Principles of Empirical Analysis

Lecture 10: Regression discontinuity design

Spring 2021
Tuukka Saarimaa

Regression discontinuity design (RDD)

Outline

- **Basic idea of regression discontinuity designs**
  - Setup and assumptions
  - Fuzzy and sharp RDD
- **Testing RDD assumptions**
  - Manipulation, covariate balance, fake cutoff placebos and other placebos
- **Applications**
- **Geographic discontinuity**

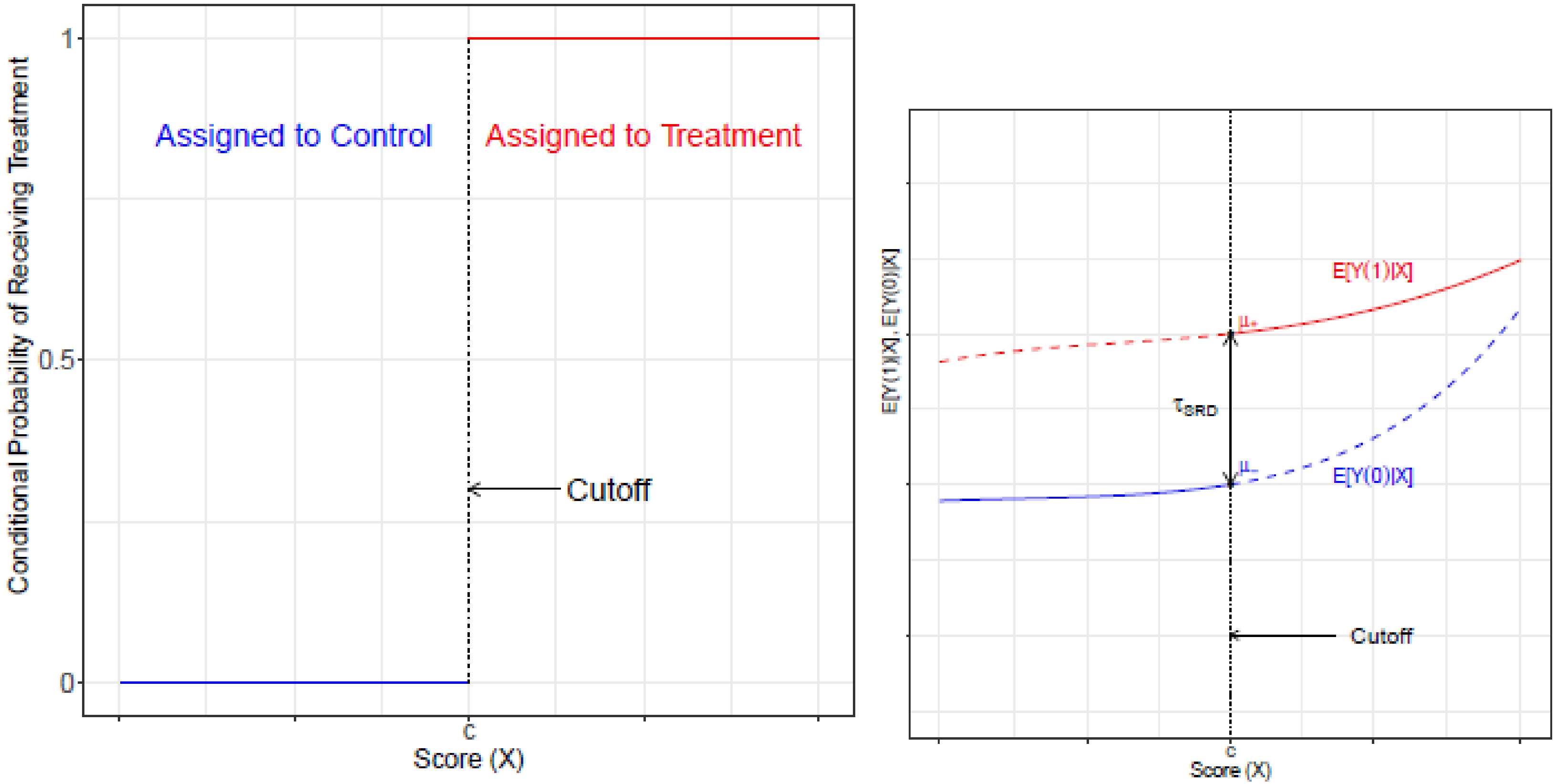
RDD

- **Introduced by Thistlethwaithe and Cambell (1960)**
  - Studied the impact of merit awards on future academic outcomes, where merit award was given if test score exceeds a cutoff
  - Idea: students can of course affect their test scores by studying, but they cannot manipulate their scores to be just above the cutoff because the cutoff is unknown to them *ex ante*
- **Reappeared and formalized in economics in late 90s and has proven to be a powerful causal tool in empirical economics and other disciplines**
  - Political science, education, epidemiology, criminology etc.
- **Strong internal validity, but very data intensive**
  - Need to have a lot of observations near the cutoff

RDD – the setup

- **RDD has three fundamental components: running variable, cutoff, and treatment**
- **Individual receives a treatment after crossing some cutoff in the running (or forcing) variable**
  - Sharp RDD: treatment received with probability zero below the threshold and probability one above threshold
  - Fuzzy RDD: The probability of receiving the treatment increases discontinuously at the threshold (imperfect compliance)
- **Assumption: the potential outcomes evolve smoothly across the cutoff**
  - If there is no precise manipulation of the running variable, observations just below the threshold are a valid control group for those just above the threshold

Sharp RDD

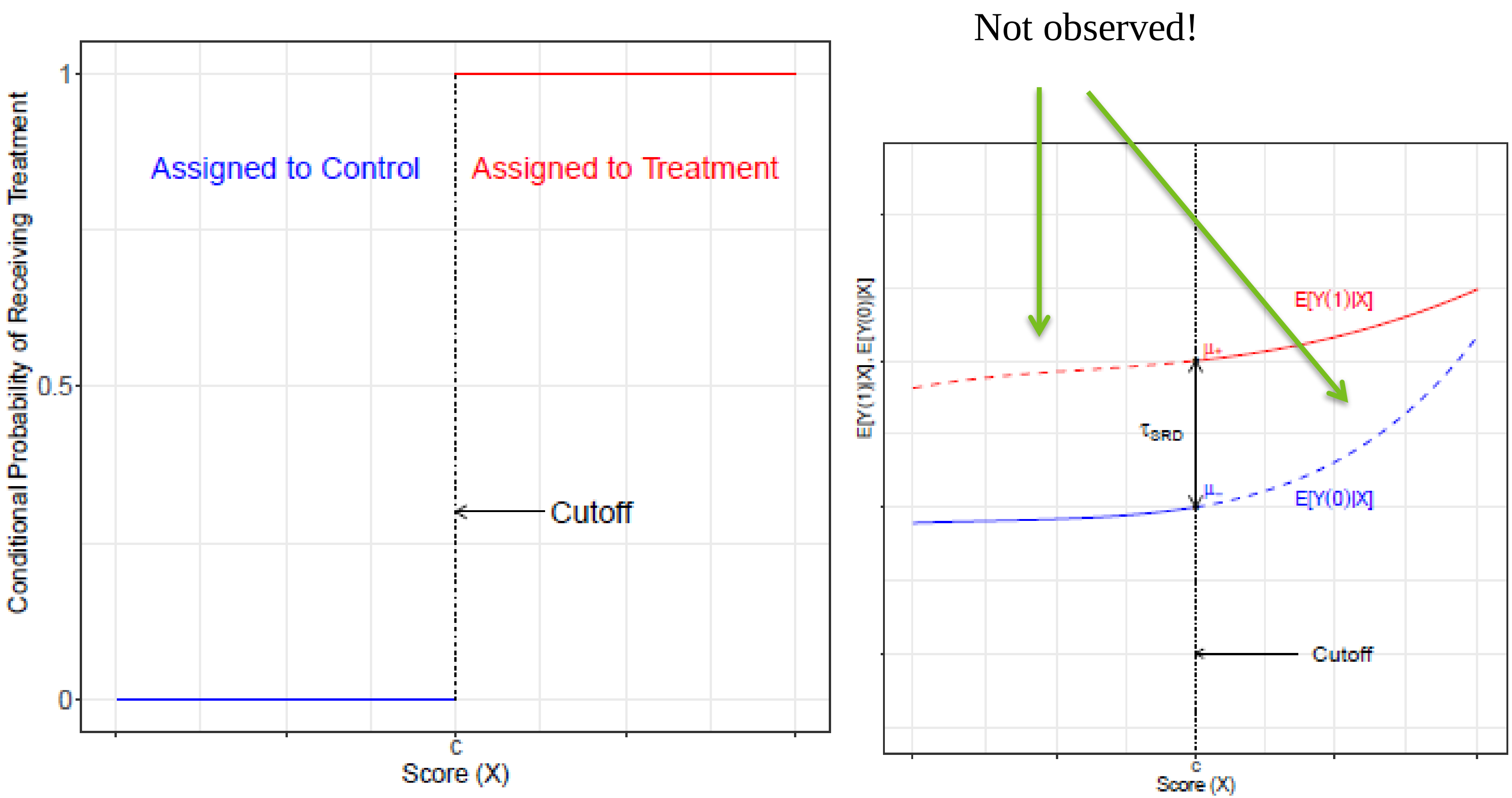


(a) Conditional Probability of Treatment

(b) RD Treatment Effect

Source: Cattaneo et al. (2019): A Practical Introduction to Regression Discontinuity Designs: Foundations.

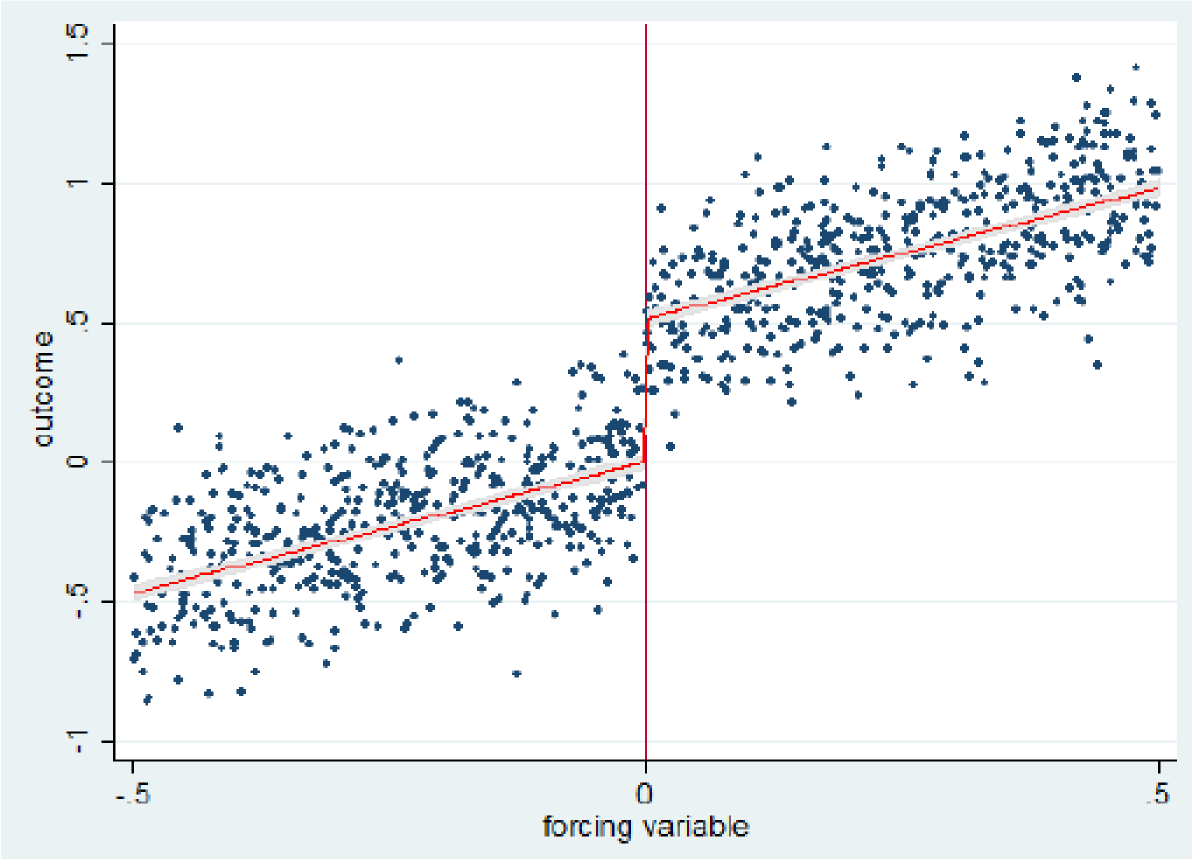
Sharp RDD



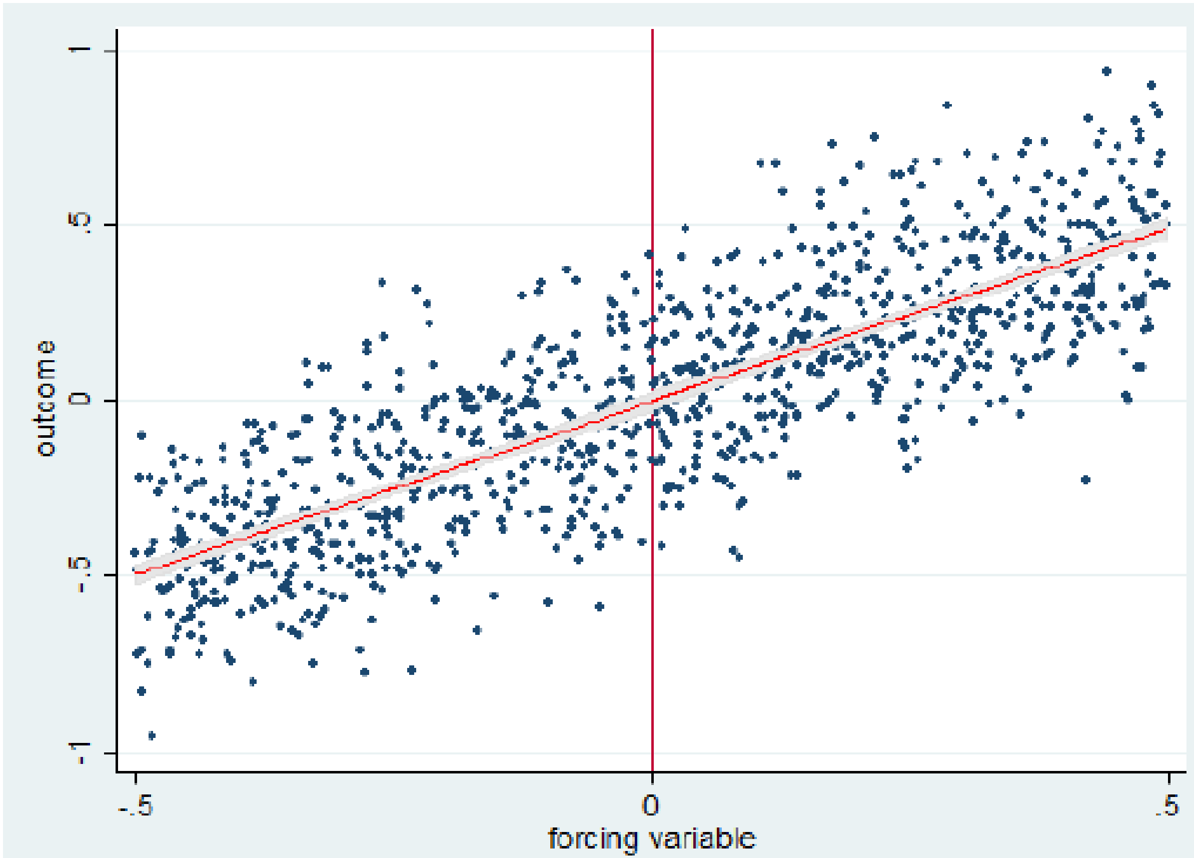
(a) Conditional Probability of Treatment (b) RD Treatment Effect

Source: Cattaneo et al. (2019): A Practical Introduction to Regression Discontinuity Designs: Foundations.

Example of a positive treatment effect



Example of a zero treatment effect



Example: Minimum legal drinking age in the US

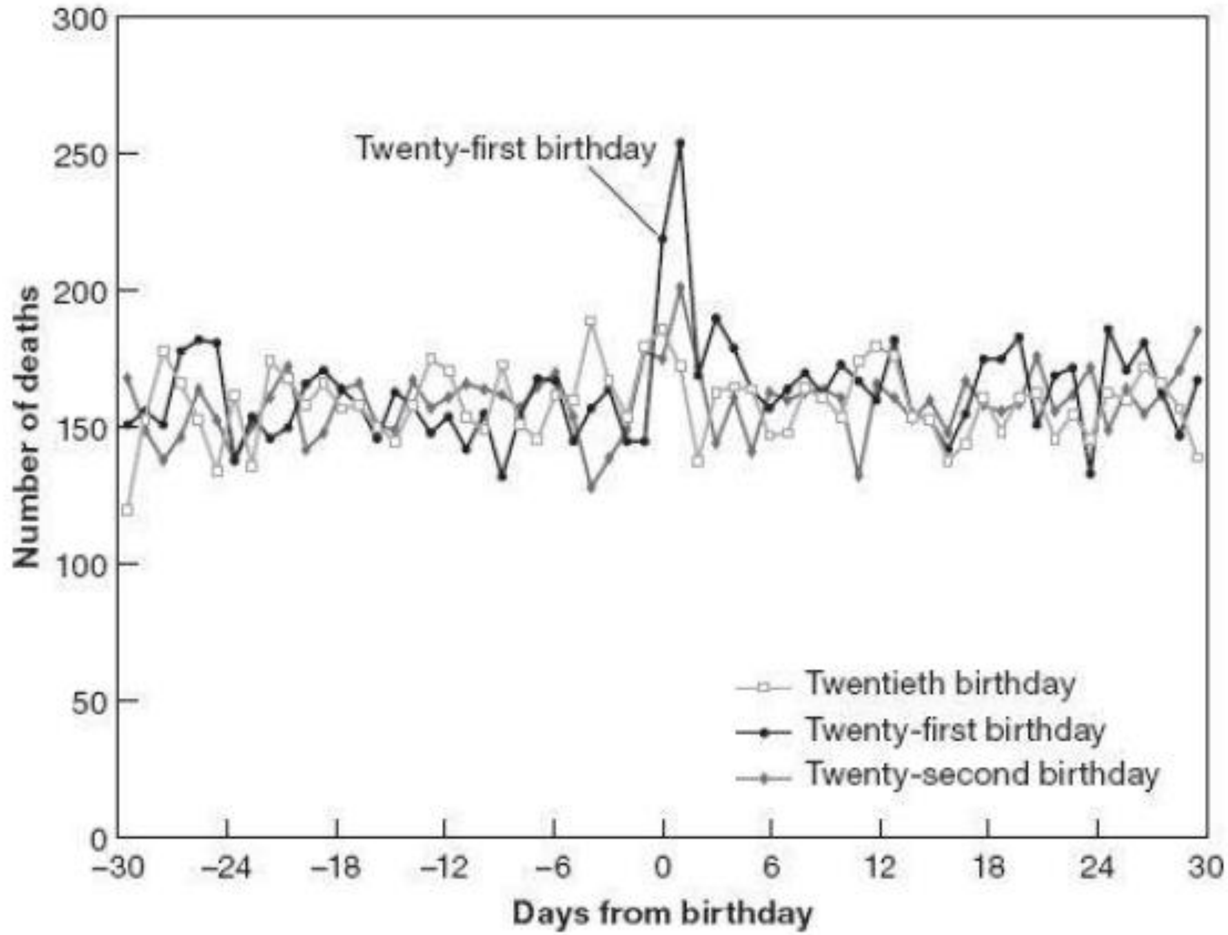
American Economic Journal: Applied Economics 2009, 1:1, 164–182
http://www.aeaweb.org/articles.php?doi=10.1257/app.1.1.164

The Effect of Alcohol Consumption on Mortality: Regression Discontinuity Evidence from the Minimum Drinking Age[†]

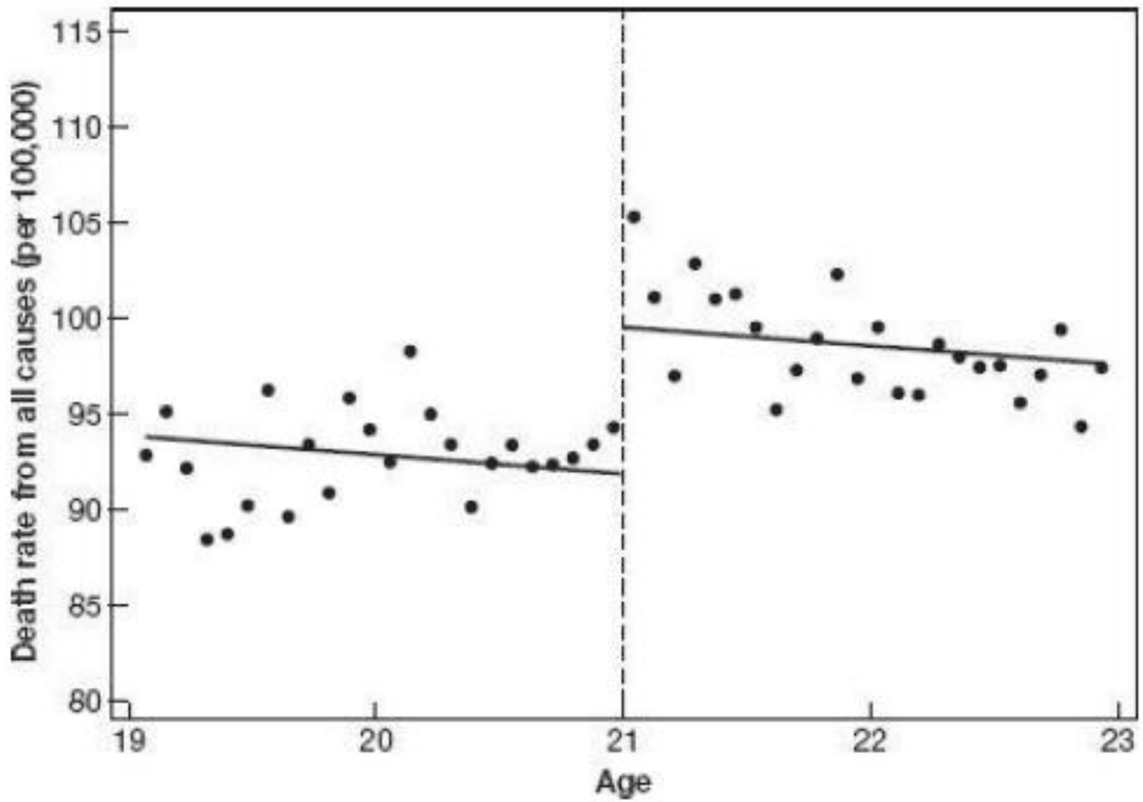
By CHRISTOPHER CARPENTER AND CARLOS DOBKIN[*]

We estimate the effect of alcohol consumption on mortality using the minimum drinking age in a regression discontinuity design. We find large and immediate increases in drinking at age 21, including a 21 percent increase in recent drinking days. We also find a discrete 9 percent increase in the mortality rate at age 21, primarily due to motor vehicle accidents, alcohol-related deaths, and suicides. We estimate a 10 percent increase in the number of drinking days for young adults results in a 4.3 percent increase in mortality. Our results suggest policies that reduce drinking among young adults can have substantial public health benefits. (*JEL* I12, I18)

FIGURE 4.1
Birthdays and funerals

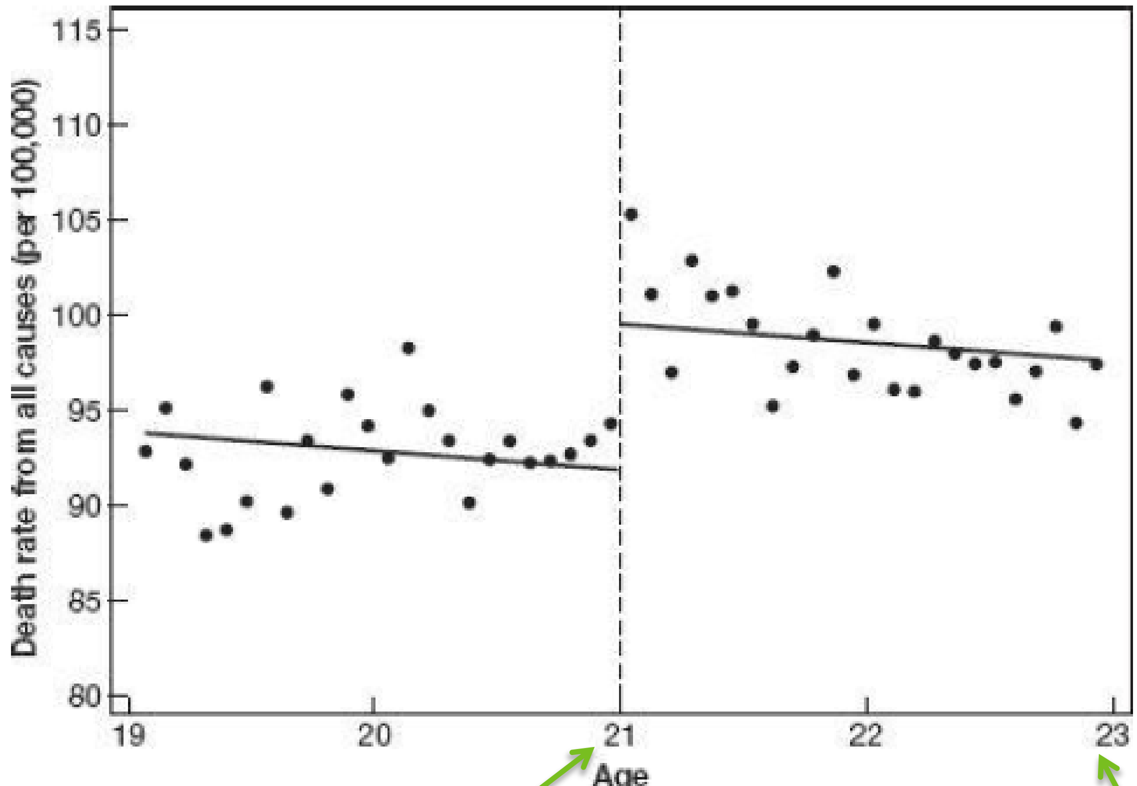


Source: Angrist & Pischke (2015): Mastering Metrics.



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).

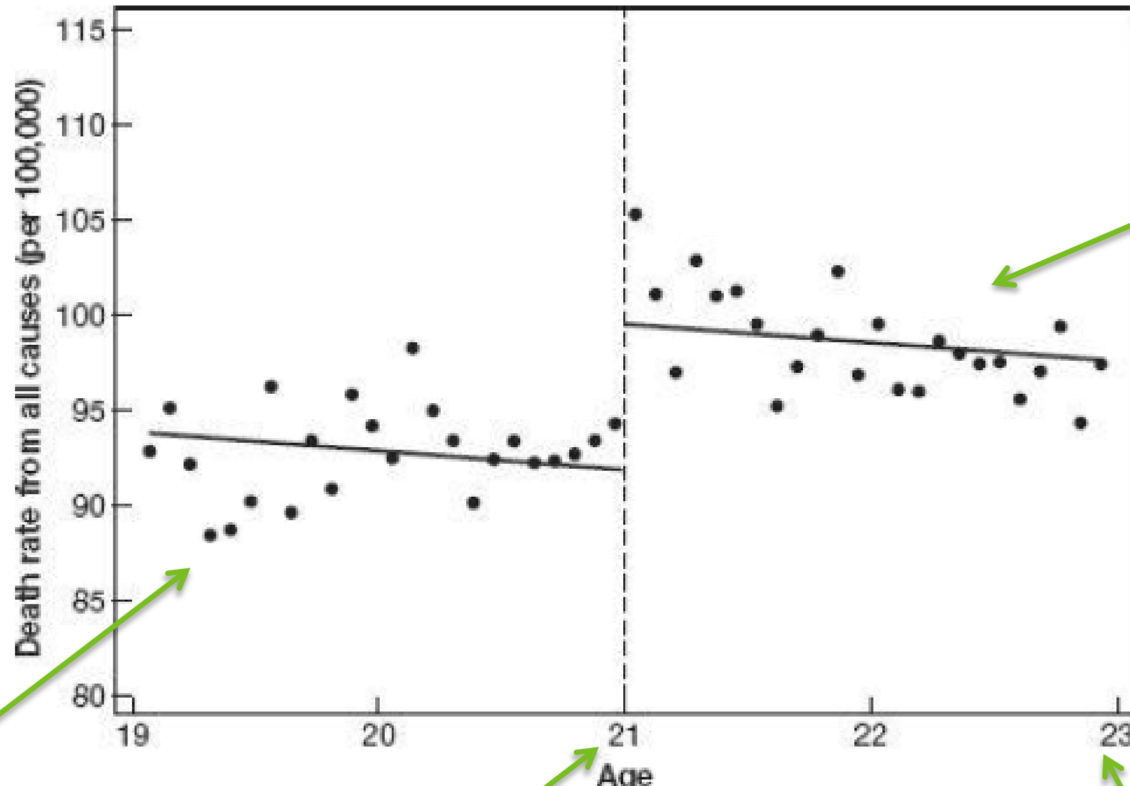
Source for the figure: Angrist & Pischke (2015): Mastering Metrics.



Cutoff (sharp!)

Running variable

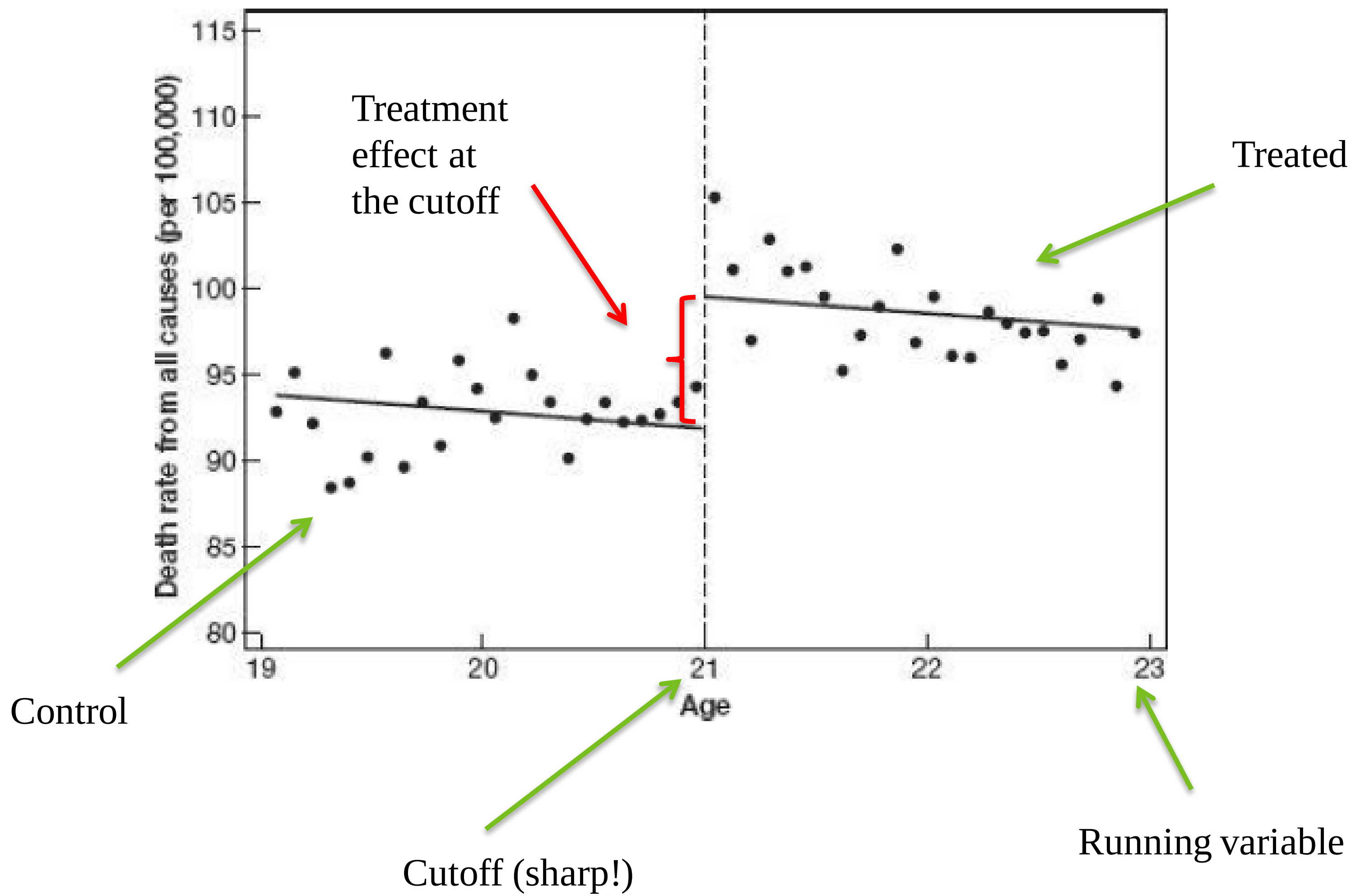
Treated

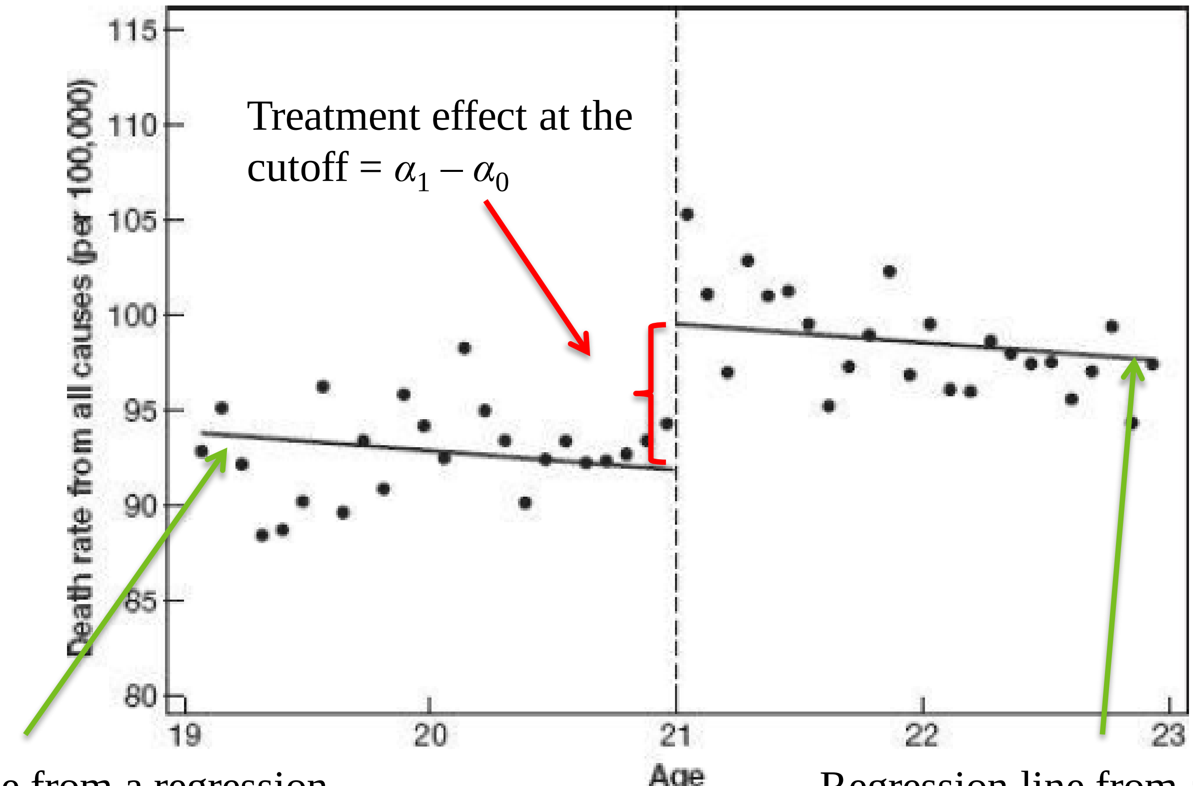


Control

Running variable

Cutoff (sharp!)


Treatment
effect at
the cutoff
Treated
Control
Cutoff (sharp!)
Running variable
Death rate from all causes (per 100,000)
115
110
105
100
95
90
85
80
19
20
21
22
23
Age

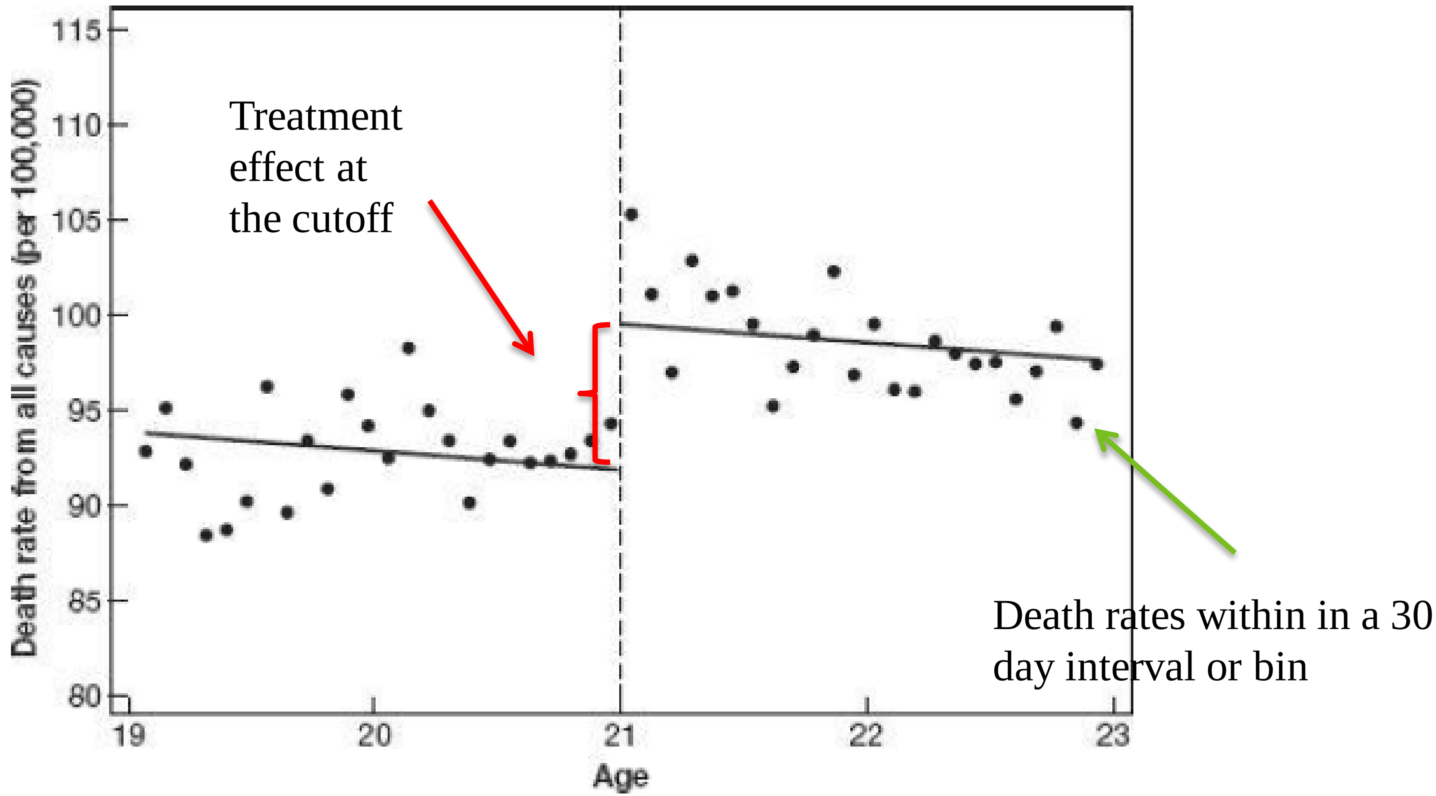


Regression line from a regression of death rates on age:

$E[death\ rate|age,\ age < 21]$
$= \alpha_0 + \beta_0 {*} age,$

Regression line from a regression of death rates on age:

$E[death\ rate|age,\ age \geq 21]$
$= \alpha_1 + \beta_1 {*} age$



- Instead of showing a scatter plot on individual level data, papers often show a scatter plot where the data is binned into smaller number of groups
- The regression lines are fitted separately for both sides of the cutoff using the individual level micro data

More results – alcohol consumption

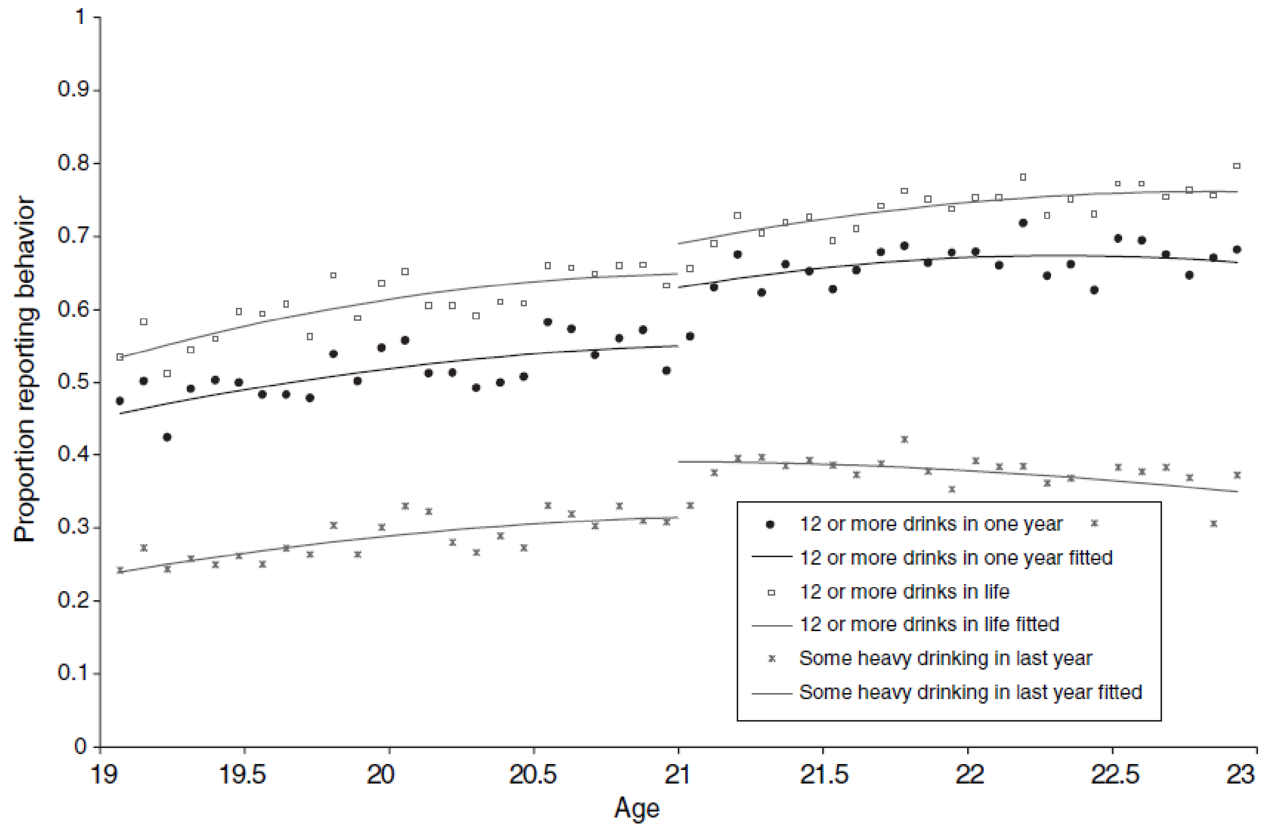


FIGURE 1. AGE PROFILE OF DRINKING PARTICIPATION

Notes: NHIS Sample Adult 1997–2005. Cells are the proportion of people in a 30-day block that report the behavior. The regression line is a second-order polynomial fitted on unweighted individual observations on either side of the age 21 cutoff.

More results – alcohol consumption

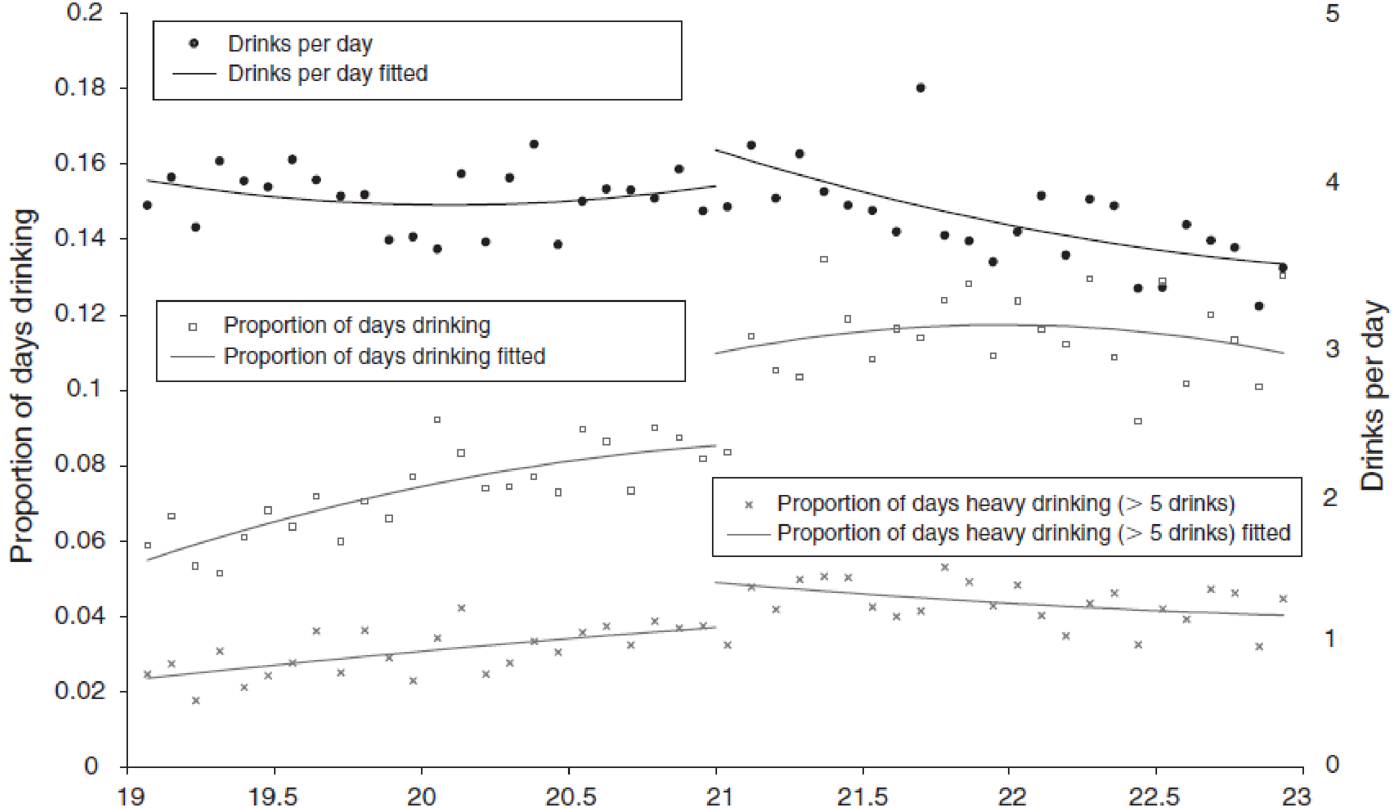


Figure 2. Age Profile of Drinking Intensity

Notes: People can report their drinking for the last week, month, or year; 71 percent of respondents used a reference period of one week or one month. Average number of drinks per day is for people who reported some drinking.

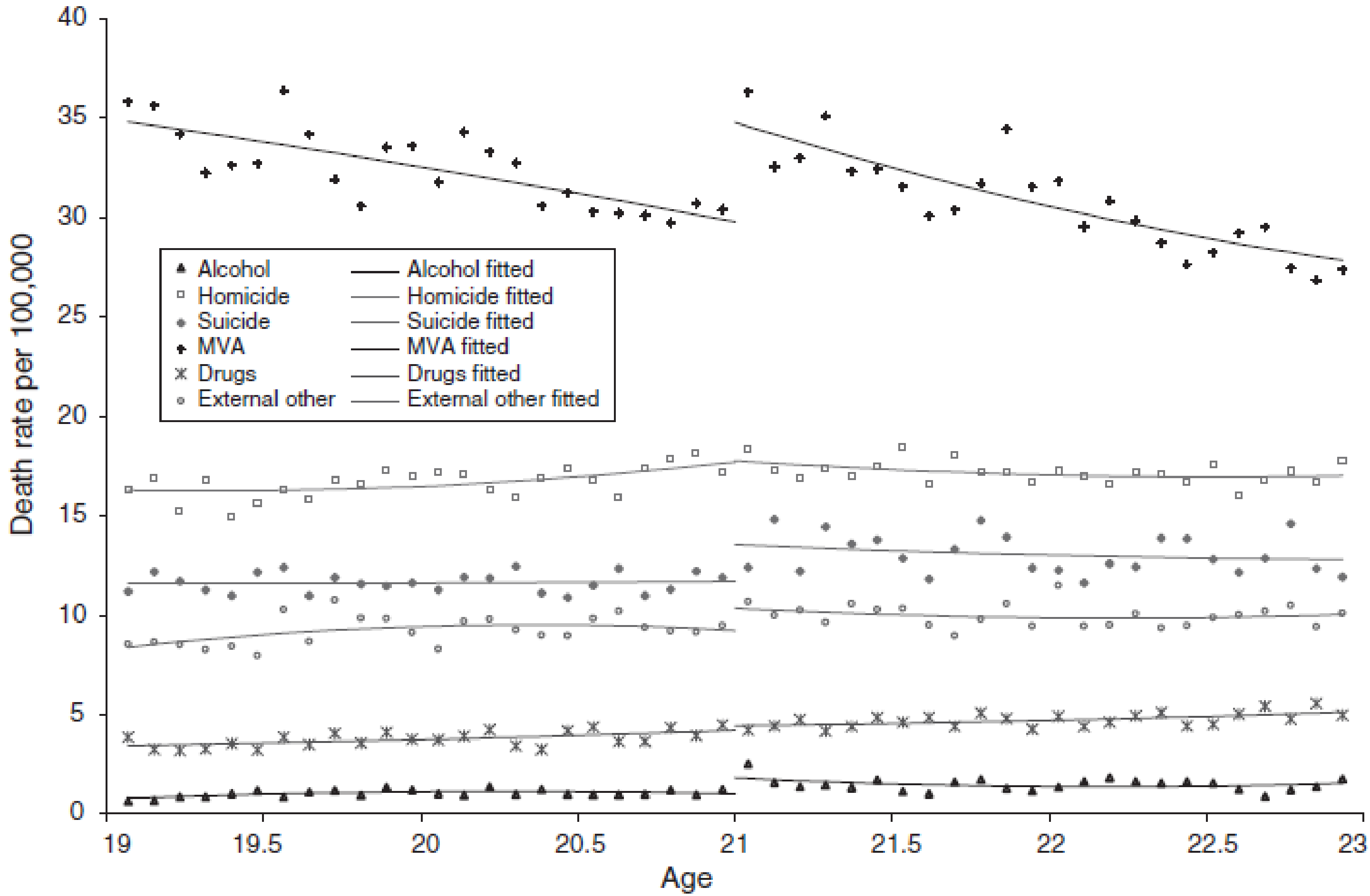


FIGURE 4. AGE PROFILES FOR DEATH RATES BY EXTERNAL CAUSE

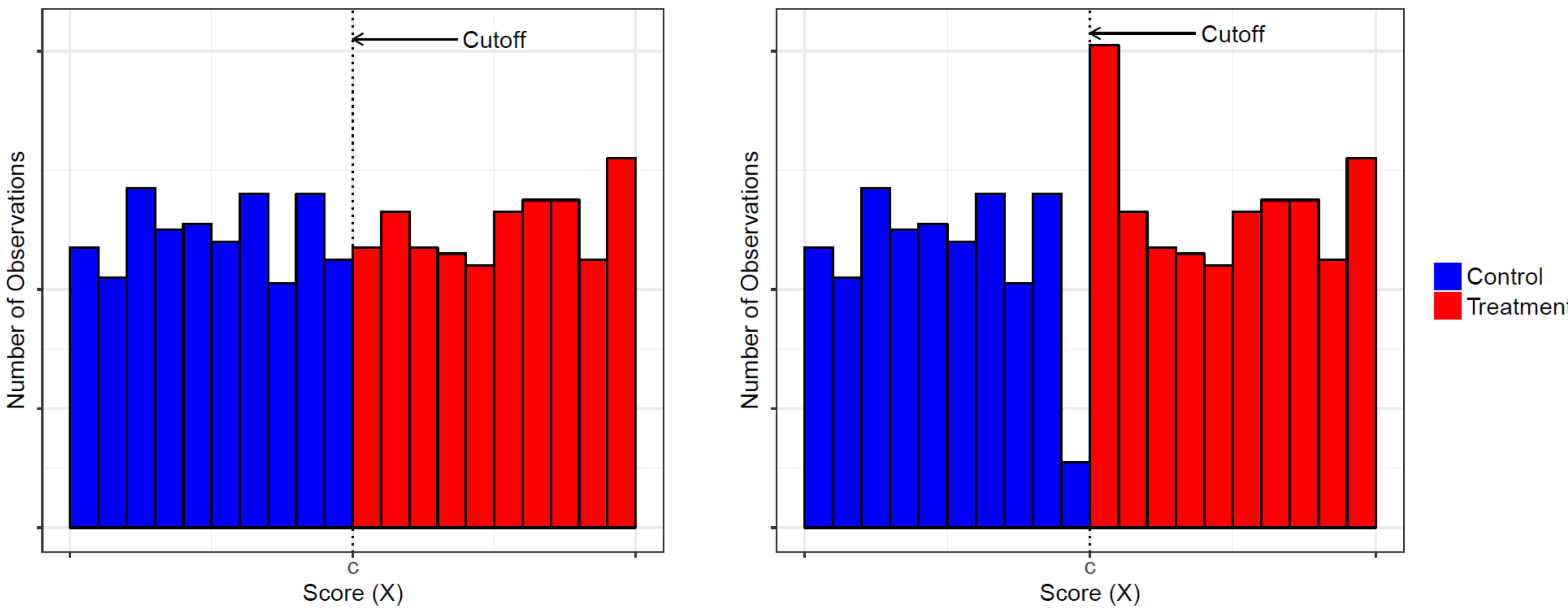
Notes: See notes to Figure 3. The categories are mutally exclusive. The order of precedence is homicide, suicide, MVA, deaths with a mention of alcohol, and deaths with a mention of drugs. The ICD-9 and ICD-10 Codes are in Appendix C.

Testing for RDD assumptions

Sorting or "manipulation" of the running variable

- **The underlying assumption in RDD is that units do not have the ability to precisely manipulate the value of the running variable they receive**
  - If they could and the treatment is something beneficial, units would want to receive the treatment and sort on the right side of the cutoff
- **With no precise manipulation, the number of treated observations just above the cutoff should be approximately the same as the number of control observations below it**
  - Test: plotting the histogram of the running variable and inspecting whether the number of observations are similar near the cutoff
  - Also, a formal statistical density test (McCrary test)

Test for sorting or "manipulation" of the running variable



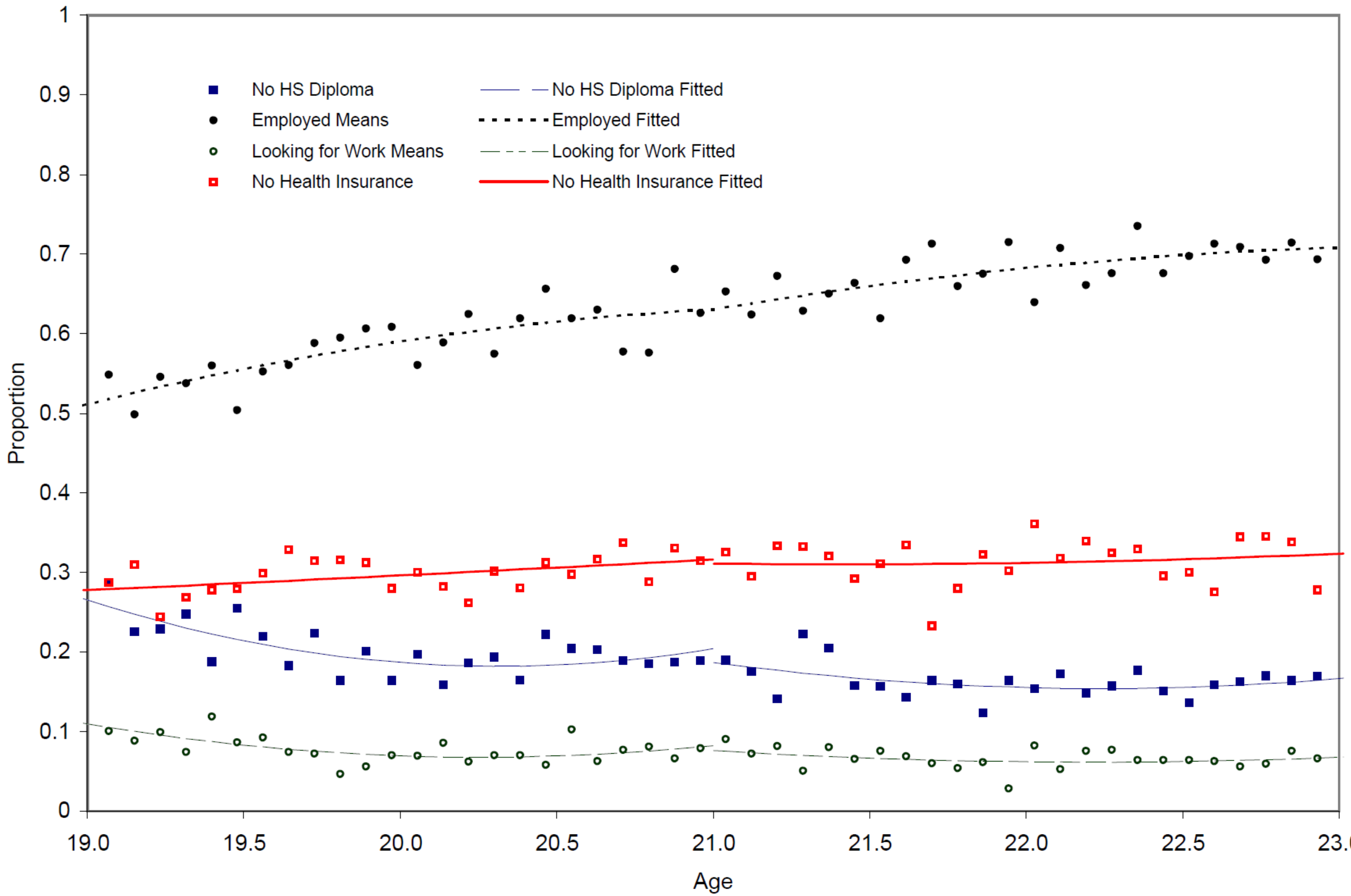
(a) No Sorting (b) Sorting

Source: Cattaneo et al. (2019): A Practical Introduction to Regression Discontinuity Designs: Foundations.

Test of predetermined covariates

- **One of the most important RDD falsification tests involves examining whether, near the cutoff, treated units are similar to control units in terms of observable characteristics**
- **Idea: if units lack the ability to precisely manipulate the running variable, there should be no systematic differences between units with similar values of the running variable**
  - Thus, except for their treatment status, units just above and just below the cutoff should be similar in all variables that could not have been affected by the treatment
- **Implementation: all predetermined covariates should be analyzed using RDD in the same way as the outcome of interest**

Test of predetermined covariates



Placebo tests

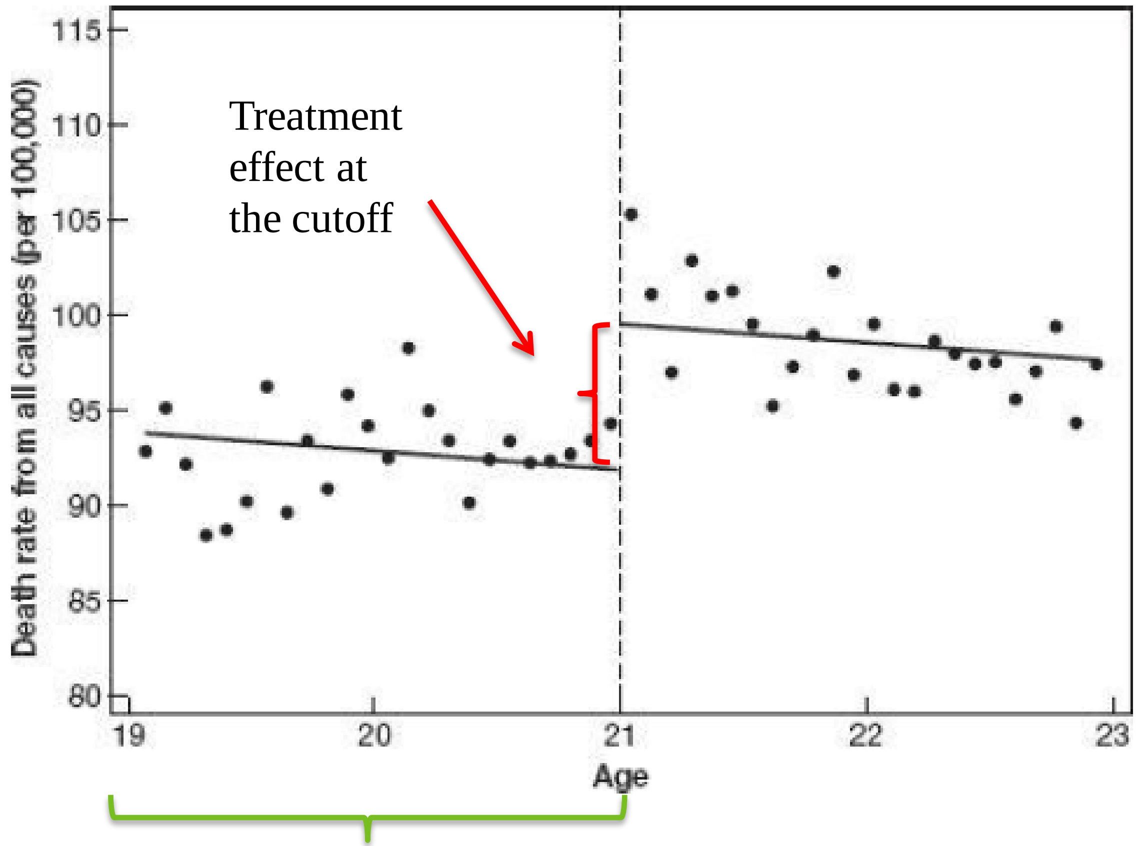
1. **Another important falsification test is to replace the true cutoff value with a fake cutoff value in the running variable**
    - A value at which the treatment status does not really change and perform estimation and inference using this “fake” cutoff
    - A significant treatment effect should occur only at the true cutoff value and not at other values where the treatment status is constant
    - No jumps in death rates at 18, 19 or 25 etc.
2. **Also, can run placebos at the true cutoff on outcomes that should not be affected by the treatment**

Local randomization interpretation

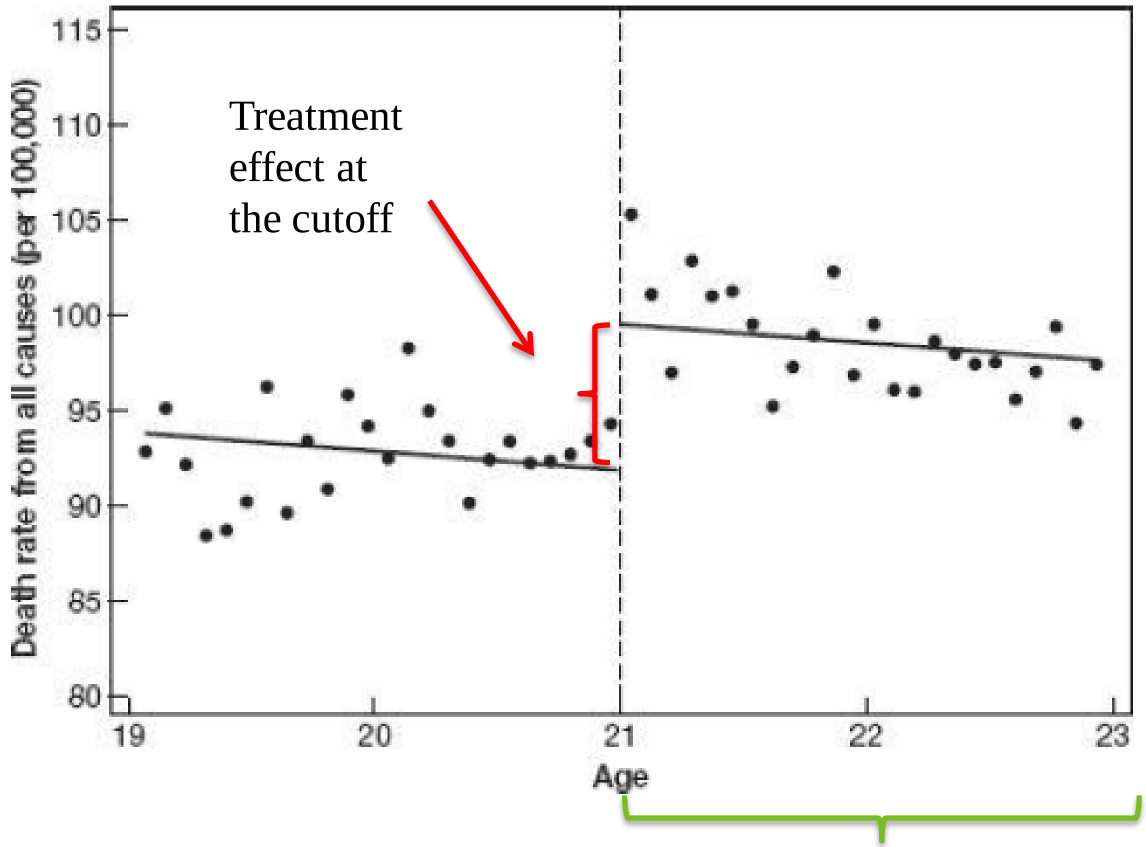
- **Given that units are unable to precisely manipulate the running variable, the RDD can be interpreted as a randomized experiment inside a window around the cutoff**
  - That is, the treatment assignment is locally random
  - Strictly speaking the assumptions for this interpretation are somewhat different than the assumptions under smoothness assumptions
- **This requires a lot of data near the cutoff, but in principle all you need to do is calculating the difference in means**

Technical issues

- **RDD is implemented using regression techniques**
- **In most cases, we do not have enough data to estimate the treatment effect simply by comparing means at the cutoff so we need to use data away from the cutoff**
- **How much data away from the cutoff should we use?**
  - In other words, how large a bandwidth should we use?
- **The choice involves a bias-variance trade-off:**
  - The closer to the cutoff you are, more likely it is that you are able estimate an unbiased causal effect
  - But at the same time variance or the standard error of your estimate is larger as you are using fewer data points



The bandwidth: the share of observations used in estimating the local linear regression:
$E[death\ rate|age, age < 21] = \alpha_0 + \beta_0 * age$



The bandwidth: the share of observations used in estimating the local linear regression:

$E[death\ rate|age\ ,\ age \geq 21] = \alpha_1 + \beta_1{*}age$

Does RDD really work?

RDD vs. RCT

- A non-experimental empirical tool meets a very important quality standard if it can reproduce the results from a randomized experiment
- Hyytinen et al. study whether RDD can, in practice, reproduce an experimental estimate obtained by utilizing data from electoral ties between two or more candidates in recent Finnish municipal elections

QUANTITATIVE ECONOMICS
JOURNAL OF THE ECONOMETRIC SOCIETY

Original Articles | Open Access

When does regression discontinuity design work? Evidence from random election outcomes

Ari Hyytinen, Jaakko Meriläinen, Tuukka Saarimaa, Otto Toivanen, Janne Tukiainen

https://onlinelibrary.wiley.com/doi/epdf/10.3982/QE864

Hyytinen et al. (2018)

- **Question: Can RDD reproduce an experimental estimate?**
- **Application: Is there a personal incumbency advantage in Finnish local elections?**
  - Do candidates who are sitting the municipal council get elected more frequently than candidates who do not?
- **RDD:**
  - Running variable: within party vote share
  - Cutoff: within list
- **Randomized treatment:**
  - Vote ties at the cutoff in which case election status has to be randomized

Party list example

- Define the pivotal number of votes as the average of the maximum number of votes among non-elected candidates and the minimum number of votes among elected candidates (here 50)
- The distance to getting elected is the number of votes of the candidate minus the pivotal number of votes
- Normalize the distance measure by dividing it by the total number of votes of the party list and multiply by 100 => "v_{ipmt}"

Votes	v_{ipmt}	E_{ipmt}
230	25.32	1
182	18.57	1
57	0.98	1
54	0.56	1
50	0.00	1
50	0.00	0
49	-0.14	0
22	-3.94	0
16	-4.78	0
1	-6.89	0

Party list example

Running variable

Votes	v_{ipmt}	E_{ipmt}
230	25.32	1
182	18.57	1
57	0.98	1
54	0.56	1
50	0.00	1
50	0.00	0
49	-0.14	0
22	-3.94	0
16	-4.78	0
1	-6.89	0

Vote ties

Election threshold

Balance test for randomized election outcomes

TABLE 1. Covariate balance tests for the lottery sample.

	Elected ($N = 671$)			Not Elected ($N = 680$)					
Variable	N	Mean	Std. Dev.	N	Mean	Std. Dev.	Difference	p-Value	p-Value (Clustered)
Vote share	671	1.54	0.69	680	1.53	0.67	0.00	0.93	0.97
Number of votes	671	41	39	680	41	38	0	0.83	0.93
Female	671	0.39	0.49	680	0.38	0.49	0.01	0.80	0.80
Age	671	45.42	11.87	680	45.69	11.54	−0.27	0.67	0.67
Incumbent	671	0.29	0.45	680	0.31	0.46	−0.02	0.34	0.35
Municipal employee	671	0.24	0.43	680	0.25	0.44	−0.01	0.62	0.62
Wage income	478	22,521	14,928	476	22,256	13,729	265	0.78	0.82
Capital income	478	2929	18,612	476	3234	12,085	−305	0.76	0.81
High professional	671	0.18	0.38	680	0.18	0.38	0.00	0.97	0.97
Entrepreneur	671	0.24	0.43	680	0.24	0.43	0.00	0.84	0.87
Student	671	0.02	0.15	680	0.03	0.16	0.00	0.76	0.76
Unemployed	671	0.06	0.24	680	0.05	0.22	0.01	0.37	0.37
University degree	537	0.13	0.34	545	0.13	0.34	0.00	0.86	0.86

Incumbency advantage using randomized election outcomes

Table 2. Experimental estimates of the personal incumbency advantage.

	(1)	(2)	(3)	(4)
Elected	0.004	0.001	−0.010	−0.010
95% confidence interval (robust)	[−0.046, 0.054]	[−0.049, 0.051]	[−0.064, 0.040]	[−0.060, 0.040]
95% confidence interval (clustered)	[−0.044, 0.053]	[−0.048, 0.050]	[−0.067, 0.047]	[−0.075, 0.055]
N	1351	1351	1351	1351
R^2	0.00	0.03	0.28	0.44
Controls	No	Yes	Yes	Yes
Municipality fixed effects	No	No	Yes	No
Municipality–year fixed effects	No	No	No	Yes

Note: Only actual lotteries are included in the regressions. Set of controls includes age, gender, party affiliation, socioeconomic status and incumbency status of a candidate, and total number of votes. Some specifications include also municipality or municipality–year fixed effects. Confidence intervals based on clustered standard errors account for clustering at the municipality level. The unit of observation is a candidate i at year t.

Incumbency advantage using RDD

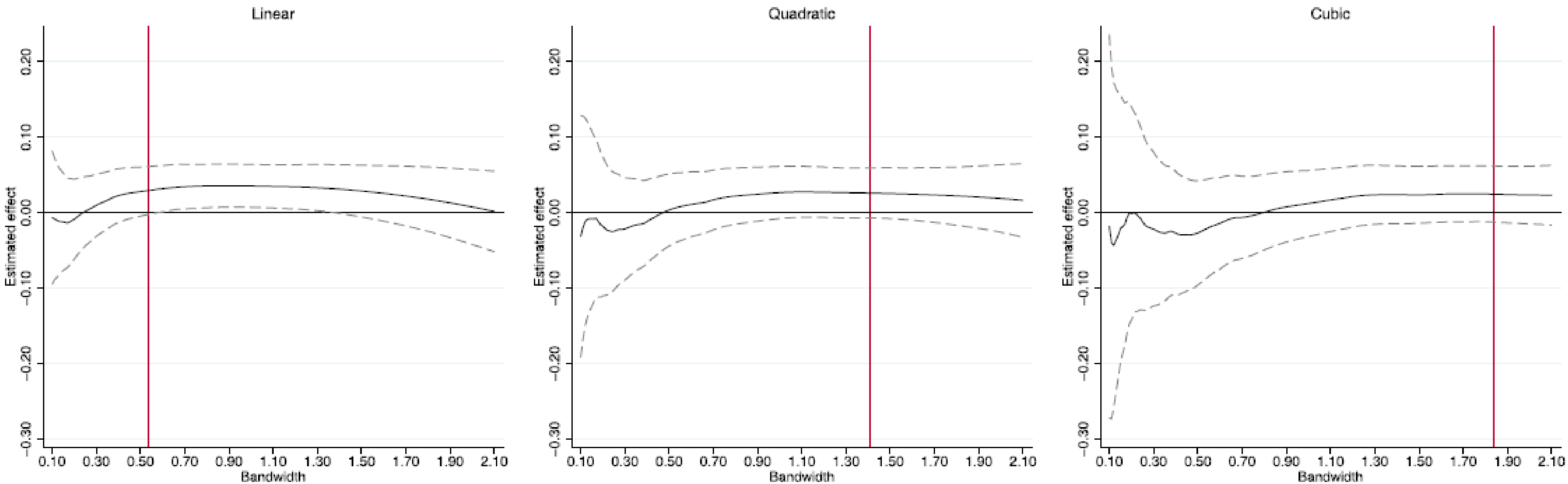


FIGURE 3. Bias-corrected RDD estimates, fixed bias bandwidth. *Notes*: Figure displays bias-corrected point estimates from local polynomial regressions with triangular kernel using various bandwidths. Dashed lines show 95% confidence intervals computed using robust standard errors. Vertical lines mark the IK bandwidth. The bias bandwidth for bias correction has been fixed to 1.14, 1.49 and 1.92 for linear, quadratic and cubic specifications, respectively.

Incumbency advantage using RDD

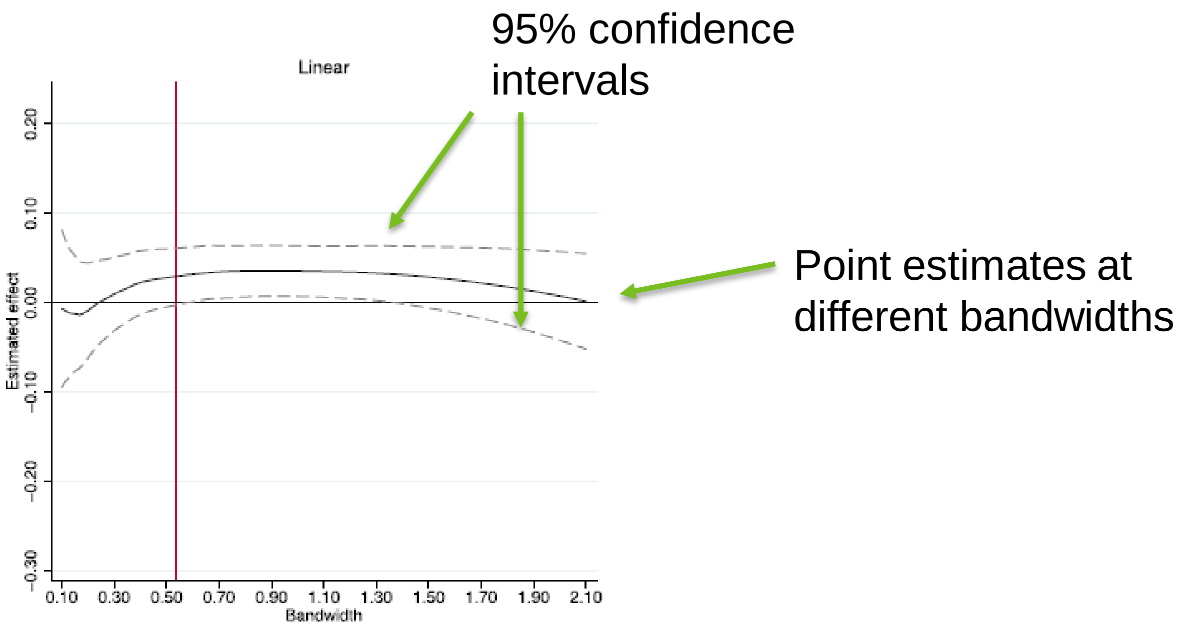


FIGURE 3. Bias-corrected RDD estimates, fixed bias bandwidth. *Notes*: Figure displays bias-corrected point estimates from local polynomial regressions with triangular kernel using various bandwidths. Dashed lines show 95% confidence intervals computed using robust standard errors. Vertical lines mark the IK bandwidth. The bias bandwidth for bias correction has been fixed to 1.14, 1.49 and 1.92 for linear, quadratic and cubic specifications, respectively.

Conclusions – Hyytinen et al. (2018)

- **These findings lead to two key conclusions:**
  1. RDD can indeed meet the replication standard in the context of close elections – reproduces the experimental benchmark.
  2. More interestingly, the results may be sensitive to the details of implementation even when the researcher has a relatively large number of observations. The recently proposed implementation approaches work better than the older ones.

Other examples

Sarvimäki, Uusitalo & Jäntti (2020)

http://www.aalto-econ.fi/sarvimaki/forced.pdf

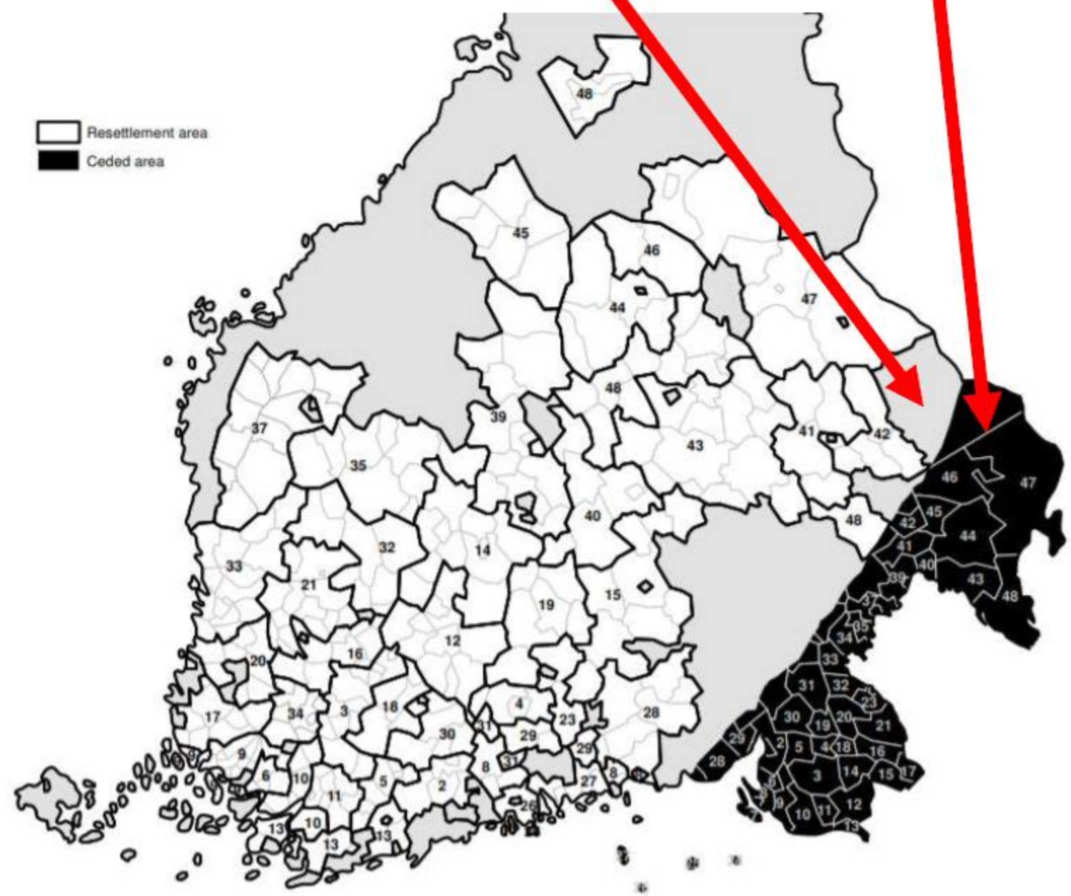
Sarvimäki et al. (2019)

- **11% of the population was forced to migrate and resettled into the remaining parts of Finland**
  - For those working in agriculture roughly one half of the population the government attempted to reconstruct the pre-war conditions as closely as possible
  - Displaced farmers were given land and assistance to establish new farms in areas that had similar soil and climate as the origin regions
  - Former neighbors were resettled close to each other in order to preserve social networks
- **Once the resettlement was completed in 1948, the displaced farmers were not subject to any special policies**
  - They received no further subsidies and, like everyone else, were free to sell and buy land and to move across locations and sectors

Sarvimäki et al. (2019)

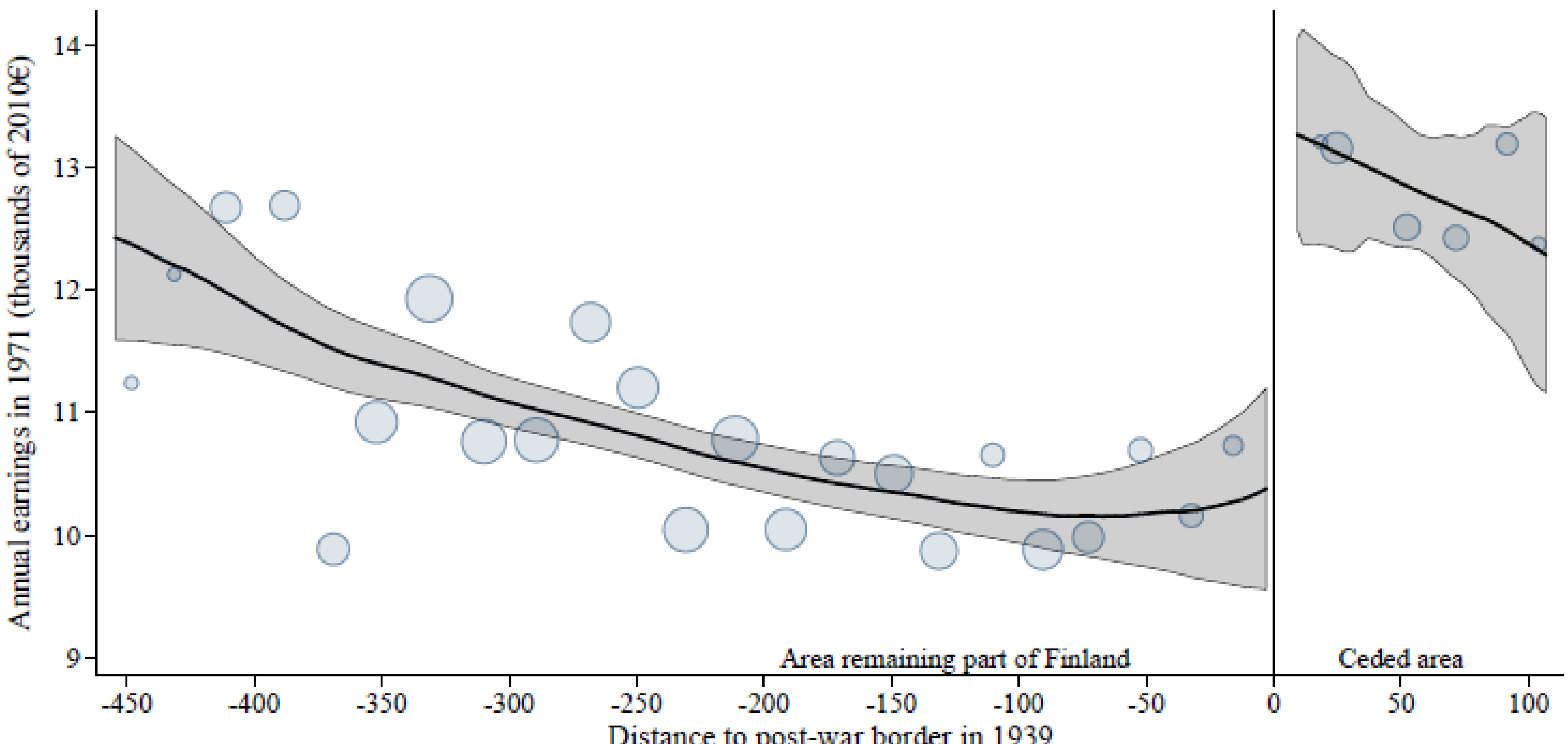
Families in these areas displaced.

Families in these areas can stay where they are.

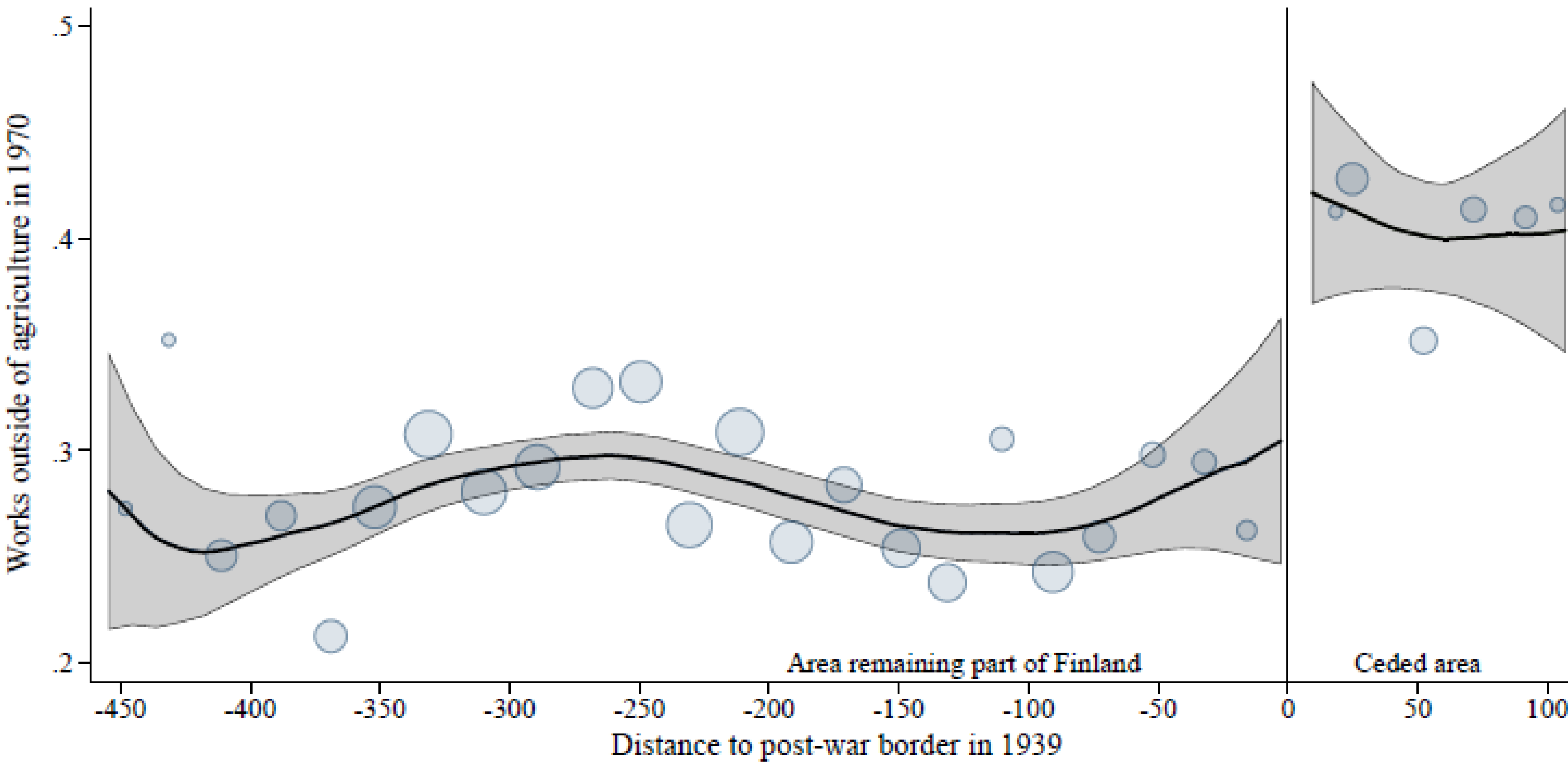


Compare the outcomes of those who live on one side of the border that separate the ceded territory and unceded territory to those on the other side of the border

Sarvimäki et al. (2019) – main results



Sarvimäki et al. (2019) – main results



Conclusions – Sarvimäki et al. (2019)

- The post war difference between displaced and non-displaced farmers suggests that forced migration increased long term income by 10–29% among men working in agriculture before the war
- Forced migration increased the likelihood of leaving agriculture between 1939 and 1970 by 10–16 percentage points from a baseline of 28%
- Increased the likelihood of moving to a city and to complete secondary education among the displaced farmers
- These results suggest that the positive impact of forced migration on the income of farmers can be attributed to an increased likelihood of leaving agriculture

Journal of Urban Economics 75 (2013) 15–28

Contents lists available at SciVerse ScienceDirect

Journal of Urban Economics

www.elsevier.com/locate/jue

Valuing school quality using boundary discontinuities

Stephen Gibbons [a], Stephen Machin [b], Olmo Silva [a,*]

[a] *Department of Geography and Environment and Centre for Economic Performance, London School of Economics, UK*

[b] *Department of Economics, University College London and Centre for Economic Performance, London School of Economics, UK*

ARTICLE INFO

Article history:
Received 28 March 2012
Revised 2 November 2012
Available online 16 November 2012

JEL classification:
C21
I20
H75
R21

Keywords:
House prices
School quality
Boundary discontinuities

ABSTRACT

Existing research shows that house prices respond to local school quality as measured by average test scores. However, higher test scores could signal higher academic value-added or higher ability, more sought-after intakes. In our research, we show that both school value-added and student prior achievement – linked to the background of children in schools – affect households' demand for education. In order to identify these effects, we improve the boundary discontinuity regression methodology by matching identical properties across admissions authority boundaries; by allowing for boundary effects and spatial trends; by re-weighting our data towards transactions that are closest to district boundaries; by eliminating boundaries that coincide with major geographical features; and by submitting our estimates to a number of novel falsification tests. Our results survive this battery of tests and show that a one-standard deviation change in either school average value-added or prior achievement raises prices by around 3%.

https://www.sciencedirect.com/science/article/pii/S0094119012000769#s0110

School choice

- **Often school choice is based on residential location**
  - Sometimes this is an explicit rule: each housing unit is tied to a particular school through catchment areas
  - Sometimes pupil attainment is freer, but residential location is still an important element in school choice (commuting costs etc.)
- **If school quality varies, we should expect this to be reflected in house prices**
  - Good schools can be accessed through the housing market
  - Hypothesis: houses with access to better schools are more expensive (*ceteris paribus*)

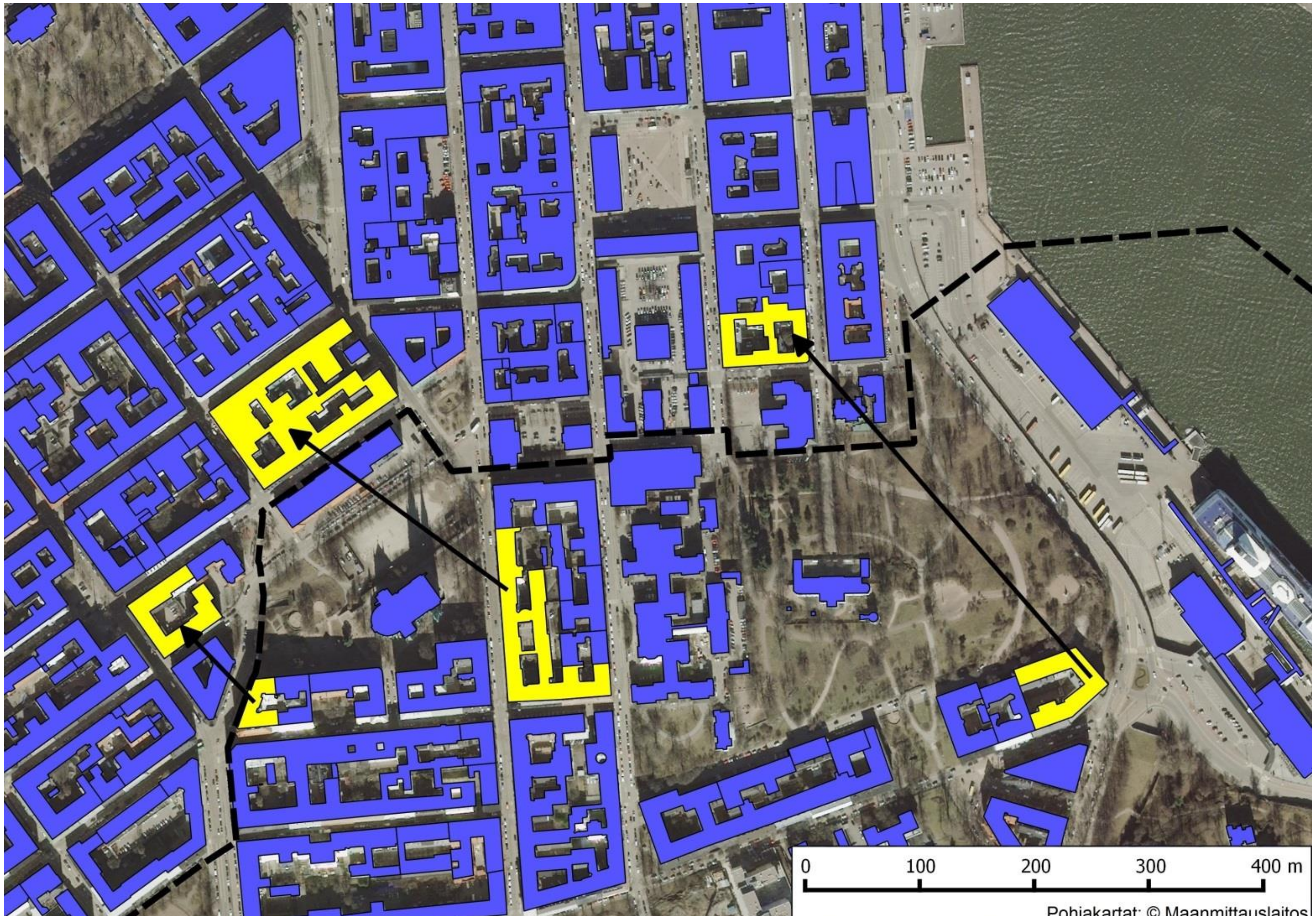
Problem in causal inference

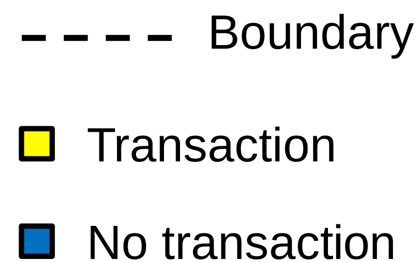
- **The housing market mechanism may lead to a correlation between housing prices and school quality, even if parents do not actually care about school quality**
  - High- and low-income households tend to be segregated and live in different neighborhoods
  - Kids of richer parents may do better in school and house prices are higher in richer neighborhoods => correlation between different measures of school quality and house prices
- **Need to find a way to plausibly fix all other neighborhood attributes that affect prices, but maintain variation in school quality**

Boundary discontinuity design

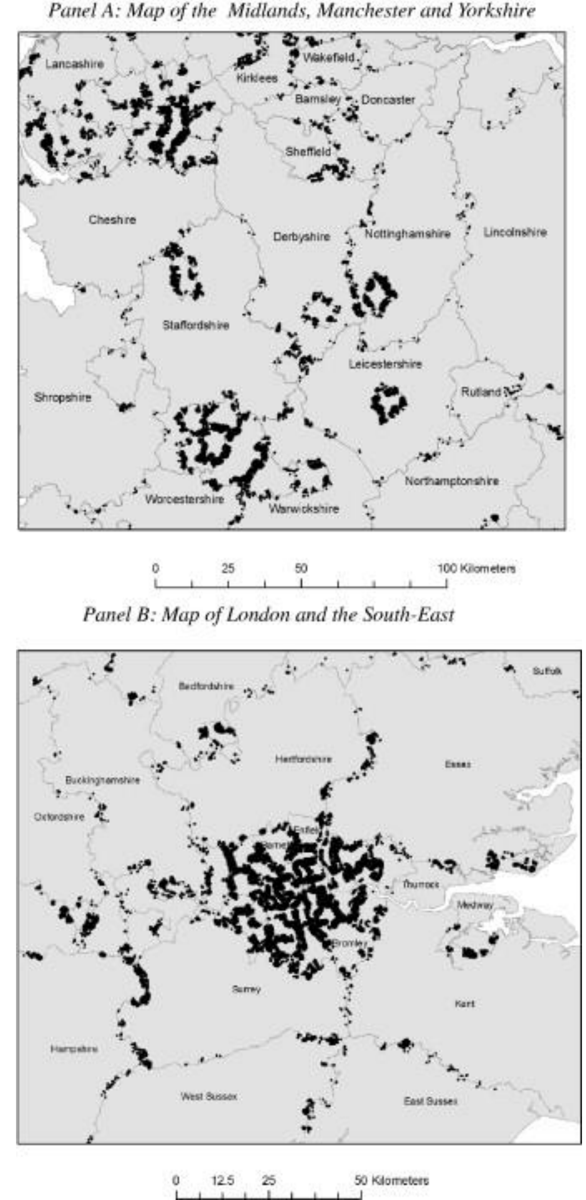
- **Solution: find areas where school quality varies, but neighborhood quality stays fixed**
  - When access to local public goods is spatially bounded there is a discrete change in space in the quality of the public good
- **In this case, a solution to this problem is to concentrate on houses at school catchment area boundaries**
  - Houses near a boundary share the same neighborhood, but the children of the residents are assigned to different schools
  - I.e. neighborhood attributes stay fixed, but there is a difference in school quality

Boundary discontinuity – example





Gibbons et al. (2013) – research design



Gibbons et al. (2013) – main result

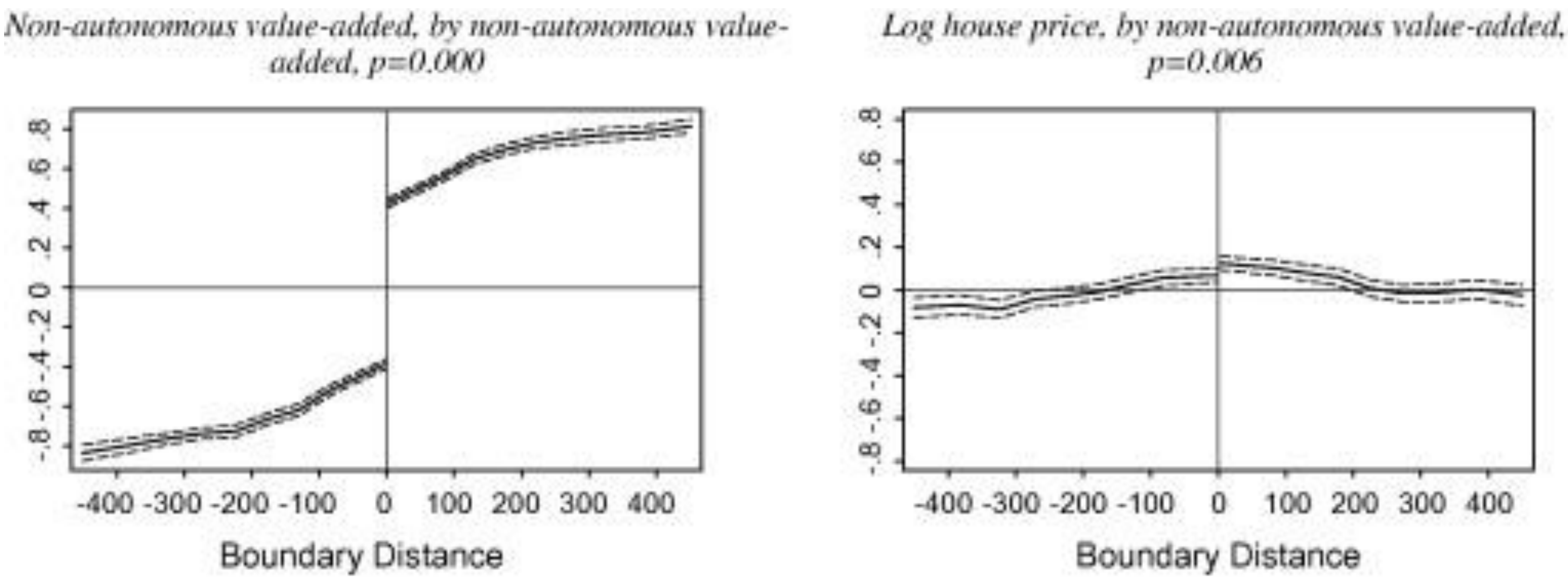
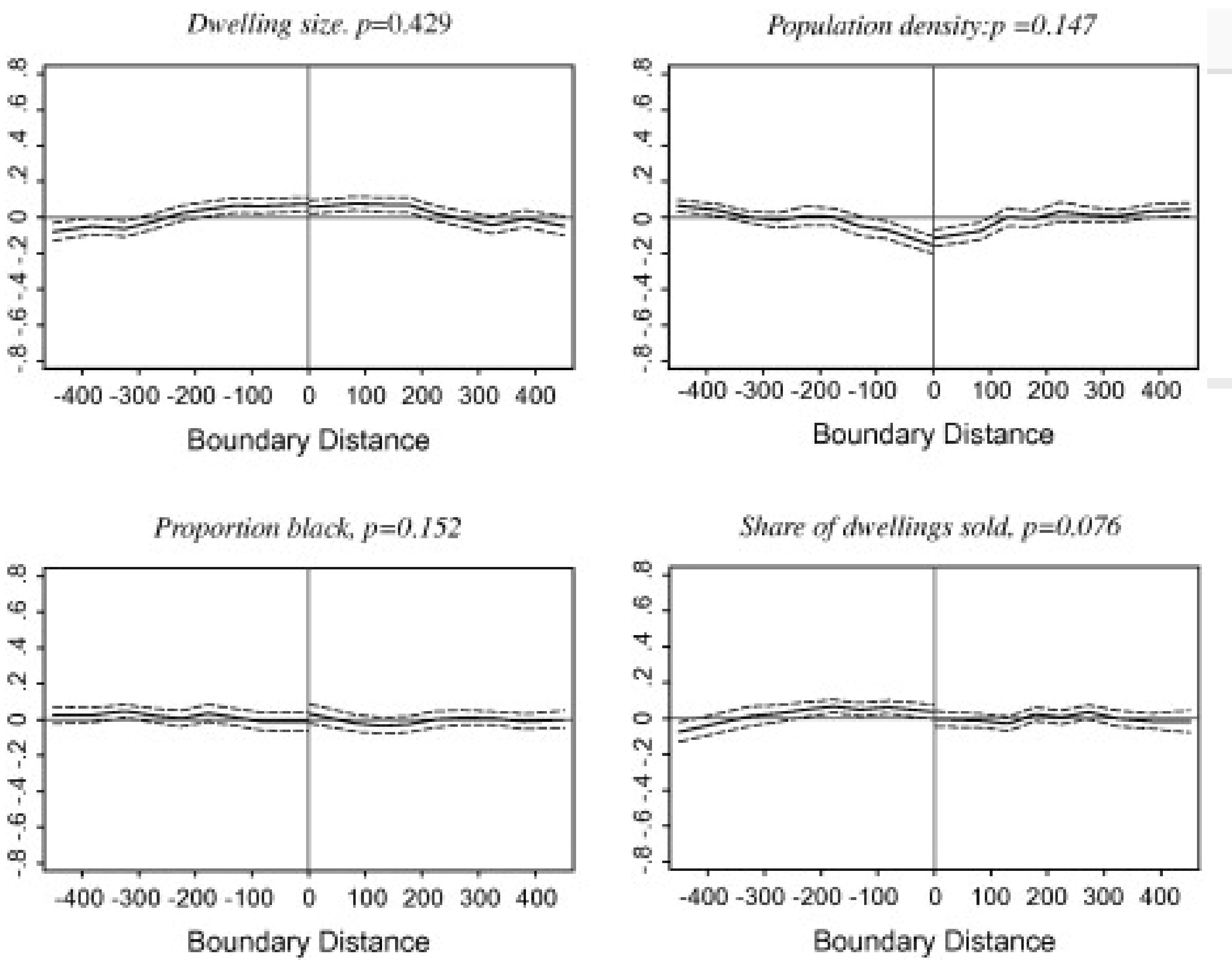
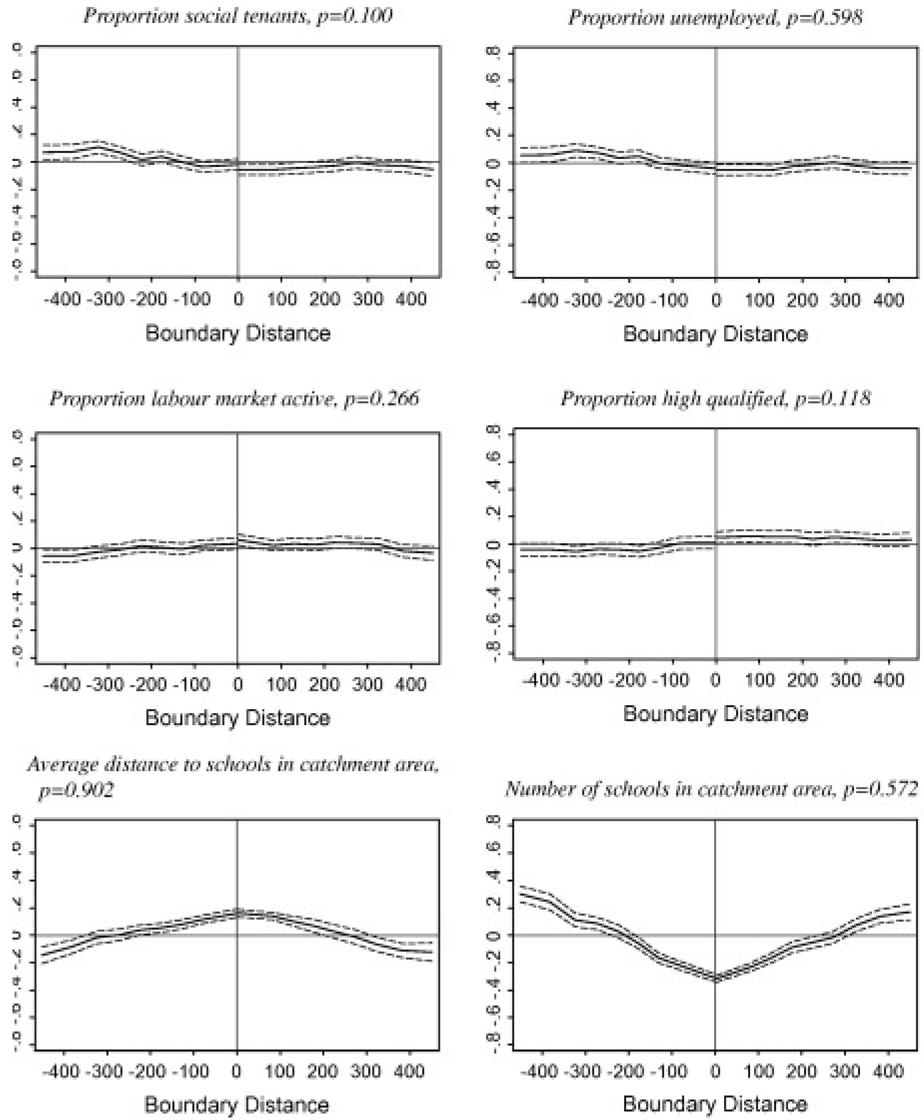


Fig. 2. Discontinuities in non-autonomous school quality and house prices. *Notes*: The scale on the *x*-axis is in metres from the boundary, at the minimum of each bin used in the regressions. The scale on the *y*-axis is in standard deviations.

Gibbons et al. (2013) – validity



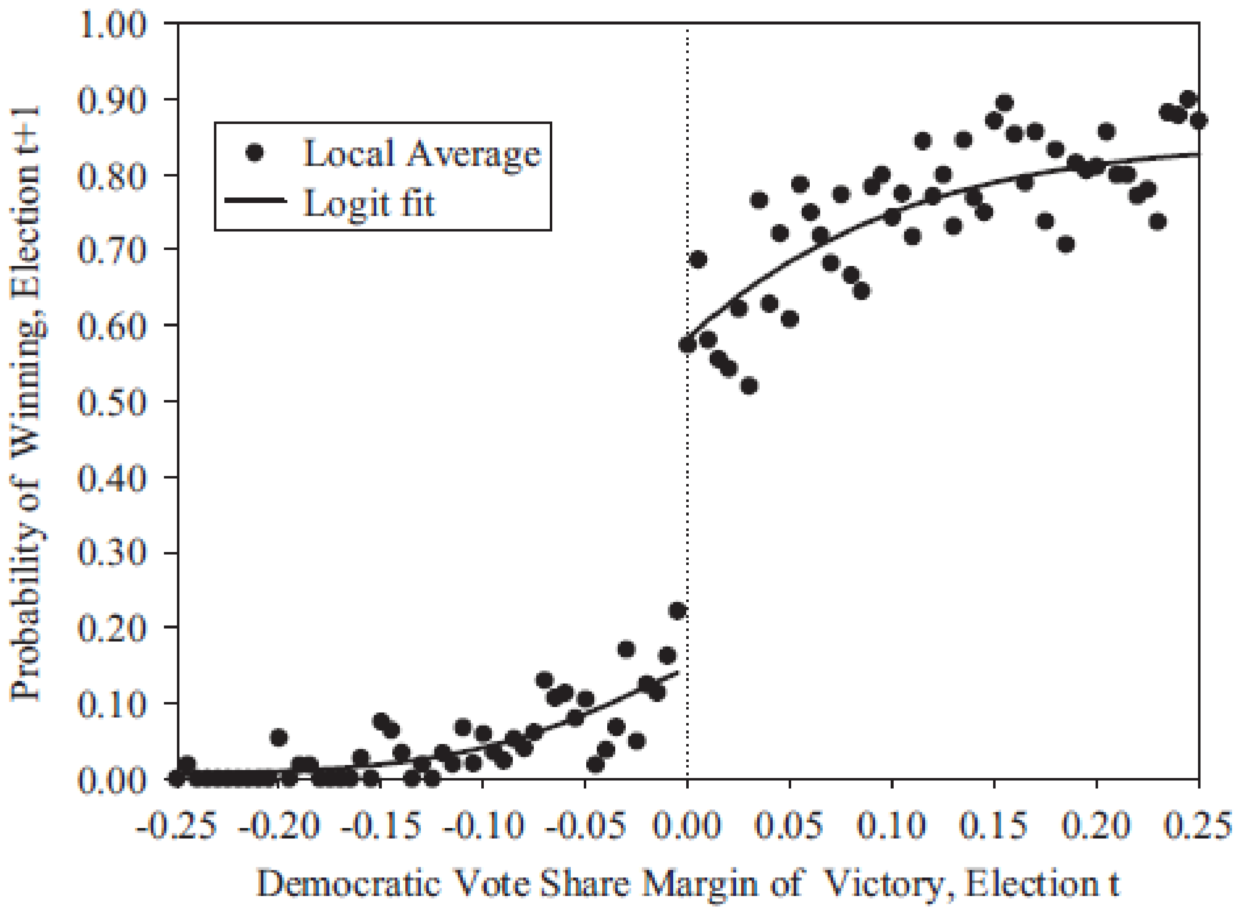
Gibbons et al. (2013) – validity



Conclusions – Gibbons et al. (2013)

- **One-standard deviation change in either age-7 to age-11 school average value-added or prior (age 7) achievement raises prices by around 3% that prioritise students who live close by**
  - There is no house price premium attached to properties close to high quality schools that do not prioritise local students
- **Back-of-the envelope calculations show that the magnitude of this house price response to school quality is plausible as a parental investment decision given the expected return in terms of future earnings of their children**
  - Harjunen et al. (2018) find similar results using data from Helsinki

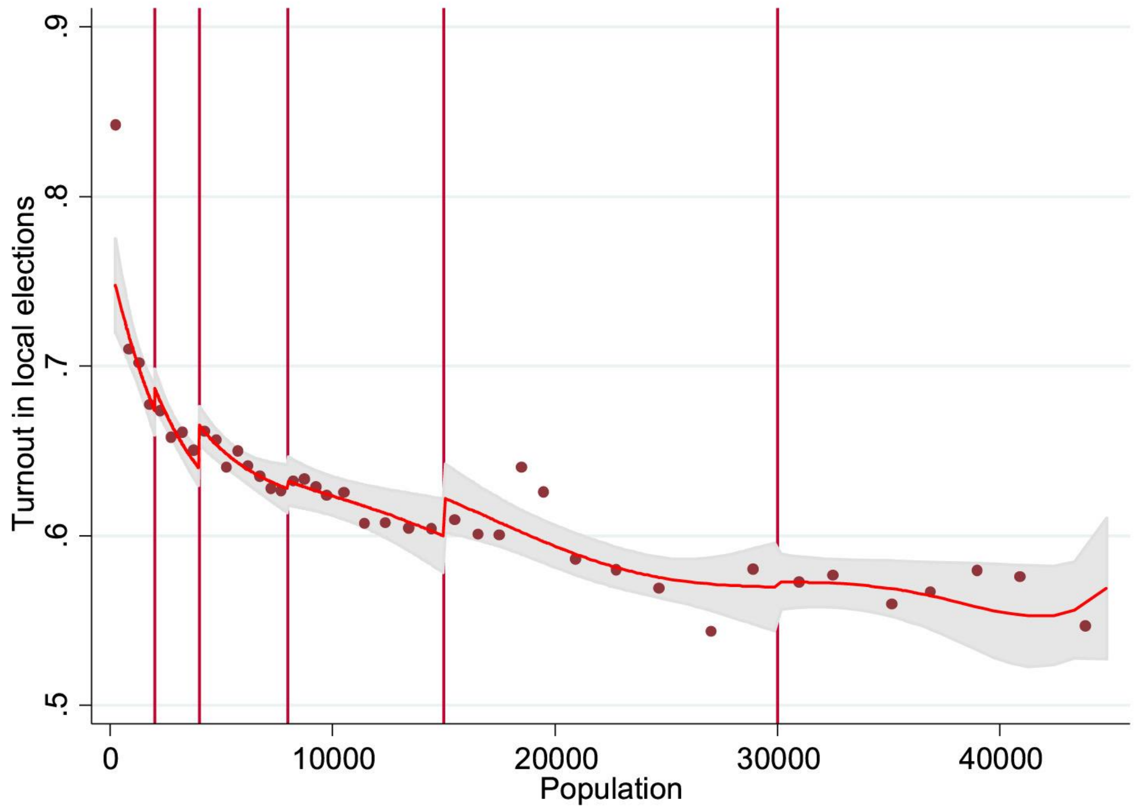
Lee (2008): incumbency advantage



https://www.sciencedirect.com/science/article/pii/S0304407607001121

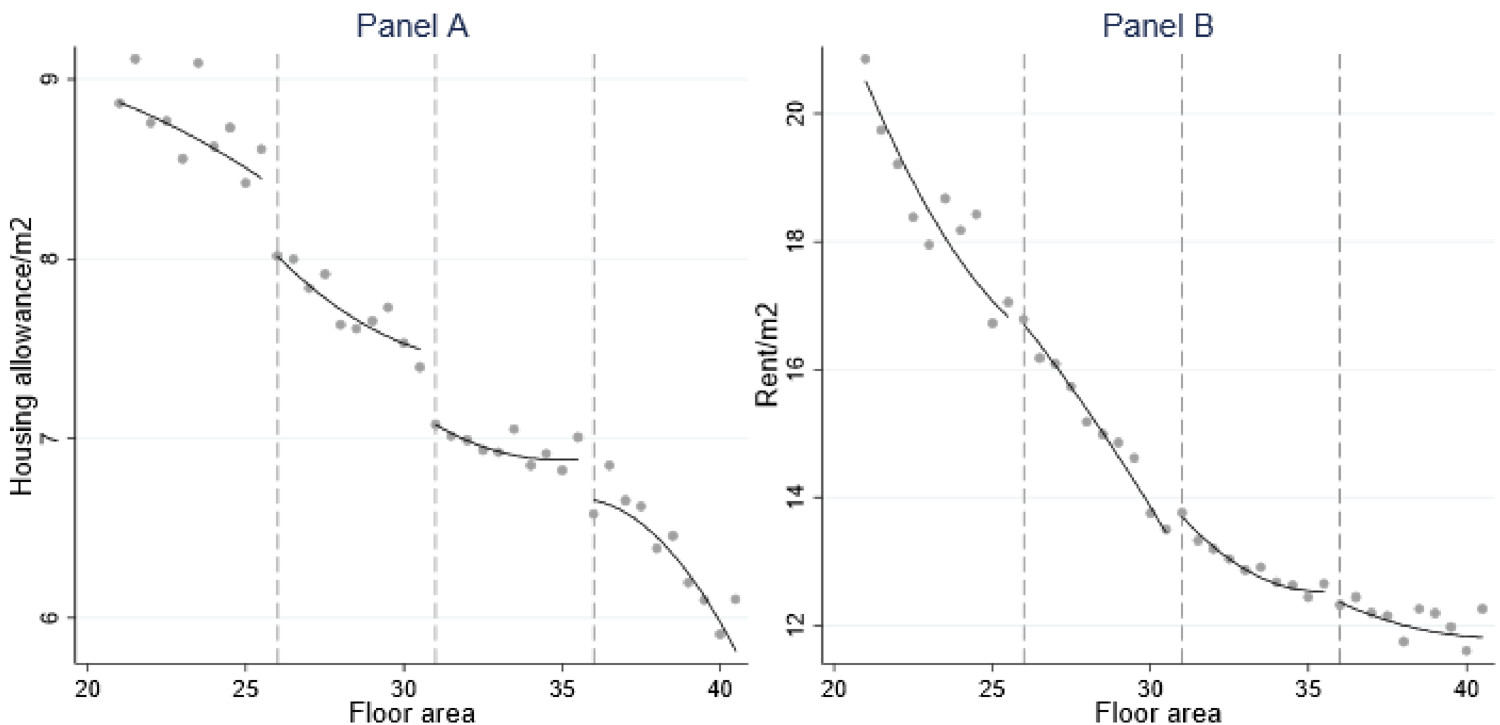
Lyytikäinen & Tukiainen (2019)

Population	Council seats	N
<2000	17 (or 15 or 13)	274
2001-4000	21	465
4001-8000	27	478
8001-15000	35	307
15001-30000	43	168
30001-45000	51	55



https://www.sciencedirect.com/science/article/pii/S0176268018303860?via%3Dihub

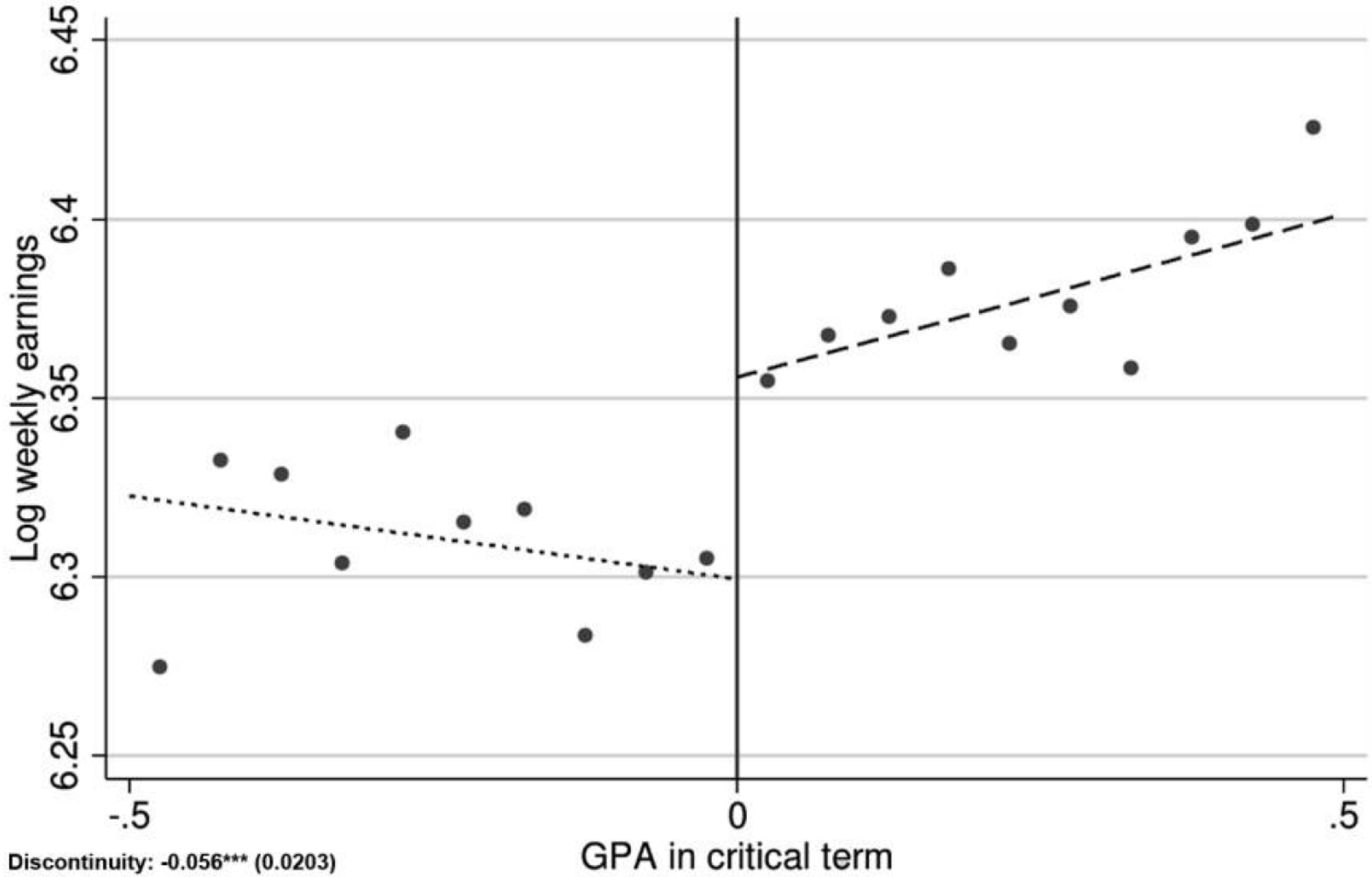
Eerola & Lyytikäinen (2020)



Notes: The figure describes the relationship of HA and floor area (Panel A) and rents of HA recipients and floor area (Panel B). The vertical dashed lines indicate the location of floor area cut-offs, where HA decreases discontinuously. The dots show mean HA/m² or mean rent/m² for floor area bins and the lines are second-order polynomials fitted separately for each interval defined by the cut-offs.

https://onlinelibrary.wiley.com/doi/epdf/10.1111/sjoe.12396

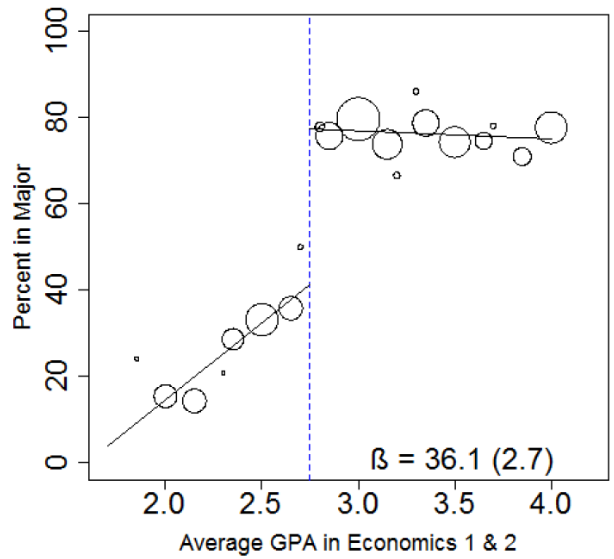
Ost, Pan & Webber (2018)



https://www.journals.uchicago.edu/doi/10.1086/696204?mobileUi=0

Bleemer & Mehta (2020)

Figure 1: The Effect of the UCSC Economics GPA Threshold on Majoring in Economics

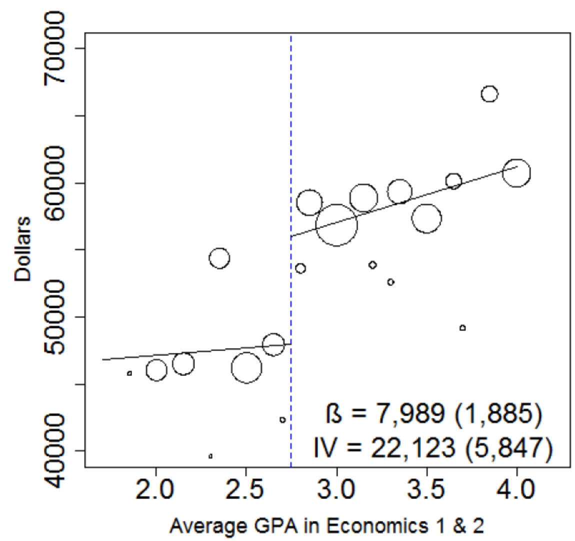


Note: Each circle represents the percent of economics majors (y axis) among 2008-2012 UCSC students who earned a given $EGPA$ in Economics 1 and 2 (x axis). The size of each circle corresponds to the proportion of students who earned that $EGPA$. $EGPA$s below 1.8 are omitted, leaving 2,839 students in the sample. Fit lines and beta estimate (at the 2.8 GPA threshold) from linear regression discontinuity specification; standard error (clustered by $EGPA$) in parentheses. Source: The UC-CHP Student Database.

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3583165

Bleemer & Mehta (2020)

Figure 2: The Effect of the UCSC Economics GPA Threshold on Annual Wages



Note: Each circle represents the mean 2017-2018 wages (y axis) among 2008-2012 UCSC students who earned a given $EGPA$ in Economics 1 and 2 (x axis). The size of each circle corresponds to the proportion of students who earned that $EGPA$. 2017-2018 wages are the mean EDD-covered California wages in those years, omitting zeroes. Wages are CPI-adjusted to 2018 and winsorized at 2% above and below. $EGPA$s below 1.8 are omitted, leaving 2,446 students with observed wages. Fit lines and beta estimate (at the 2.8 GPA threshold) from linear regression discontinuity specification and instrumental variable specification (with majoring in economics as the endogenous variable); standard errors (clustered by $EGPA$) in parentheses. Sources: The UC-CHP Student Database and the CA Employment Development Department.

RDD recap

- **Idea:**
  - If a rule determines treatment due to a hard to predict cutoff, we can use the rule to estimate a causal effect without an RCT
- **Assumption:**
  - Potential outcomes (impossible to observe) develop smoothly across the cutoff
- **Testing for design validity:**
  - Density tests, covariate balance test, placebos
- **Challenges:**
  - Requires a lot of observations near the cutoff
  - Limited external validity