

USING MAIMONIDES' RULE TO ESTIMATE THE EFFECT OF CLASS SIZE ON SCHOLASTIC ACHIEVEMENT*

JOSHUA D. ANGRIST AND VICTOR LAVY

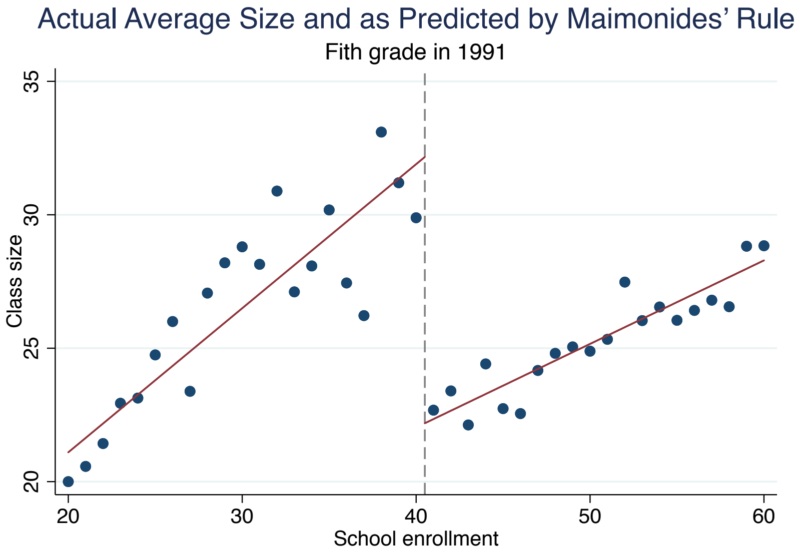
The twelfth century rabbinic scholar Maimonides proposed a maximum class size of 40. This same maximum induces a nonlinear and nonmonotonic relationship between grade enrollment and class size in Israeli public schools today. Maimonides' rule of 40 is used here to construct instrumental variables estimates of effects of class size on test scores. The resulting identification strategy can be viewed as an application of Donald Campbell's regression-discontinuity design to the class-size question. The estimates show that reducing class size induces a significant and substantial increase in test scores for fourth and fifth graders, although not for third graders.

Maimonides (1138-1204)

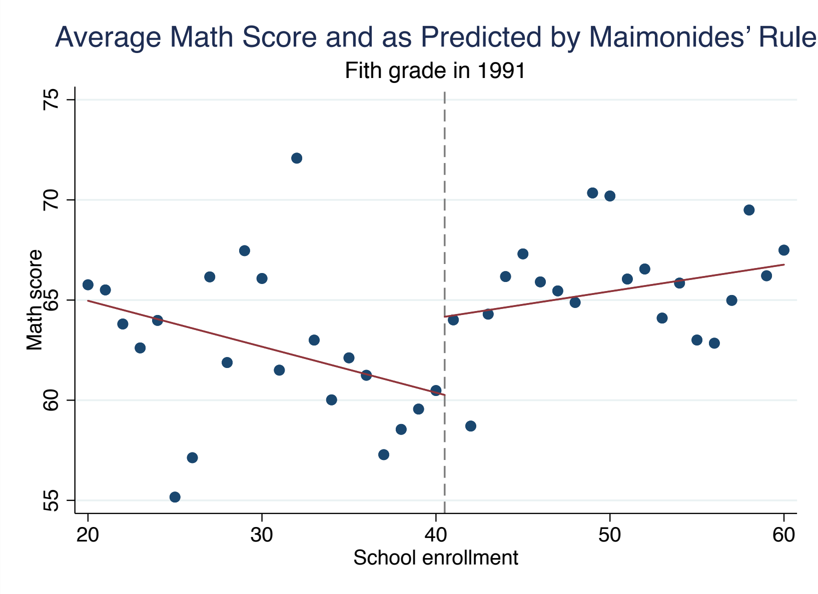


Regression Discontinuity

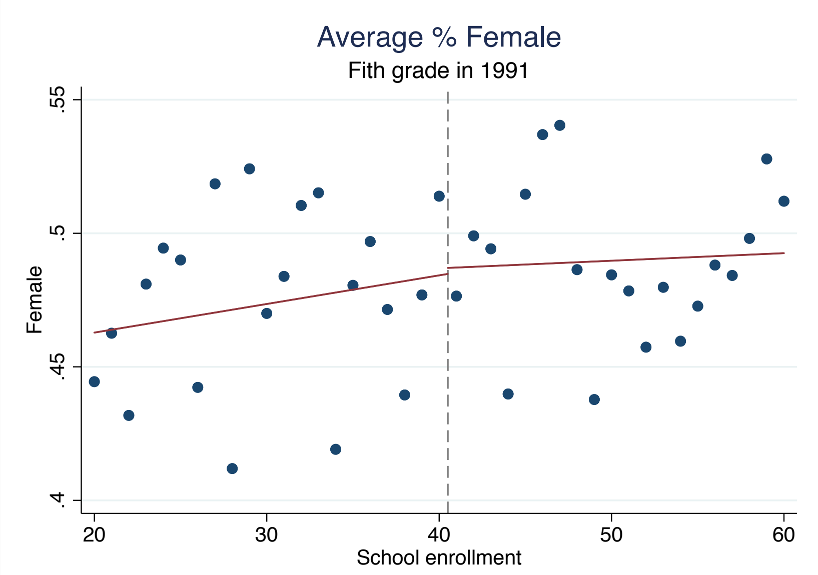
First stage: Y = treatment



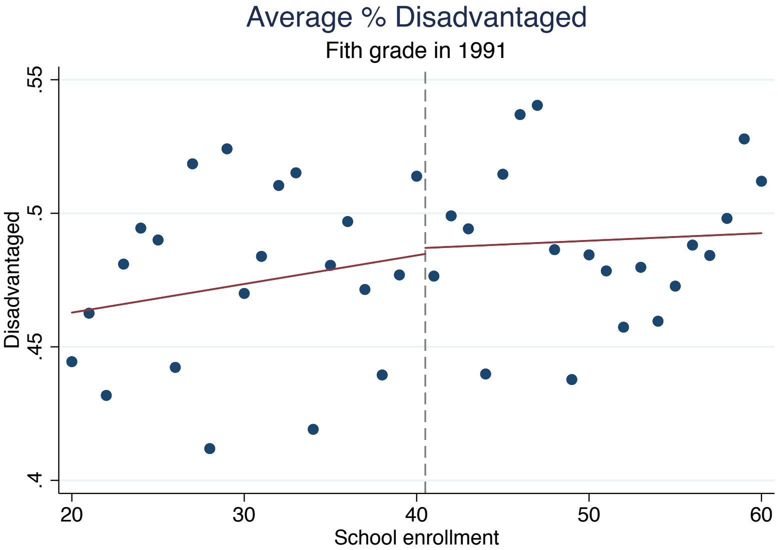
Second stage: Y = outcome



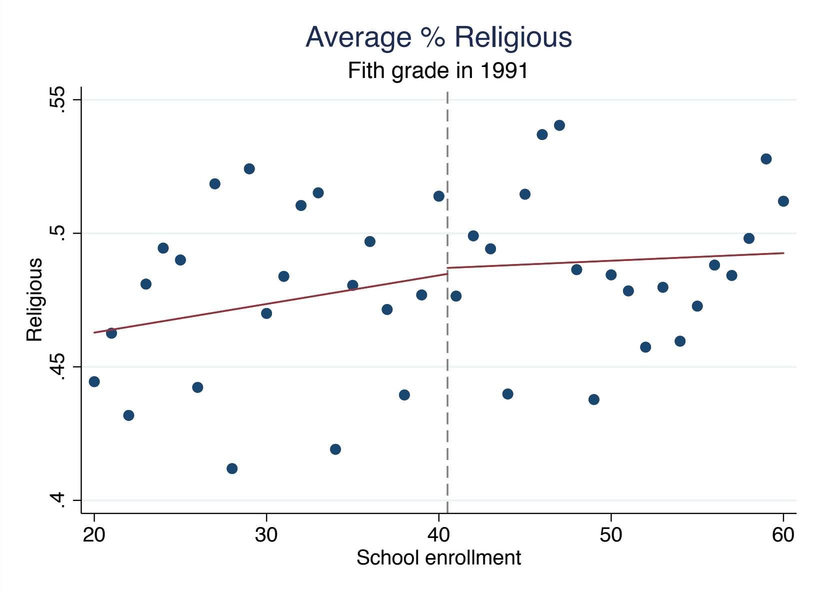
Balance test (1)



Balance test (2)



Balance test (3)



Introduction

Potential outcomes

- *Factual* vs. *Counterfactual*

$$Y_i = T_i \cdot Y_i(1) + (1 - T_i) \cdot Y_i(0)$$

  - T_i: a dummy variable indicating whether individual i receives treatment ($T_i = 1$) or not ($T_i = 0$)
  - $Y_i(1)$: the outcome of individual i if she receives treatment
  - $Y_i(0)$: the outcome of individual i if she does not receive treatment

- A valid causality question must involve well-defined causes (treatments, manipulations), and the counterfactuals should be unambiguously defined.

Fundamental problem of causal inference

- *Individual treatment effect*

$$\tau_i = Y_i(1) - Y_i(0)$$

- Causality is defined by potential outcomes, not by realized (observed) outcomes
- We can only observe one of the two potential outcomes
    - ▷ Missing data problem: Any statistical method dealing with treatment effects necessarily imputes the counterfactual part of the data.

Selection bias in observed outcomes

- Holland (1986):

$$
\begin{aligned}
&\mathbf{E}[Y_i(1)|T_i = 1] - \mathbf{E}[Y_i(0)|T_i = 0] \\
&= \underbrace{\mathbf{E}[Y_i(1)|T_i = 1] - \mathbf{E}[Y_i(0)|T_i = 1]}_{\tau_{\text{ATT}}} + \underbrace{\mathbf{E}[Y_i(0)|T_i = 1] - \mathbf{E}[Y_i(0)|T_i = 0]}_{\text{selection bias}}
\end{aligned}
$$

- Roy model:

$$
\begin{aligned}
\text{Potential Outcomes:} \quad & Y_i(0) = \mathbf{X}_i\beta(0) + u_i(0) \\
& Y_i(1) = \mathbf{X}_i\beta(1) + u_i(1) \\
\text{Selection/Assignment Mechanism:} \quad & \mathbf{1}_{\{T_i=1\}} = F(\mathbf{X}_i\gamma) + \epsilon_i
\end{aligned}
$$

  - ▷ The identification is:

$$\mathbf{X}_i \perp (u_i(0), u_i(1), \epsilon_i)$$

Causal inference designs

1. By knowledge of **Assignment Mechanism**
   - Random assignment (RCT)
   - Regression discontinuity (RD)
2. By **Self-Selection**
   - Difference-in-differences (DID)
     - Influence of "other factors" fixed
   - Selection on unobservables and instrumental variables (IV)
     - Conditional on covariates, instrument "as good as randomly assigned" (uncorrelated with potential outcomes)
     - Another structural approach: Heckman selection model
   - Selection on observables and matching (Matching)
     - Conditional on covariates, treatment "as good as randomly assigned"

Sharp Regression Discontinuity

Outline

- RD Designs
  - Sharp, fuzzy, kink, fuzzy kink
  - Analogies with experiments
  - Multi-cutoff, multi-variable
- Graphical presentation and falsification

Four facts of RD Designs

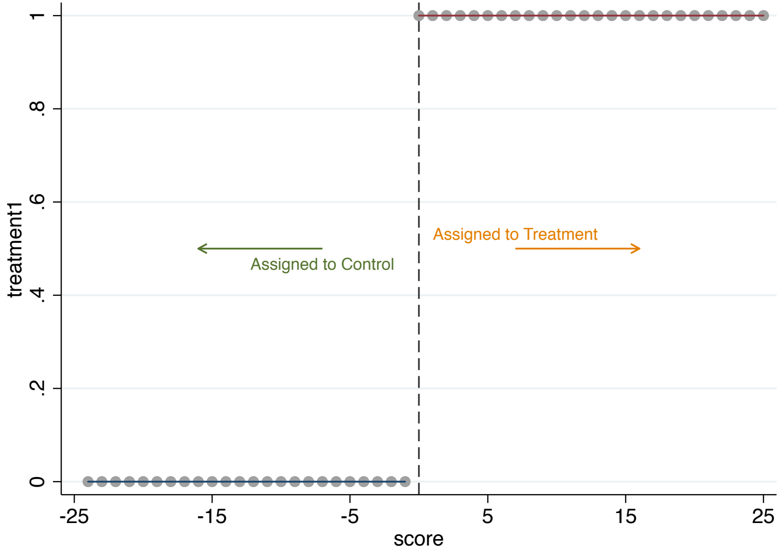
1. Simple and objective: Requires little information, if design available
2. Might be viewed as a "local" randomized trial
3. Easy to falsify, easy to interpret
4. *Careful*: very local!

Score, cutoff, treatment

- Units receive a score (“running variable”)
- A treatment is assigned based on the score and a known cutoff
- The treatment
  - is given to units whose score is greater than the cutoff
  - is withheld from units whose score is less than the cutoff
- Under some assumptions, the abrupt change in the probability of treatment assignment allows us to learn about effect of treatment

Sharp Regression Discontinuity Design

Treatment assignment



Sharp Regression Discontinuity Design

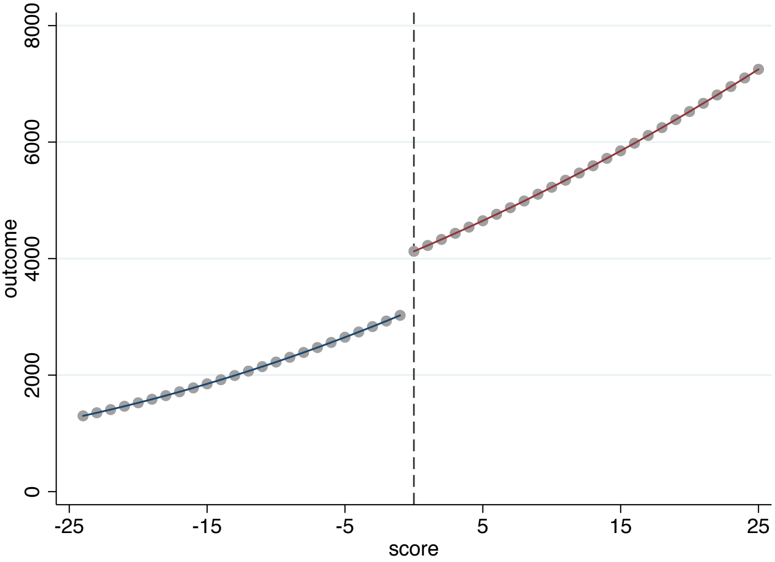
Fundamental problem of causal inference

- n units, indexed by $i = 1, 2, ..., n$
- Unit's score is X_i, treatment is $T_i = 1(X_i \geq \bar{x})$
- Two potential outcomes:
  - ▷ $Y_i(1)$: outcome that would be observed if i received treatment
  - ▷ $Y_i(0)$: outcome that would be observed if i received control
- The *observed* outcome:

$$Y_i = \begin{cases} Y_i(0) & \textit{if} \quad X_i < \bar{x} \\ Y_i(1) & \textit{if} \quad X_i \geq \bar{x} \end{cases}$$

Sharp Regression Discontinuity Design

Outcome



Sharp Regression Discontinuity Design

Treatment effect

- A special situation occurs at the cutoff $X = \bar{x}$, the only point at which we may "almost" observe both curves
- Two groups of units:
  - with score equal to $\bar{x}$, $Xi = \bar{x}$ $\rightarrow$ treated
  - with with score barely below $\bar{x}$, $X = \bar{x} - \varepsilon$ $\rightarrow$control

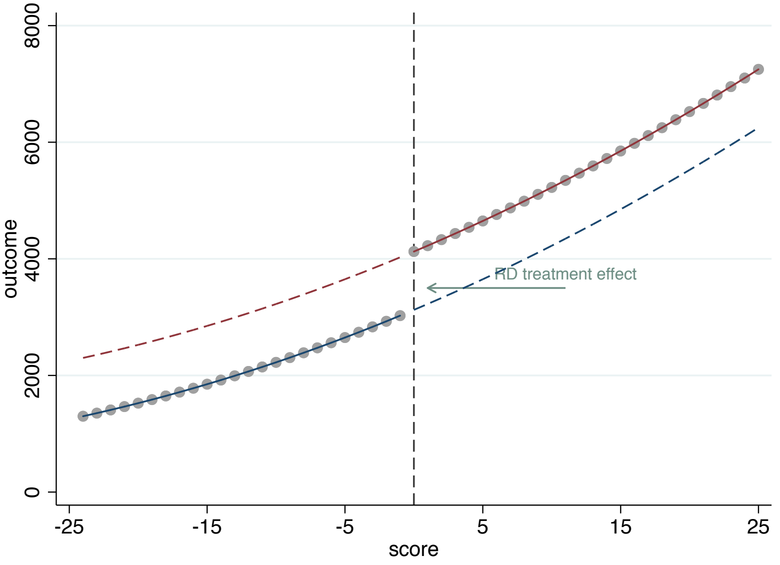
Sharp Regression Discontinuity Design

Treatment effect

- A special situation occurs at the cutoff $X = \bar{x}$, the only point at which we may "almost" observe both curves
- Two groups of units:
  - ▷ with score equal to $\bar{x}$, $Xi = \bar{x}$ $\rightarrow$ treated
  - ▷ with with score barely below $\bar{x}$, $X = \bar{x} - \varepsilon$ $\rightarrow$control
- Local randomization
  - ▷ Yet if values of the average potential outcomes at $\bar{x}$ are not abruptly different from their values at points near $\bar{x}$
  - ▷ these two sets of units would be identical except for their treatment status
- Local average treatment effect: Vertical distance at $\bar{x}$

Sharp Regression Discontinuity Design

Treatment effect



RCT vs. RD

- RCT (Experimental design)
  - Treatment
    - $T_i \in \{0, 1\}$, T_i independent of $(Y_i(0), Y_i(1), X_i)$
  - Average treatment effect (ATE)

$$\tau_{\text{ATE}} = \mathbf{E}[Y_i(1) - Y_i(0)] = \mathbf{E}[Y_i(1)|T_i = 1] - \mathbf{E}[Y_i(0)|T_i = 0]$$

RCT vs. RD

- RCT (Experimental design)
  - Treatment
    - $T_i \in \{0, 1\}$, T_i independent of $(Y_i(0), Y_i(1), X_i)$
  - Average treatment effect (ATE)

$$\tau_{\text{ATE}} = \mathbf{E}[Y_i(1) - Y_i(0)] = \mathbf{E}[Y_i(1)|T_i = 1] - \mathbf{E}[Y_i(0)|T_i = 0]$$

- RD (Quasi-experimental design)
  - Treatment
    - $T_i \in \{0, 1\}$, $T_i = \mathbf{1}(X_i \geq \bar{x})$
  - Local average treatment effect at the cutoff (LATE)

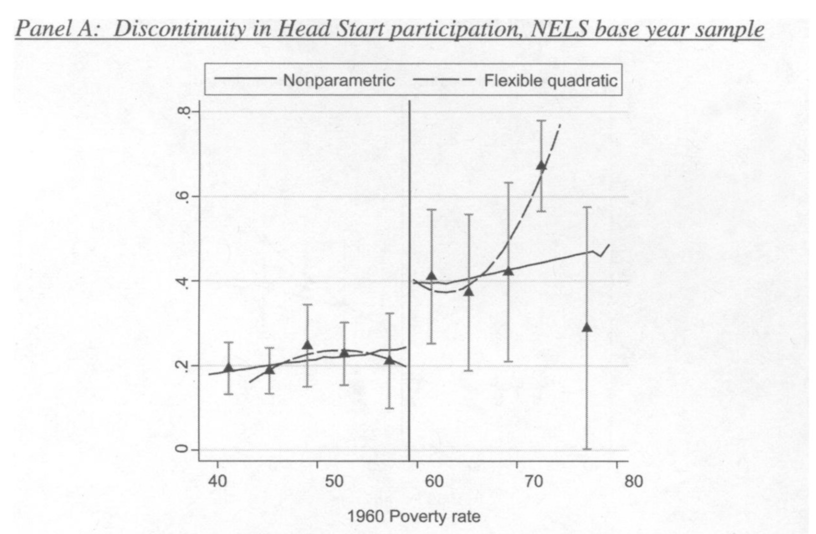
$$\tau_{\text{SRD}} = \mathbf{E}[Y_i(1) - Y_i(0)|X_i = \bar{x}] = \lim_{x \downarrow \bar{x}} \mathbf{E}[Y_i(1)|X_i = x] - \lim_{x \uparrow \bar{x}} \mathbf{E}[Y_i(0)|X_i = x]$$

Example 1: Head Start (Ludwig and Miller, 2007 QJE)

Ludwig and Miller, 2007 QJE

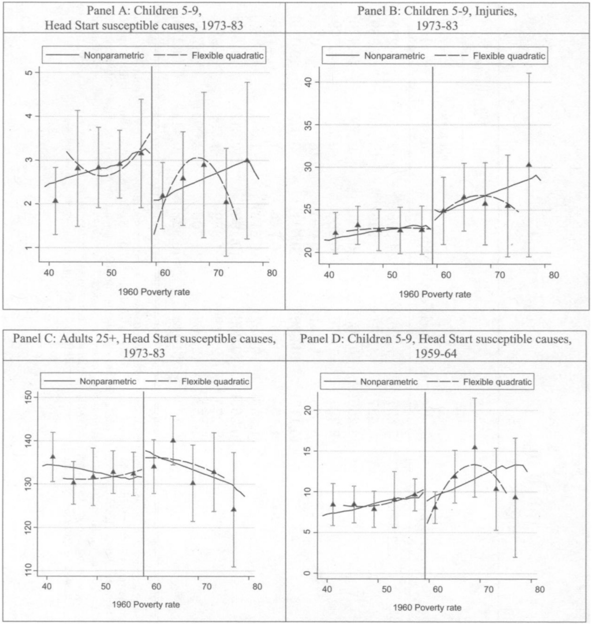
- **Question:** Impact of Head Start on infant mortality
- **Data**
  - Y_i = child mortality 5 to 9 years old
  - T_i = whether county received Head Start assistance
  - X_i = 1960 poverty index ($\bar{x}$ = 59.1984)
- **Potential outcomes**
  - $Y_i(0) =$ child mortality if had not received Head Start
  - $Y_i(1) =$ child mortality if had received Head Start

First stage



Panel A: Discontinuity in Head Start participation, NELS base year sample
Nonparametric
Flexible quadratic
0
.2
.4
.6
.8
40
50
60
70
80
1960 Poverty rate

Second stage



Example 2: School Choice

Exam schools in the U.S.

Econometrica, Vol. 82, No. 1 (January, 2014), 137–196

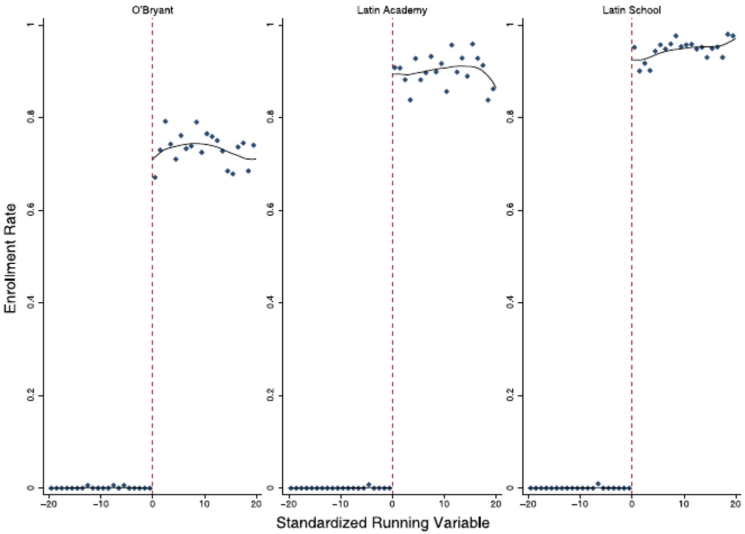
THE ELITE ILLUSION: ACHIEVEMENT EFFECTS AT BOSTON AND NEW YORK EXAM SCHOOLS

BY ATILA ABDULKADIROĞLU, JOSHUA ANGRIST, AND PARAG PATHAK[1]

Parents gauge school quality in part by the level of student achievement and a school's racial and socioeconomic mix. The importance of school characteristics in the housing market can be seen in the jump in house prices at school district boundaries where peer characteristics change. The question of whether schools with more attractive peers are really better in a value-added sense remains open, however. This paper uses a fuzzy regression-discontinuity design to evaluate the causal effects of peer characteristics. Our design exploits admissions cutoffs at Boston and New York City's heavily over-subscribed exam schools. Successful applicants near admissions cutoffs for the least selective of these schools move from schools with scores near the bottom of the state SAT score distribution to schools with scores near the median. Successful applicants near admissions cutoffs for the most selective of these schools move from above-average schools to schools with students whose scores fall in the extreme upper tail. Exam school students can also expect to study with fewer nonwhite classmates than unsuccessful applicants. Our estimates suggest that the marked changes in peer characteristics at exam school admissions cutoffs have little causal effect on test scores or college quality.

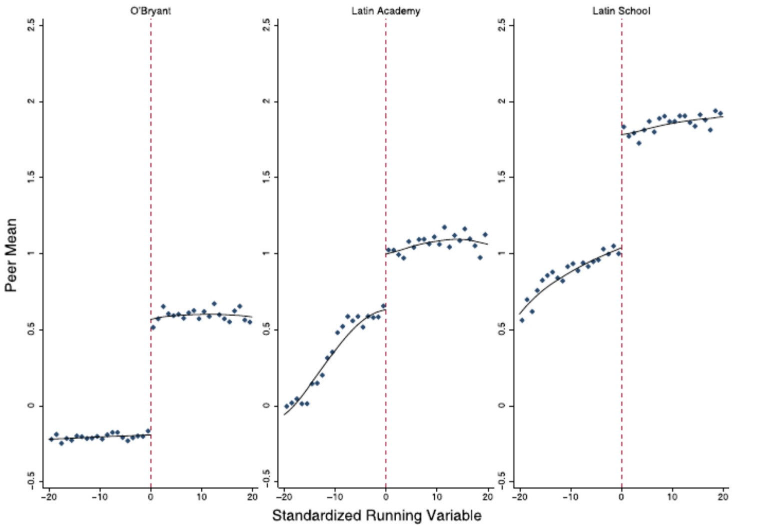
KEYWORDS: Peer effects, school choice, deferred acceptance, selective education.

First stage



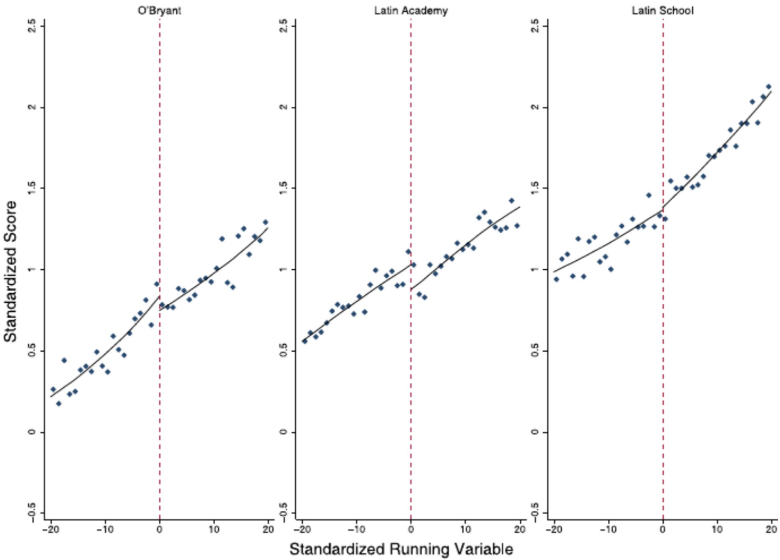
(b) Enrollment at each Boston exam school

First stage



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

Second stage



(a) 7th and 8th grade math at Boston exam schools for 7th grade applicants

High school choice in China

Economics of Education Review 47 (2015) 128–142

Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

The achievement and course-taking effects of magnet schools: Regression-discontinuity evidence from urban China

Thomas Dee[a], Xiaohuan Lan[b,*]

[a] Center for Education Policy Analysis, Stanford University, 520 Galvez Mall, CERAS Building, 5th Floor, Stanford, CA 94305-3001, United States
[b] China Center for Economic Studies, Research Institute of Chinese Economy, Fudan University, 600 Guoquan Road, Shanghai 200433, China

A R T I C L E I N F O

Article history:
Received 28 November 2013
Revised 11 March 2015
Accepted 19 May 2015
Available online 3 June 2015

JEL:
I20
I24

Keywords:
Magnet school
Regression discontinuity design
China

A B S T R A C T

We examine the effects of attending elite magnet schools on the subsequent academic performance of high-school students in urban China. Using a novel data set of the students who entered high school from 2006 to 2008 in a Chinese city, our fuzzy regression discontinuity estimates exploit the threshold values of the high school entrance exam scores. Passing the thresholds significantly reduces the financial cost and raises the probability of attending a magnet school. However, attending such an elite school does not meaningfully improve the academic performance of the marginal student.

First stage

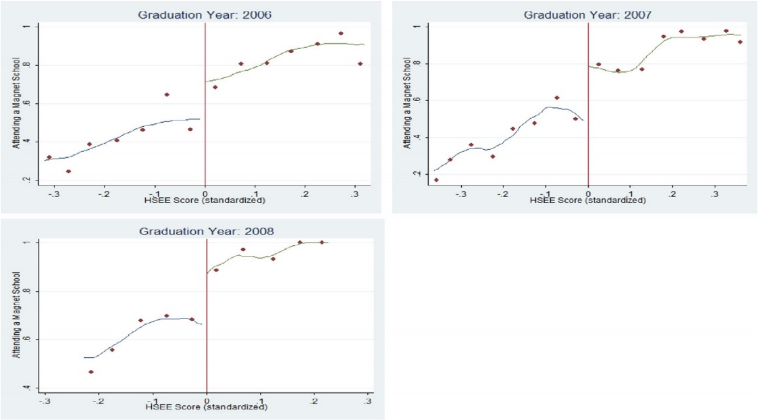


Fig. 2. Probability of attending a magnet school and HSEE scores.

Balance tests

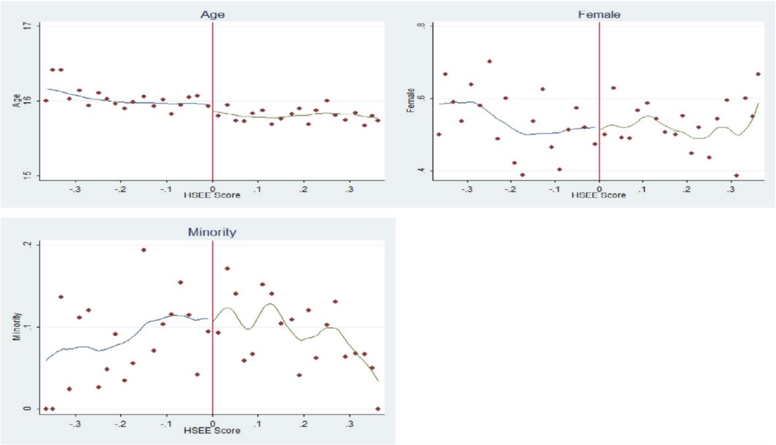


Fig. 3. Students' characteristics and HSEE scores.

Second stage

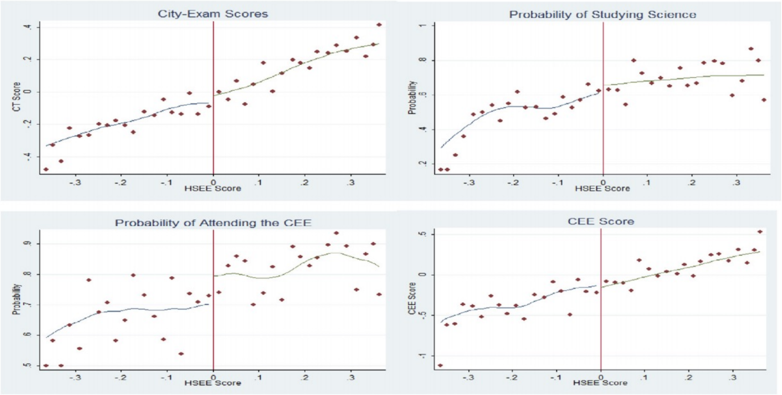


Fig. 4. High school outcomes and HSEE scores.

Results are different in rural China

Journal of Comparative Economics 43 (2015) 825–843

Contents lists available at ScienceDirect

Journal of Comparative Economics

journal homepage: www.elsevier.com/locate/jce

ELSEVIER

Magnet high schools and academic performance in China: A regression discontinuity design☆

Albert Park [a], Xinzheng Shi [b,*], Chang-tai Hsieh [c], Xuehui An [d]

[a] HKUST, Hong Kong
[b] Tsinghua University, China
[c] Chicago Graduate School of Business, United States
[d] National Center for Education Development Research, China Ministry of Education, China

A R T I C L E I N F O

Article history:
Received 9 August 2014
Revised 24 May 2015
Available online 10 November 2015

JEL Classification:
I21
I28
O53

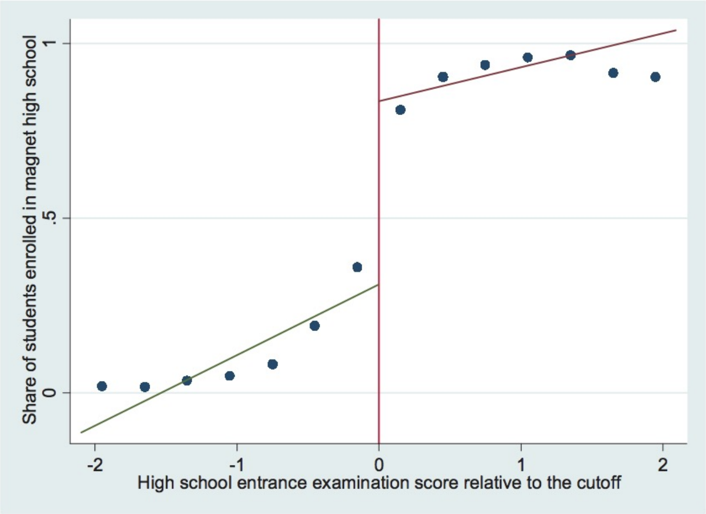
Keywords:
Magnet high school
Regression discontinuity design

A B S T R A C T

Park, Albert, Shi, Xinzheng, Hsieh, Chang-tai, and An, Xuehui—Magnet high schools and academic performance in China: A regression discontinuity design

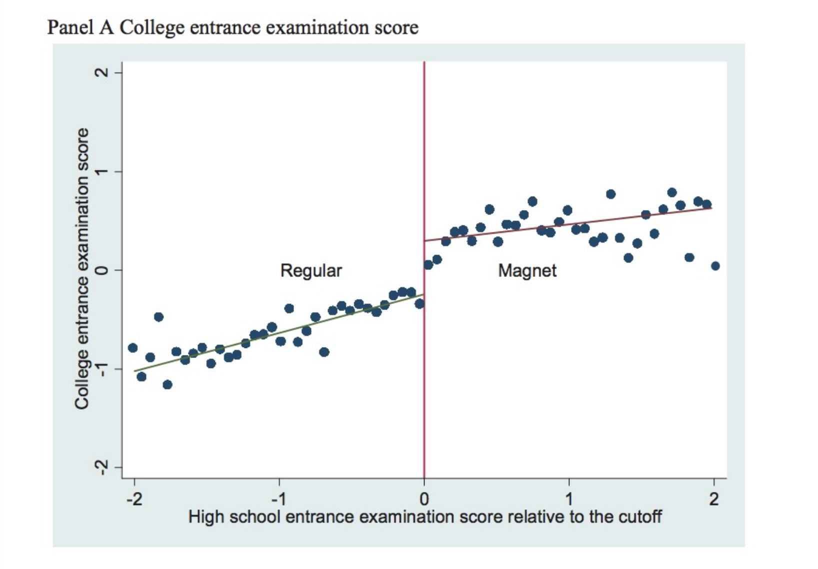
This paper investigates the impact of high school quality on students' educational attainment using a regression discontinuity research design based on entrance examination score thresholds that strictly determine admission to the magnet high schools. Using data from rural counties in Western China, we find that attending a magnet high school significantly increases students' college entrance examination scores and the probability of being admitted to college. *Journal of Comparative Economics* **43** (4) (2015) 825–843. HKUST, Hong Kong; Tsinghua University, China; Chicago Graduate School of Business, United States; National Center for Education Development Research, China Ministry of Education, China.

First stage



Share of students enrolled in magnet high school
0
.5
1
-2
-1
0
1
2
High school entrance examination score relative to the cutoff

Second stage



Potential reason 1: Ability tracking within school

Magnet Classes and Educational Performance: Evidence from China

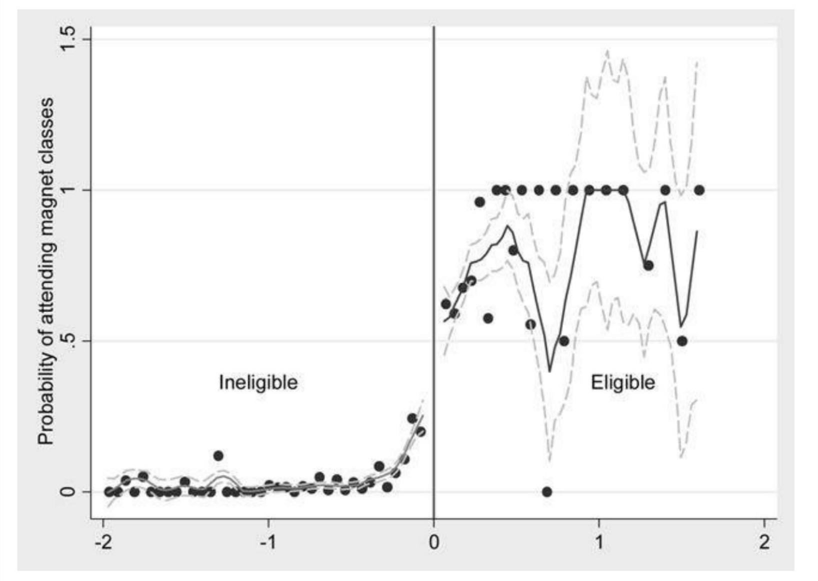
MINGMING MA
University of Southern California

XINZHENG SHI
Tsinghua University

I. Introduction

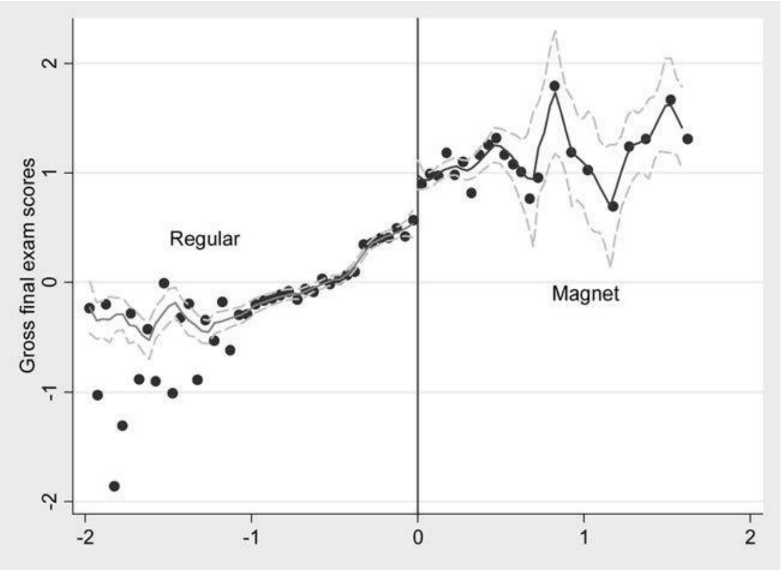
Ability tracking, a concept that originated in developed countries (Figlio and Page 2002), has been increasingly observed in developing countries, including China.[1] In China, middle school graduates are required to take an entrance examination for admission to high school. Although the practice is illegal, most, if not all, high schools in China track students on the basis of academic ability judged primarily by their test scores (usually their high school entrance examination scores) and then place them accordingly into magnet and regular classes.[2]

First stage



Probability of attending magnet classes
1.5
1
.5
0
Ineligible
Eligible
-2
-1
0
1
2

Second stage



Potential reason 2: School difference and selection

第 13 卷第 4 期
2015 年 10 月

北京大学教育评论
Peking University Education Review

Vol. 13, No. 4
October 2015

重点高中能否提高学生的学业成绩

——基于 F 县普通高中的一个断点回归设计研究

王　骏　孙志军

(北京师范大学　经济与工商管理学院/首都教育经济研究院,北京　100875)

摘　要　本文使用 F 县两届普通高中学生全样本数据,利用普通高中招生录取方式对学生能否进入重点高中产生的外生影响,根据断点回归设计的原理,研究了重点高中对学生学业成绩的影响。估计结果表明,就理科生来看,重点高中学生的高考总成绩、数学和语文成绩都高于一般高中,但从数值上看,这一差异并不大;就文科生来看,重点高中与一般高中无显著差别。这说明,重点高中对学生学业成绩仅产生了微弱的正面影响。但由于学习能力和学习基础较好的学生更愿意选择理科,重点高中对不同学科类别学生学业成绩的影响反映的可能是重点高中对不同学习能力和学业基础学生学业成绩的影响。不同学科对学校资源的依赖性不同、学校内部在不同学科资源配置的偏好不同也可能导致这种影响的学科差异。此外,重点高中对女生学业成绩的影响更大,对城市学生高考成绩的影响明显大于农村学生,但对农村学生数学和语

Two cutoffs

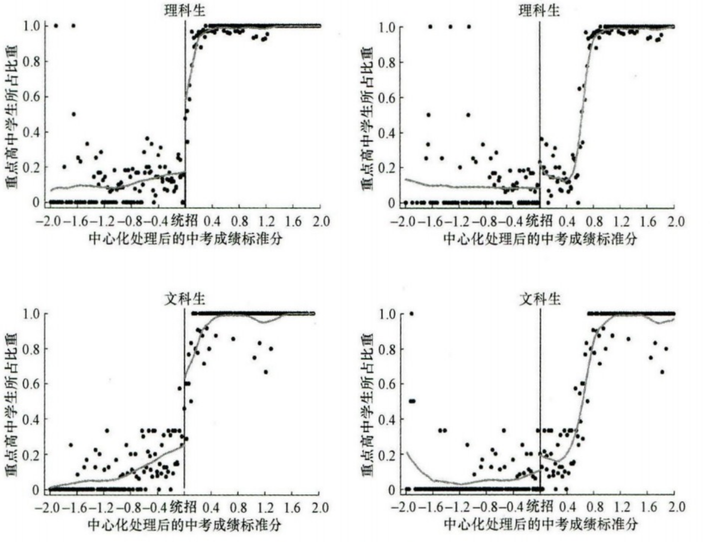


图 1　断点示意图

STEM vs. non-STEM tracks

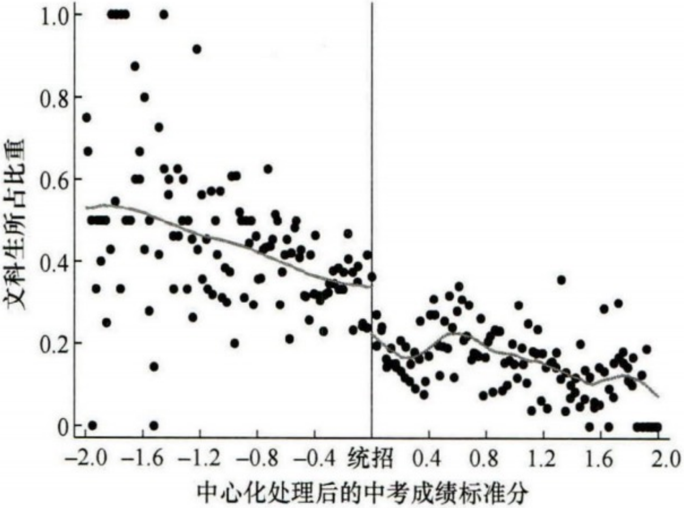


图 5　中考成绩与学生学科类别

Example 3: College Entrance Exam

Cutoff in College Entrance Exam

Just Above the Cutoff: The Return to Elite Education in China *

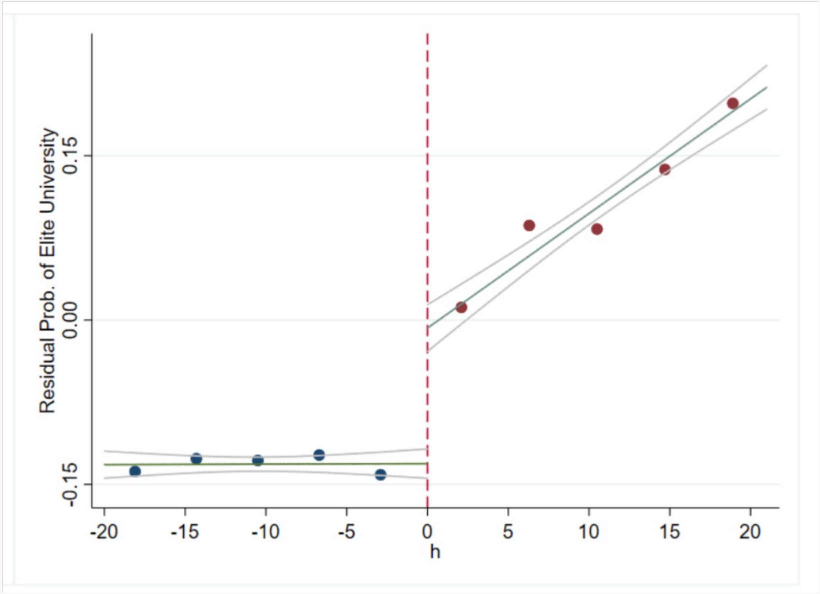
Ruixue Jia† and Hongbin Li‡

January 16, 2020

Abstract

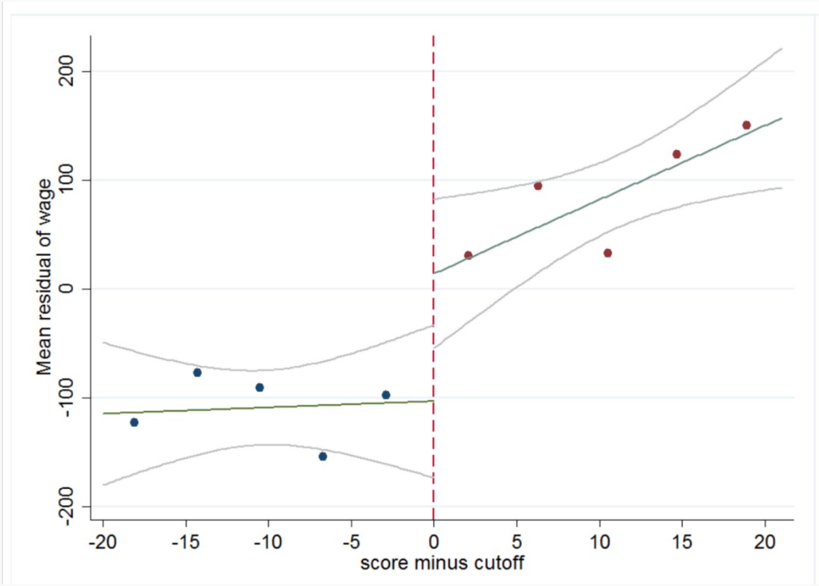
China's National College Entrance Exam (*Gaokao*) is considered the most important life-changing opportunity for students and their families. We estimate the return to entrance to an elite college by utilizing *Gaokao* cutoff scores designated by the government. Employing multiple waves of college student surveys we conducted during 2010-15, we find gaining access to an elite college thanks to a few extra exam points raises a young person's first-job wages by 30-45%. We also find that those just above the cutoff have peers with higher scores and better social networks. Underlying our findings is a hierarchical college system under the control of the government.

First stage



Residual Prob. of Elite University
0.15
0.00
-0.15
-20
-15
-10
-5
0
5
10
15
20
h

Second stage



The RD Family

Key empirical points

- RD designs exploit "variation" near the cutoff
- Graphical analysis is very useful: validation & falsification
- Need to work with data near cutoff
    - bandwidth or window selection
- Covariates and density of running variable should be similar near cutoff
- Zero "overlap" so extrapolation is unavoidable (local or global).
- Causal effect is different (in general) than RCT

Estimands and Identification

- Parameters of interests:
  - Sharp RD (SRD) and Fuzzy RD (FRD)
  - Kink RD (KRD) and Kink Fuzzy RD (KFRD)
  - Multiple scores RD and Geographic RD
  - Pooled RD v.s. Multiple Cutoff RD
- Inference methods roughly the same
- Falsification methods more different in each case

Sharp/Fuzzy RD

- Sharp RD - perfect compliance
  - every unit with score above $\bar{x}$ receives treatment
  - every unit with score below $\bar{x}$ receives control
- Fuzzy RD - imperfect compliance
  - probability of receiving treatment changes at $\bar{x}$, but not necessarily from 0 to 1
  - some units with score above $\bar{x}$ may decide not to take up treatment

Sharp/Fuzzy Kink RD

- A treatment or policy is assigned on the basis of a score via a formula that relates the assignment variable to the treatment
- The formula has a “kink” point ($\bar{x}$) at which it changes discontinuously
- We expect a change in slope at $\bar{x}$, instead of a change in intercept

Formula-based UI benefits

Econometrica, Vol. 83, No. 6 (November, 2015), 2453–2483

NOTES AND COMMENTS

INFERENCE ON CAUSAL EFFECTS IN A GENERALIZED REGRESSION KINK DESIGN

BY DAVID CARD, DAVID S. LEE, ZHUAN PEI, AND ANDREA WEBER[1]

We consider nonparametric identification and estimation in a nonseparable model where a continuous regressor of interest is a known, deterministic, but kinked function of an observed assignment variable. We characterize a broad class of models in which a sharp “Regression Kink Design” (RKD or RK Design) identifies a readily interpretable treatment-on-the-treated parameter (Florens, Heckman, Meghir, and Vytlacil (2008)). We also introduce a “fuzzy regression kink design” generalization that allows for omitted variables in the assignment rule, noncompliance, and certain types of measurement errors in the observed values of the assignment variable and the policy variable. Our identifying assumptions give rise to testable restrictions on the distributions of the assignment variable and predetermined covariates around the kink point, similar to the restrictions delivered by Lee (2008) for the regression discontinuity design. Using a kink in the unemployment benefit formula, we apply a fuzzy RKD to empirically estimate the effect of benefit rates on unemployment durations in Austria.

First stage

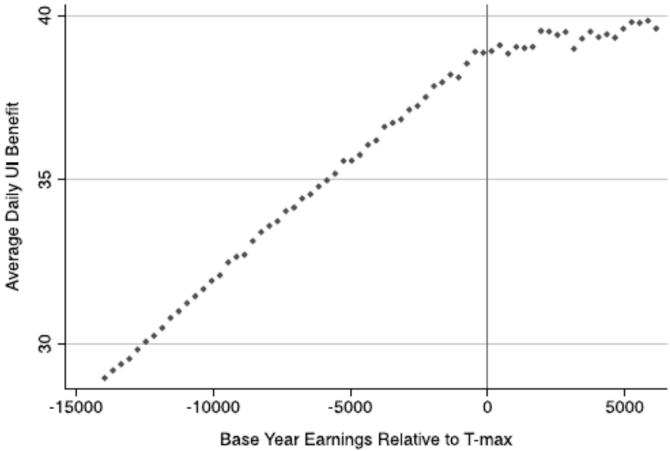


FIGURE 2.—Daily UI benefits.

Second stage

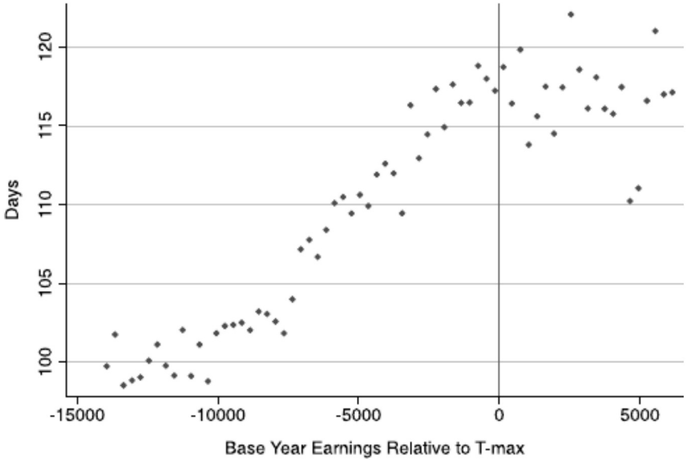


FIGURE 3.—Unemployment duration.

Multiple scores RD

The Stata Journal (*xxxx*) *vv*, Number *ii*, pp. 1–24

Analysis of Regression Discontinuity Designs with Multiple Cutoffs or Multiple Scores

Matias D. Cattaneo
Princeton University
Princeton, NJ
cattaneo@princeton.edu

Rocio Titiunik
Princeton University
Princeton, NJ
titiunik@princeton.edu

Gonzalo Vazquez-Bare
UC Santa Barbara
Santa Barbara, CA
gvazquez@econ.ucsb.edu

Abstract. We introduce the `Stata` (and `R`) package `rdmulti`, which includes three commands (`rdmc`, `rdmcplot`, `rdms`) for analyzing Regression Discontinuity (RD) designs with multiple cutoffs or multiple scores. The command `rdmc` applies to non-cumulative and cumulative multi-cutoff RD settings. It calculates pooled and cutoff-specific RD treatment effects, and provides robust bias-corrected inference procedures. Post estimation and inference is allowed. The command `rdmcplot` offers RD plots for multi-cutoff settings. Finally, the command `rdms` concerns multi-score settings, covering in particular cumulative cutoffs and two running variables contexts. It also calculates pooled and cutoff-specific RD treatment effects, provides robust bias-corrected inference procedures, and allows for post-estimation estimation and inference. These commands employ the `Stata` (and `R`) package `rdrobust` for plotting, estimation, and inference. Companion `R` functions with the same syntax and capabilities are provided.

Keywords: st0001, regression discontinuity designs, multiple cutoffs, multiple scores, local polynomial methods.

This version: April 23, 2020

Multiple scores RD

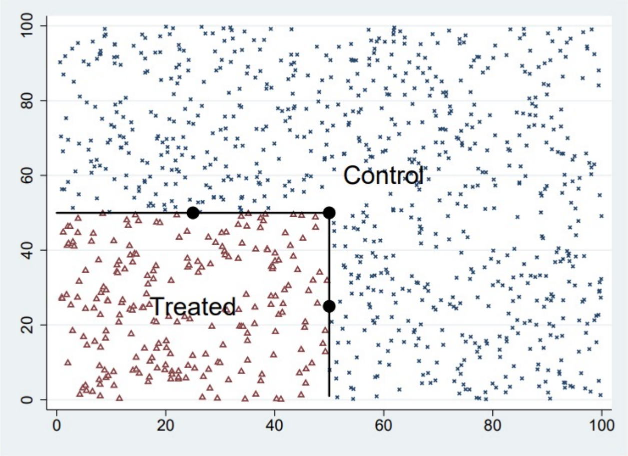


Figure 4: Bivariate score.

Geographic RD

Econometrica, Vol. 78, No. 6 (November, 2010), 1863–1903

THE PERSISTENT EFFECTS OF PERU'S MINING *MITA*

BY MELISSA DELL[1]

This study utilizes regression discontinuity to examine the long-run impacts of the *mita*, an extensive forced mining labor system in effect in Peru and Bolivia between 1573 and 1812. Results indicate that a *mita* effect lowers household consumption by around 25% and increases the prevalence of stunted growth in children by around 6 percentage points in subjected districts today. Using data from the Spanish Empire and Peruvian Republic to trace channels of institutional persistence, I show that the *mita*'s influence has persisted through its impacts on land tenure and public goods provision. *Mita* districts historically had fewer large landowners and lower educational attainment. Today, they are less integrated into road networks and their residents are substantially more likely to be subsistence farmers.

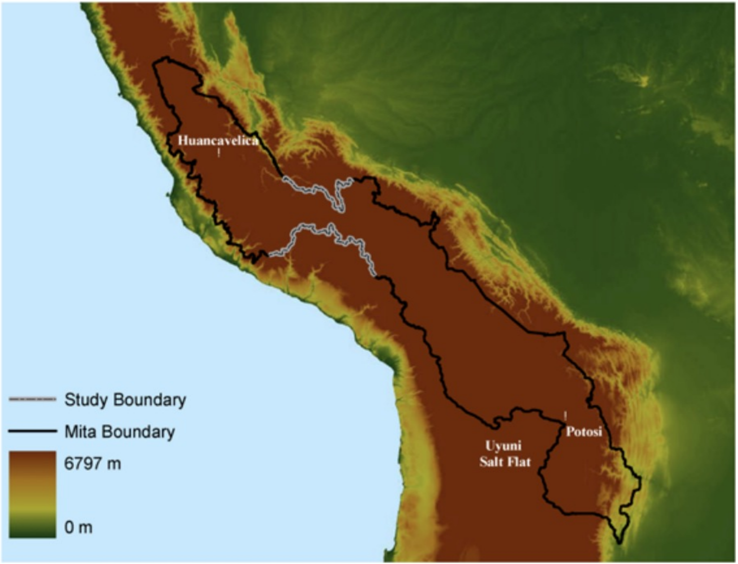
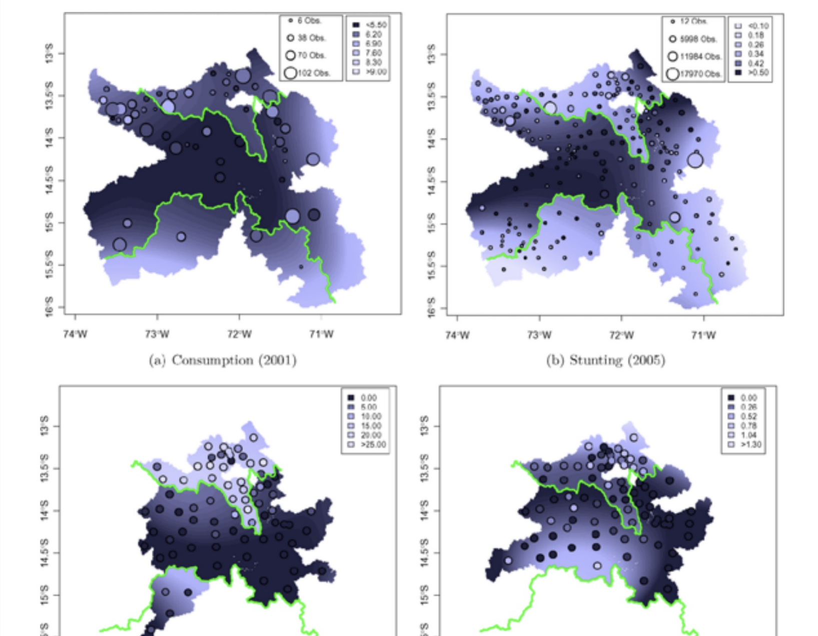


FIGURE 1.—The *mita* boundary is in black and the study boundary in light gray. Districts falling

Geographic RD



(a) Consumption (2001)

(b) Stunting (2005)

Multiple Cutoff RD

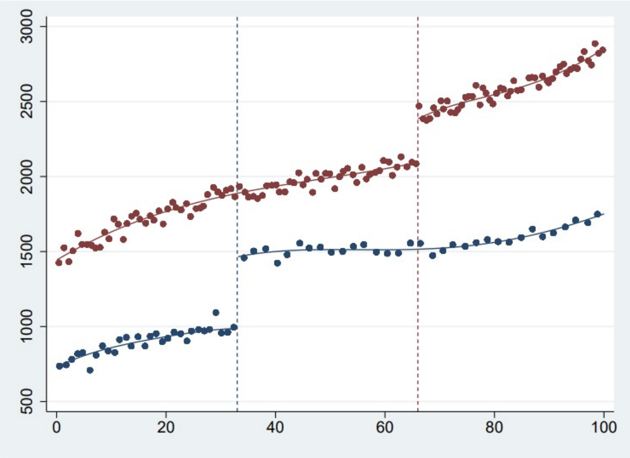


Figure 1: Multiple RD plot.

Graphical and falsification methods

- Always (beautifully) plot data: main advantage of RD designs!
- Plot outcomes
- Plot covariates
- Plot density of X_i (manipulation tests: continuity at cutoff)
- Plot placebo outcomes (0 RD treatment effects)

RD packages

https://sites.google.com/site/rdpackages/

RD Software Packages

- **Home**
- rdrobust
- rdlocrand
- rddensity
- rdmulti
- rdpower
- Replication Files

This work was supported in part by the National Science Foundation through grants SES-1357561 and SES-1459931.

Software available in R and Stata:

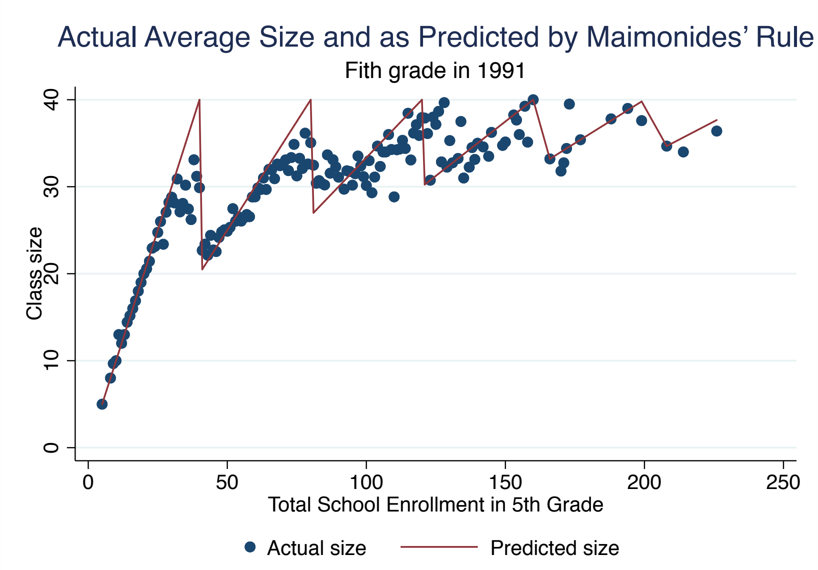
- **`rdrobust`**: inference and graphical procedures using local polynomial and partitioning regression methods.
- **`rdlocrand`**: finite-sample inference using local randomization methods.
- **`rddensity`**: manipulation testing using local polynomial density methods.
- **`rdmulti`**: estimation, inference, RD Plots, and extrapolation with multiple cutoffs and multiple scores.
- **`rdpower`**: power and sample size calculations using robust bias-corrected local polynomial inference.

Contributors:

Sebastian Calonico, Columbia University
Matias D. Cattaneo, Princeton University
Max H. Farrell, University of Chicago
Michael Jansson, UC Berkeley and CREATES
Xinwei Ma, UC San Diego
Rocio Titiunik, Princeton University
Gonzalo Vazquez-Bare, UC Santa Barbara

Practice Example: Class Size

Angrist & Lavy, 1999 QJE



What is the effect of class size on test scores?

1. What is an ideal RCT to answer this question?
2. Whether and how an RD design will help?
3. Plot: 1st stage
4. Plot: 2nd stage
5. Plot: Covariates
6. Plot: Manipulation tests
7. Estimates: Parametric
8. Estimates: Non-parametric

Thanks!

Acknowledgments

I have borrowed from various sources to prepare this lecture, including
Matias Cattaneo's Econ 675,
Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2019). A Practical Introduction to Regression Discontinuity Designs: Foundations. Cambridge University Press,
Angrist and Pischke's MHE, and

https://www.scunning.com/causalinference_norap.pdf.