Financial Econometrics II – Cross Section and Panel Data

Regression discontinuity design

Andreas Fuster
Swiss Finance Institute @ EPFL

SFI Léman PhD program – 2025/6, Lecture 4

Regression discontinuity

- **Regression Discontinuity Design (RDD)** is another quasi-natural experiment technique.

- RDD takes advantage of a **known cutoff** or threshold determining treatment assignment or the probability of receiving treatment.
  - For some variable, x, an observation is treated if $x \geq x'$ (or $x \leq x'$, but for exposition I will assume treatment happens for x above threshold – everything would of course be symmetric)

- Cutoff creates a discontinuity in the treatment recipiency rate at that point.

Regression discontinuity

- In an RDD, assignment to treatment and control is not random, but whether **individual** observation is treated is assumed to be as good as random.

- Randomized variation is a consequence so long as agents are **unable to precisely control the assignment** variable near the cutoff.
  - Therefore, whether an observation *x* falls immediately above or below the cutoff *x'* is random

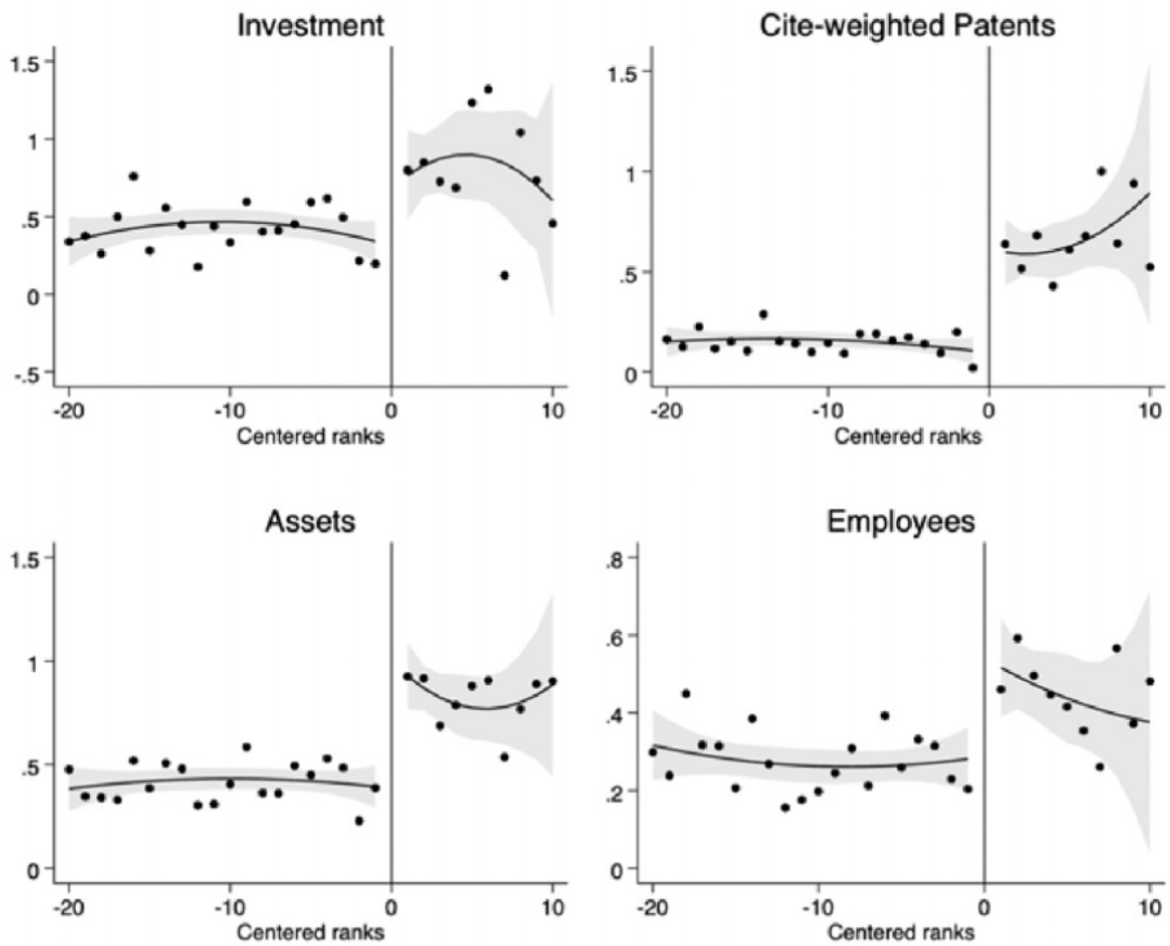
- **Appeal**: (i) relies on relatively mild assumptions relative to other non-experimental techniques; (ii) easily conveyed graphically
- **Limitations**: (i) can be sensitive to estimation details; (ii) very "local" treatment effects; (iii) typically need large sample size

Example

- We want to study the effects of government-sponsored R&D grants on the innovation activity (and other outcomes) of SMEs
- Endogeneity problem: these grants are **not** randomly assigned – firms have to apply and the government will give to the most promising applicants → OLS would be biased
- But: if you can observe the ranking of the projects, you can compare the outcomes of firms that "just got the grant" to those of firms that "just didn't"

- A paper by Santoleri et al. (RESTAT 2024) does so based on EU funding competitions over 2014-17
- Observe for any application its rank in the competition and whether got the grant

Example



Source: “The Causal Effects of R&D Grants: Evidence from a Regression Discontinuity”, Santoleri et al., Review of Economics and Statistics 2024

- Centered ranks > 0 got the grant, < 0 did not
- The treatment effect is estimated from the jump at 0
  - Will see below how to estimate fitted lines & how to do this in regression framework
- Firm-level outcomes (as of 2019) quite clearly affected by receipt of the grant

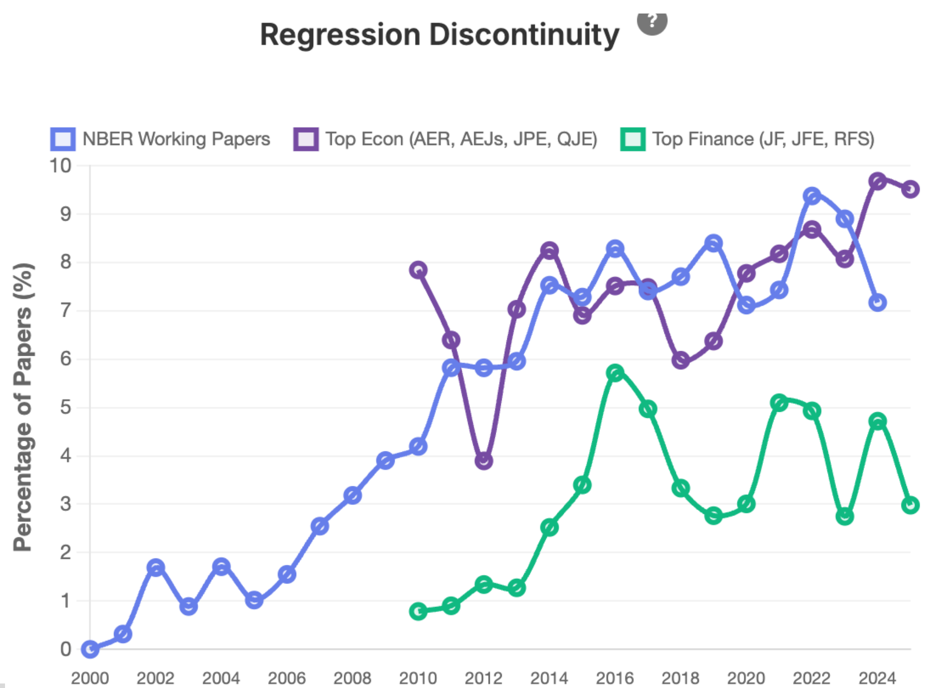
Other examples & popularity

- Many types of cutoffs and applications in finance:

1. Credit scores
2. Eligibility for certain government programs, etc.
3. Thresholds in financial covenants
4. 50% threshold in elections/votes

- RDD have become very popular in econ, but a bit less in finance

(Chart from https://paulgp.com/econlit-pipeline/dashboard.html, cf. lecture 1)



Overview

- **Sharp and fuzzy RDD**
- Estimating RDD
- Checking validity

RDD – definitions

- x is typically called **forcing variable** (or running variable, or score)
- x' is called **threshold**
- $y(0)$ is the (potential) outcome variable absent treatment
- $y(1)$ is the (potential) outcome variable with treatment

- We distinguish two types of RDD:

  — **Sharp RDD:** Assignment to treatment solely based on a cutoff value of an observable variable

  — **Fuzzy RDD:** value of *x* above threshold ($x \geq x'$) increases **probability** of treatment

Sharp RDD

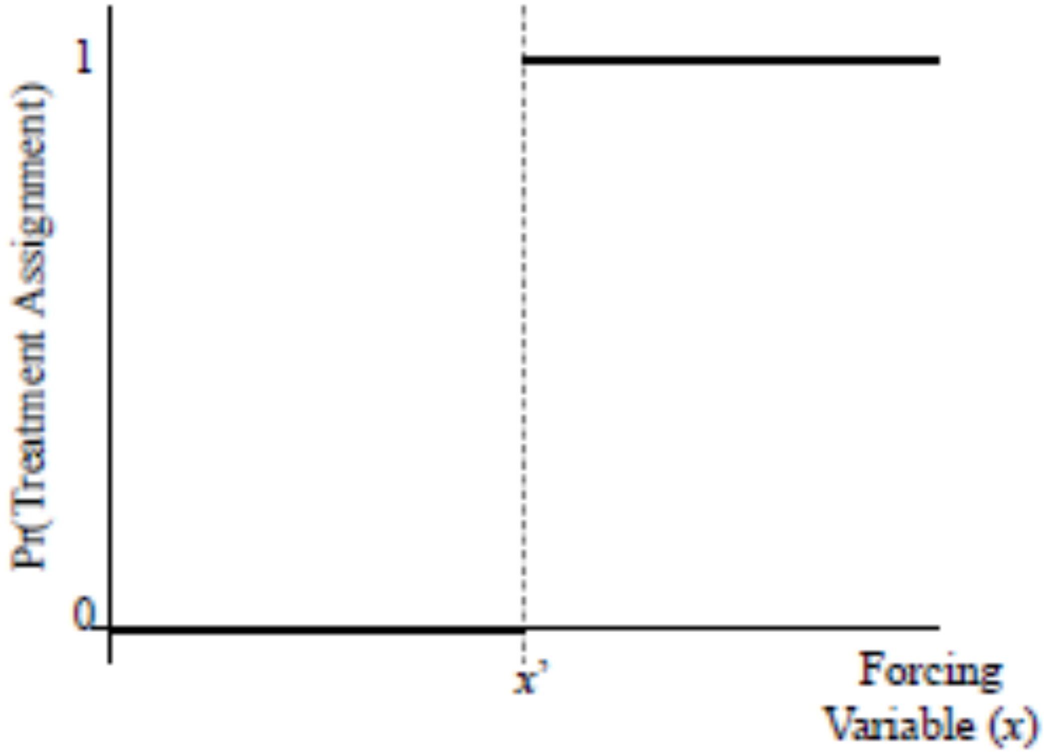
- **Assumption 1:** Assignment to treatment occurs through known and deterministic decision rule:

$$d = d(x) = \begin{cases} 1 & if\ x \geq x' \\ 0 & otherwise \end{cases}$$

- Where x is the forcing variable and x' is the threshold.
- It is important that there exist x around the threshold value.

- Example – Chava and Roberts (2008): a firm's status changes from "not in violation of covenants" to "in violation" when debt-to-EBITDA ratio crosses a threshold.

Sharp RDD



- Source: Roberts and Whited (2012)

Sharp and fuzzy RDD

- **Assumption 2: local continuity**
- Both potential outcomes $y(0)$ and $y(1)$, conditional on forcing variable x, are continuous at the threshold x'.

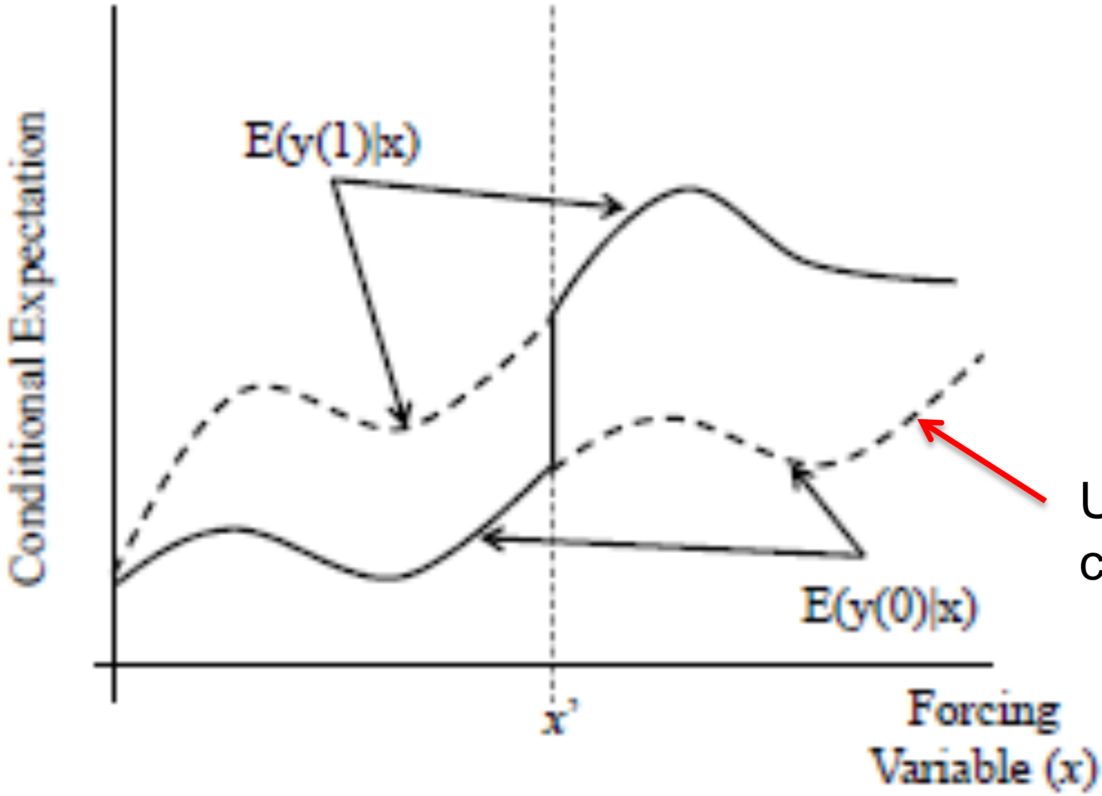
- **Interpretation**: If there were **no treatment**, *y* would be a smooth function around the threshold *x'*

- **In other words**: The average potential outcome is similar for subjects close to but on different sides of the threshold, i.e. in the absence of treatment, outcomes would be similar.

- This assumption needs to hold for both **sharp and fuzzy RDD**.

Sharp and fuzzy RDD

Local continuity ensures that the only reason for different outcomes around the threshold is the treatment



Unobserved counterfactual

- Source: Roberts and Whited (2012)
- Implication: we will only learn about treatment effect for subjects close to threshold

Fuzzy RDD

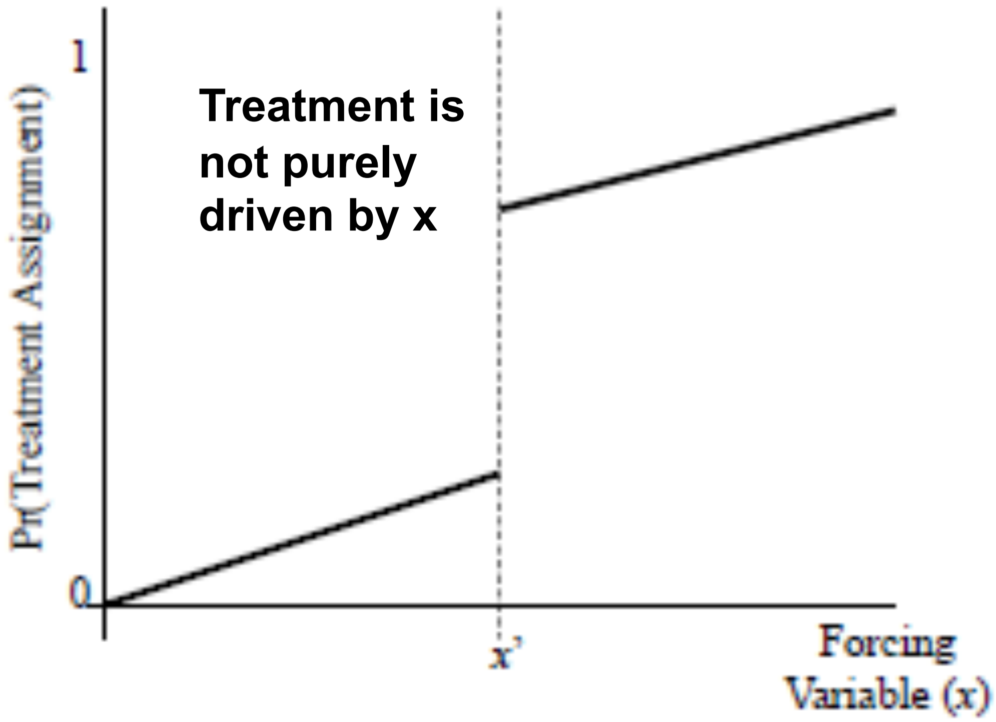
- **Assumption 1':** Assignment to treatment occurs in a ***stochastic*** manner where the ***probability*** of assignment has a known discontinuity at x'

$$0 < \lim_{x \downarrow x'} \Pr(d = 1|x) - \lim_{x \uparrow x'} \Pr(d = 1|x) < 1$$

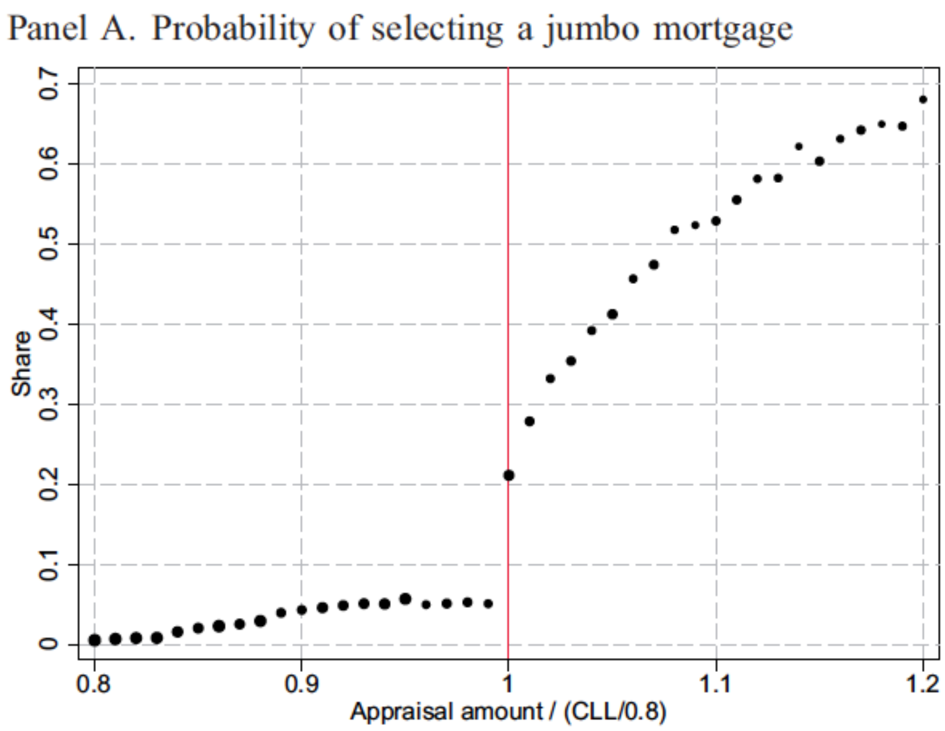
- What we need is a jump in the probability of treatment of less than one at the cutoff.

- Effectively, this will provide us with an **instrumental variable** for treatment d
  - if the jump at the threshold is very small, we have a weak IV

Fuzzy RDD



Source: Roberts and Whited (2012)



Source: Fuster and Vickery (2015)

Where is threshold coming from?

- Usually the threshold *x'* is chosen based on some "institutional features"
  - e.g. voting rules, contracts, published guidelines,…
- But in some cases, can use a data-driven approach – i.e. look for discontinuities without knowing ex-ante where they are
- Examples: Agarwal et al. (QJE 2018); Argyle et al. (RFS 2020) – both using credit score cutoffs of consumer lenders
- Note: this is mostly useful as "first stage", not for main outcome
  - otherwise would worry (even more) about data mining

Overview

- Sharp and fuzzy RDD
- **Estimating RDD**
- Checking validity

Estimation – graphical analysis

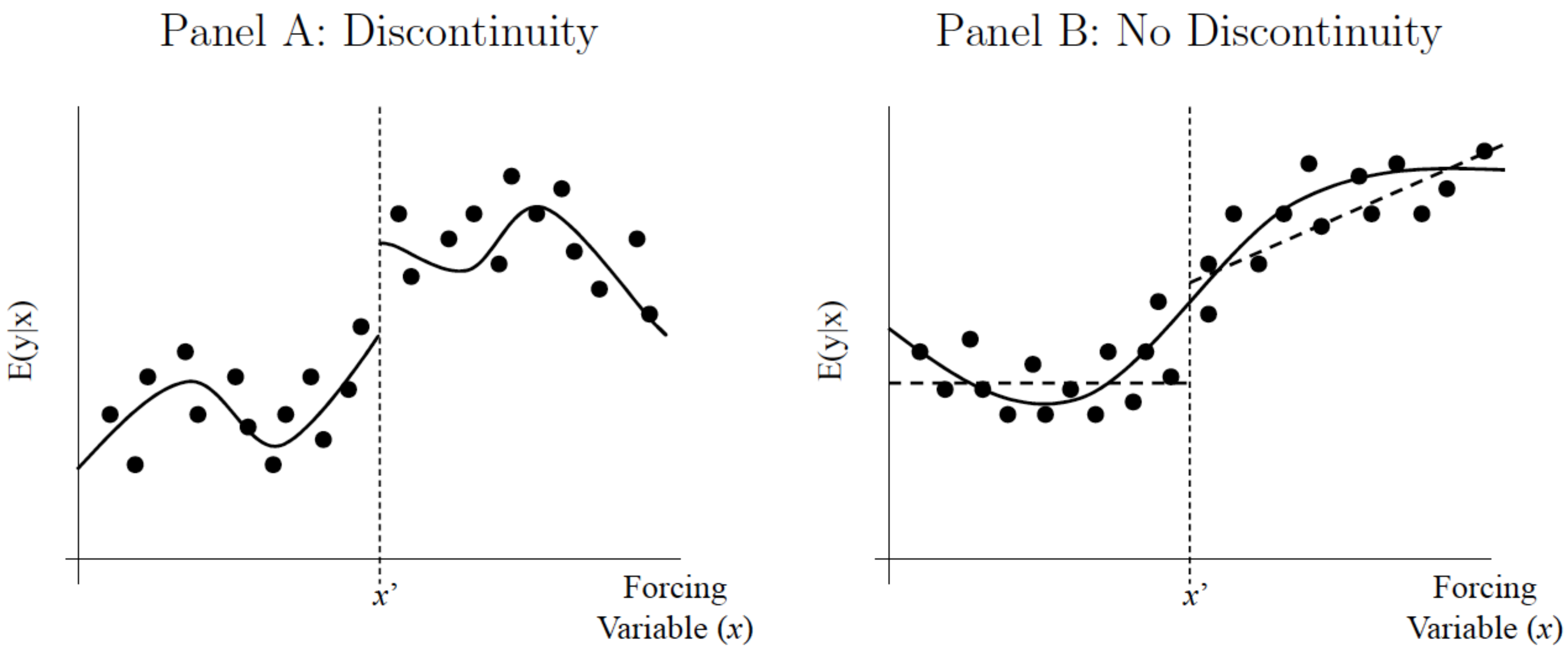
s:fi

- A good place to start is **graphical analysis**. This allows us to visually inspect whether a discontinuity exists at *x'*.

- Divide the domain of *x* into bins (as for a histogram).
- Make sure that the cutoff threshold *x'* does NOT fall into a bin.
  - This allows to make sure that treatment and control observations are not mixed together into one bin by the researcher, though this may occur naturally in a fuzzy RDD.

- Then calculate the average *y* in each bin and plot the average for each bin.
  - For fuzzy RDD: also plot the average d for each bin.
- These plotted averages represent a non-parametric estimate of $E(y|x)$

Estimation – graphical analysis

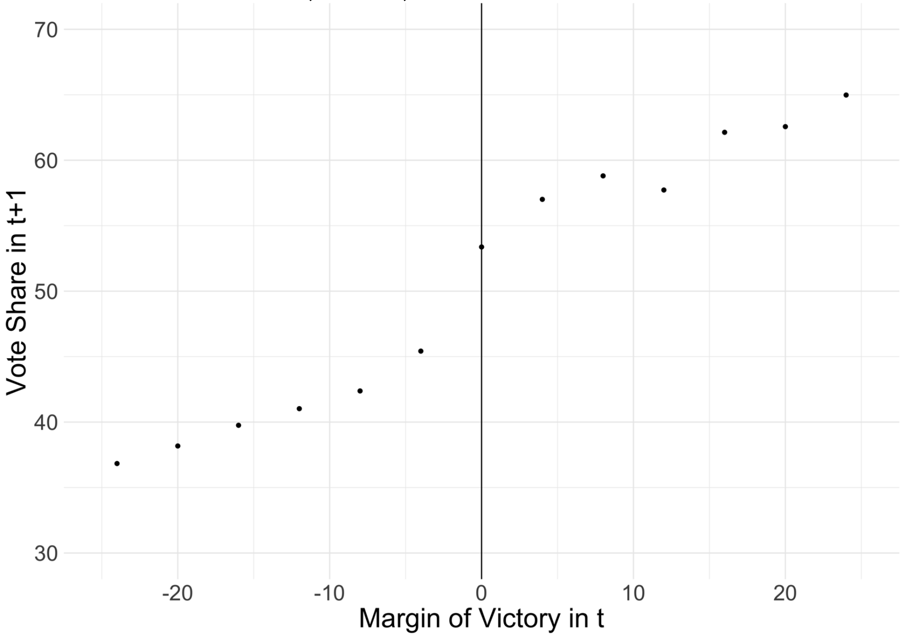


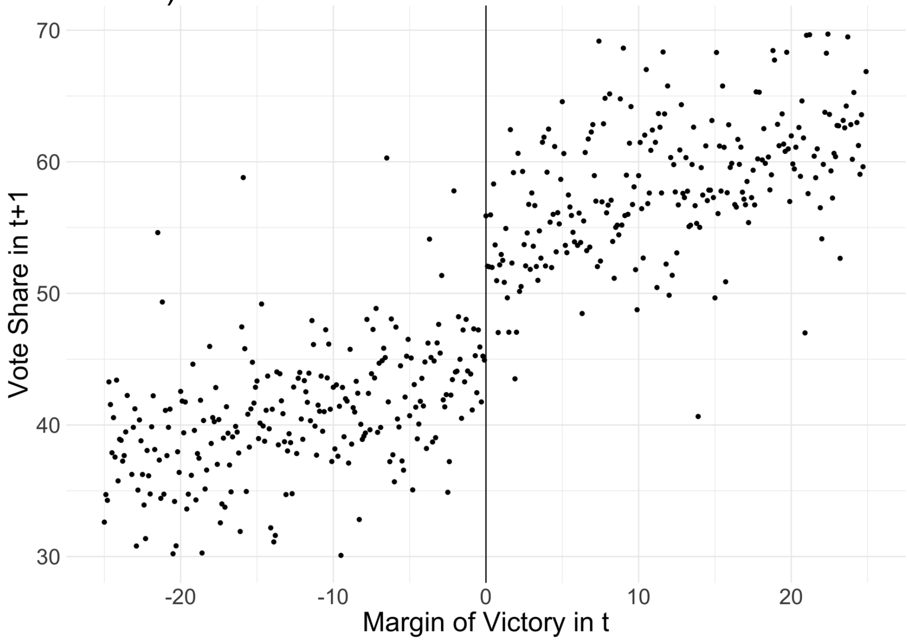
Source: Roberts and Whited (2012)

How to pick number and width of bins?

Left: bins of 4%. Right: bins of 0.1%.

(Source: https://github.com/paulgp/applied-methods-phd/blob/main/lectures/21_regression_discontinuity_2.pdf, based on data from Lee, 2008, who studies incumbents' advantage in elections)





Trade-off: bias (more bins helps get closer to the "true" conditional local means) vs. noise (fewer bins increases observations within bins, lowering the SE for a bin)

How to pick number and width of bins?

- Second decision is whether to make equal-spaced bins, or by quantile (reflecting the underlying distribution)
  - or: could make equal-spaced but with size of dots reflecting # of observations
- See “rdplot” in https://rdpackages.github.io/rdrobust/ for a theoretically motivated approach – good as a baseline choice
- (More broadly, Calonico et al.’s https://rdpackages.github.io/ contains a number of state-of-the-art packages for Stata, R, and Python. For overview: https://rdpackages.github.io/references/Cattaneo-Idrobo-Titiunik_2020_CUP.pdf and NBER 2021 methods lectures: https://www.nber.org/conferences/si-2021-methods-lecture-causal-inference-using-synthetic-controls-and-regression-discontinuity)
  - see also https://blogs.worldbank.org/impactevaluations/how-should-you-draw-rdd-graph (summarizing Korting et al. QJE 2023 paper on visual inference in RD)

Estimation

- Recall that we are relying on the local continuity assumption around the threshold x'

- We therefore want to estimate the outcome Y at the threshold x' “coming from below” vs. “coming from above”, with the difference being the estimated treatment effect

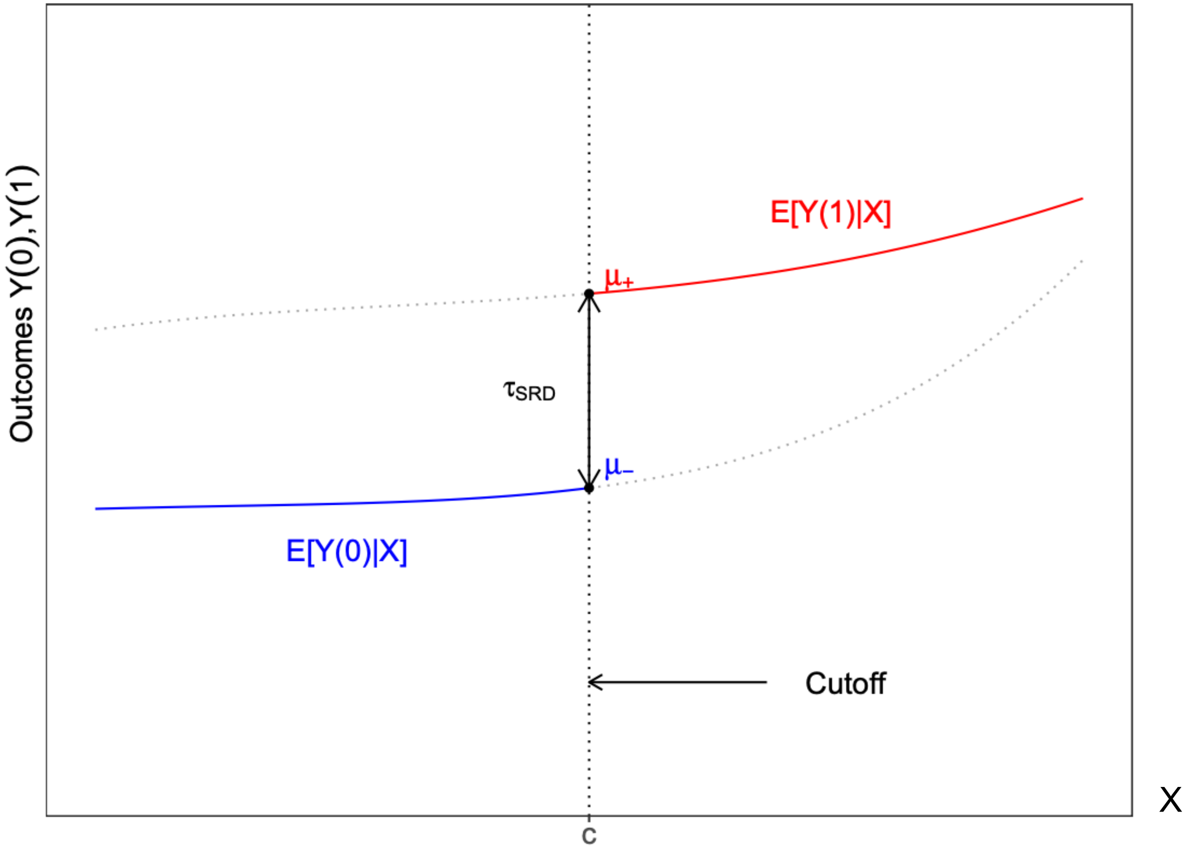
- The problem is that we typically only have limited data right around x'

- Need to use “nearby observations” and extrapolate based on estimated function
  - — how “nearby”? what function to estimate?

Estimation

$$\tau_{\text{SRD}} = \underbrace{\mathbb{E}[Y_i(1) - Y_i(0)|X_i = c]}_{\text{Unobservable}} = \underbrace{\lim_{x \downarrow c} \mathbb{E}[Y_i|X_i = x]}_{\text{Estimable}} - \underbrace{\lim_{x \uparrow c} \mathbb{E}[Y_i|X_i = x]}_{\text{Estimable}}$$

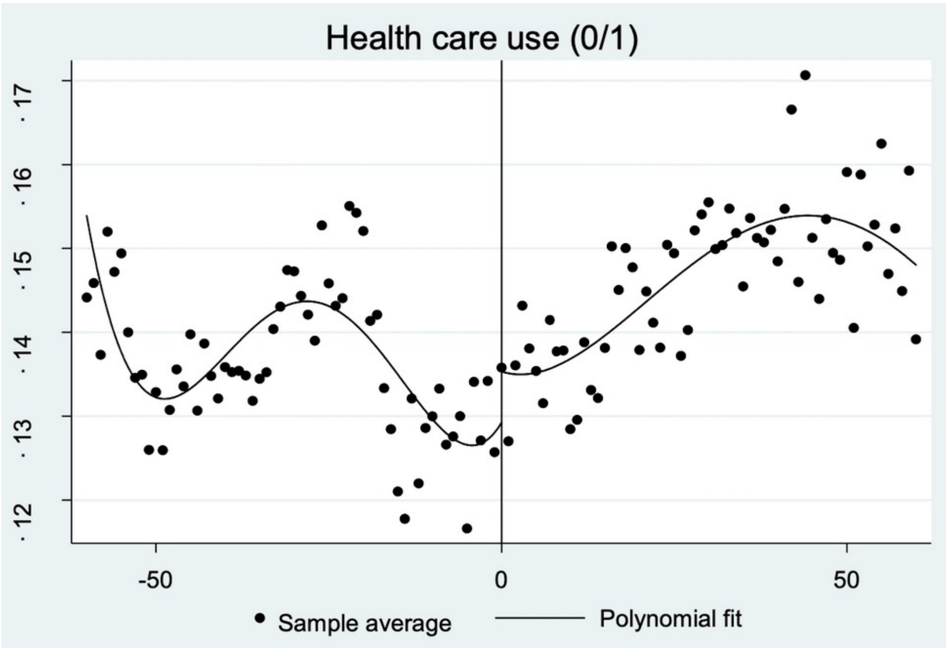
Object of interest: ATE at cutoff

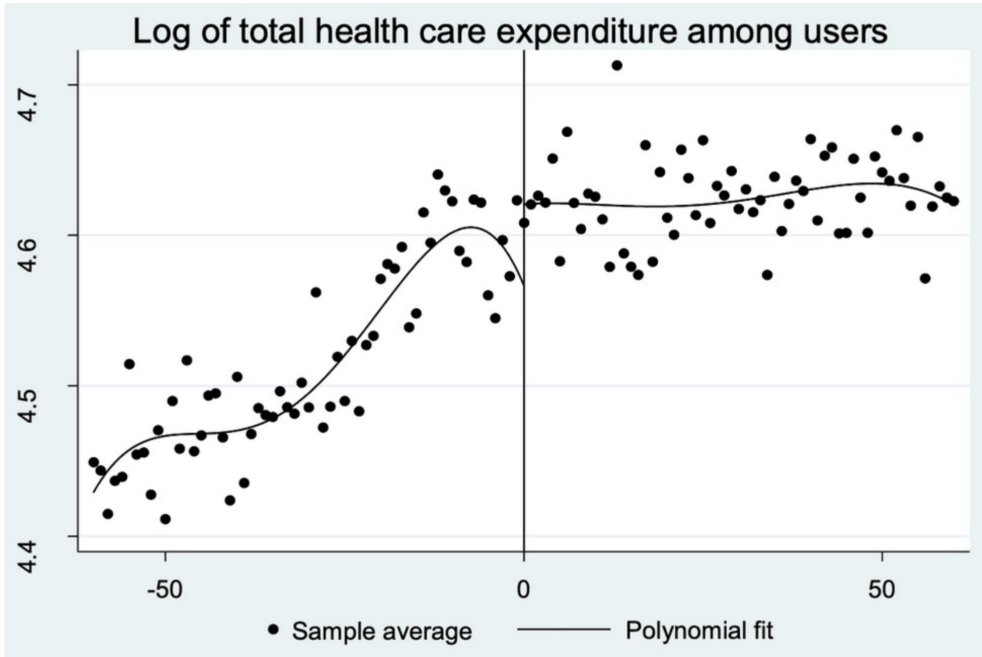


Source: https://conference.nber.org/confer/2021/SI2021/ML/RDD.pdf

Bad idea for estimation: global high-order polynomials

- High (e.g. 4th) order polynomials are fine for visualizing the variation across full range of x, but a bad idea for estimation, due to **overfitting** and **bad properties especially at endpoints**, which we are most interested in. (See Gelman and Imbens, JBES 2019)





https://statmodeling.stat.columbia.edu/2020/12/27/rd-gullible/

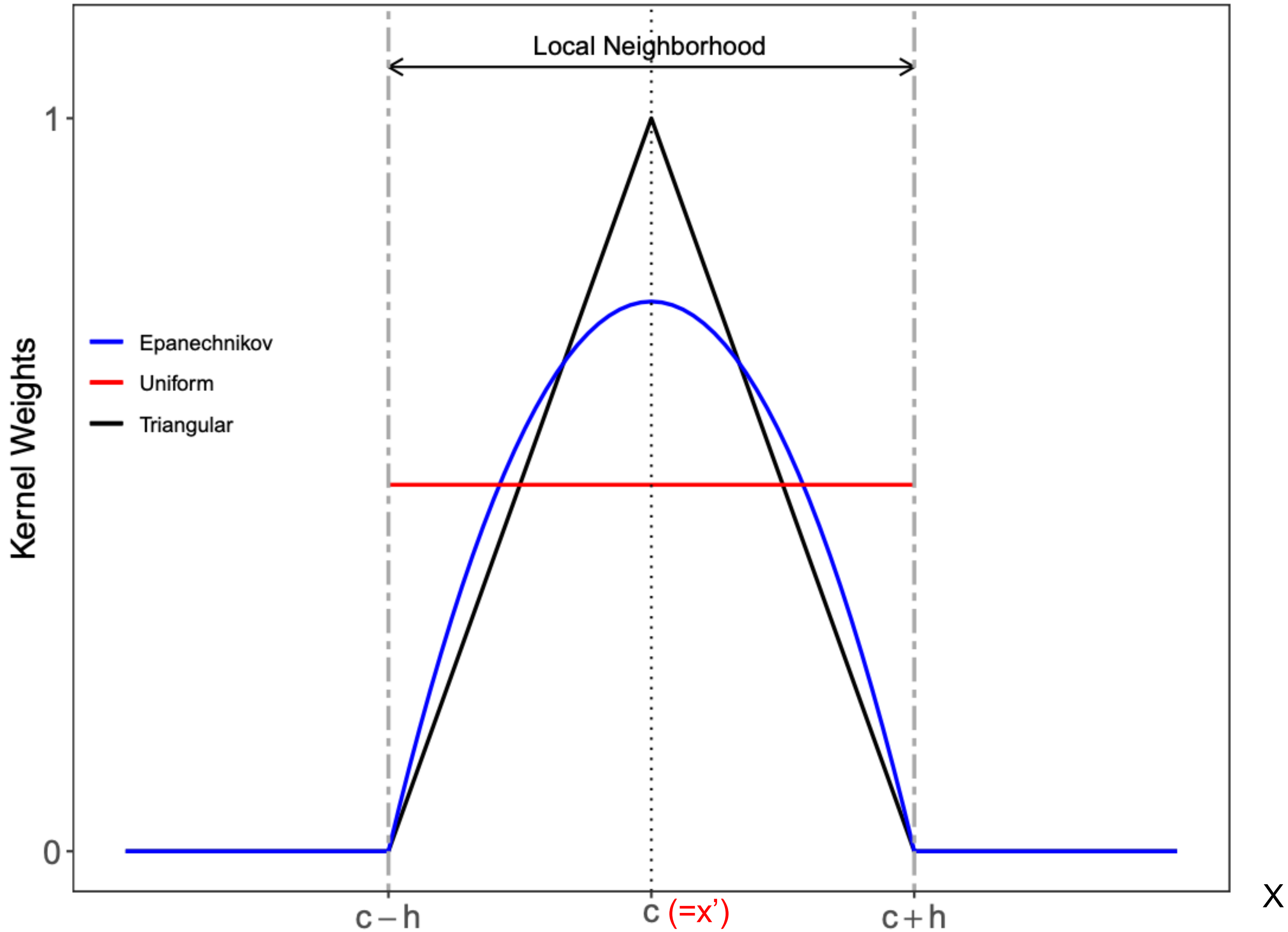
Local polynomial regression

- Instead, preferred approach is to use low-order polynomials (typically *p* = 1, meaning linear regression, or *p* = 2)…
- …in an area close to the threshold ("bandwidth" *h*)…
- …potentially giving more weight to observations near the threshold ("regression kernel" *K*)
- E.g. linear regression (*p* = 1) – WLS with weights K(.):

$$y_i = \alpha + \beta d_i + \gamma^b (x_i - x') + \gamma^a d_i(x_i - x') + \varepsilon_i \quad \text{for } |x_i - x'| \leq h$$

- d_i is indicator for $x \geq x'$, β equals the treatment effect

Kernels



Source: https://conference.nber.org/confer/2021/SI2021/ML/RDD.pdf

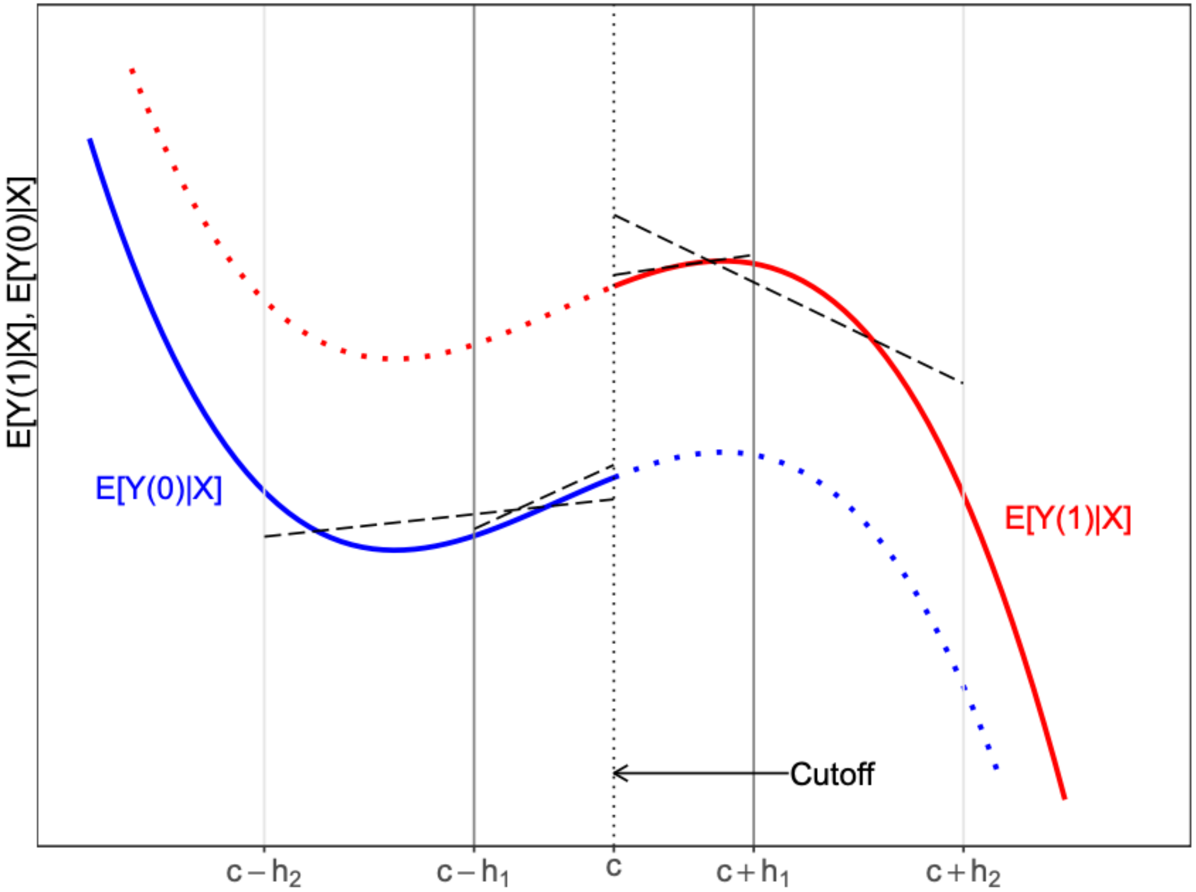
Triangular and uniform are most common choices (they both have certain optimality properties).

Uniform is simplest (no weighting), so may be sensible baseline choice.

Worth trying both to assess robustness.

Bandwidth choice

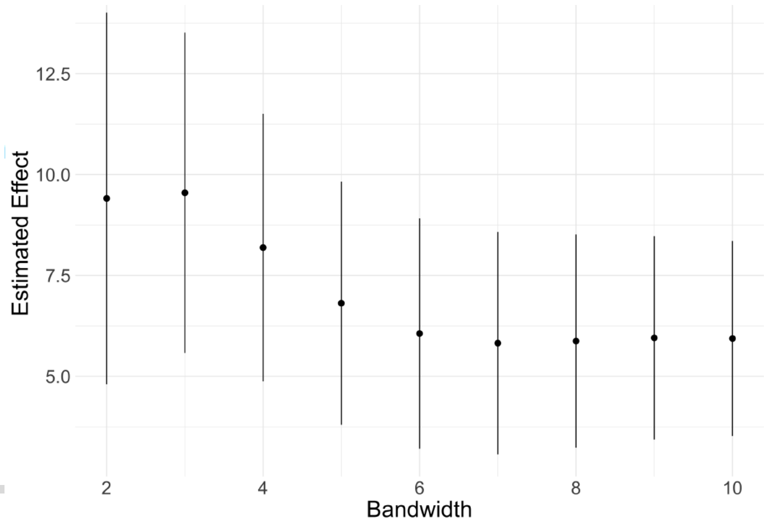
- Intuitive tradeoff: larger bandwidth increases bias but reduces variance



Source: https://conference.nber.org/confer/2021/SI2021/ML/RDD.pdf

Bandwidth choice

- Choice of bandwidth can be very important for size and precision of estimated effect
- As baseline, should go with one of the data-driven choices that have been shown to be "optimal" in some way:
  - — Calonico, Cattaneo, and Titiunik (2014): MSE-optimal
  - — Calonico, Cattaneo, and Farrell (2020): inference-optimal
- Still good idea to show robustness to alternatives – can do graphically:



Inference

- Calonico et al.: MSE-optimal bandwidth choice features bias, which should be incorporated when doing inference

- They advocate "robust bias correction" when reporting confidence intervals (for details, see section 4.3 of https://rdpackages.github.io/references/Cattaneo-Idrobo-Titiunik_2020_CUP.pdf)

- "rdrobust" package (https://rdpackages.github.io/rdrobust/) does this automatically

Estimation of fuzzy RDD

- In a **fuzzy RDD,** not all observations above the threshold are treated and not all below are untreated.

- When $x > x'$ there is just an **increase in the probability** of treatment.

- This means we can use the indicator for being above x' as an instrument for treatment. **Fuzzy RDD is just 2SLS** (the so-called "**Wald estimator**") and gets us

$$\beta = \frac{\lim_{x \downarrow x'} E(y|x) - \lim_{x \uparrow x'} E(y|x)}{\lim_{x \downarrow x'} E(d|x) - \lim_{x \uparrow x'} E(d|x)}.$$

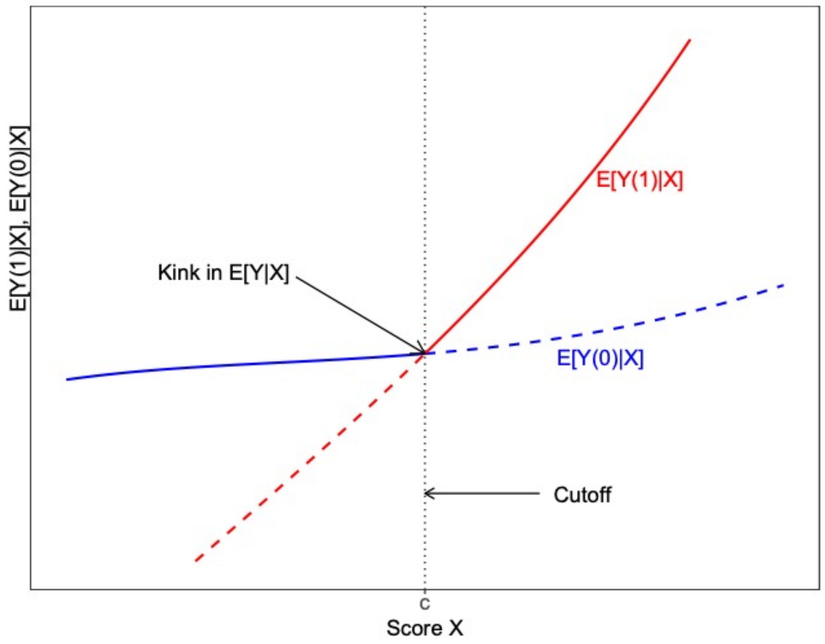
- If treatment effects are heterogeneous, this is the LATE (i.e. the average treatment effect for "compliers")

Estimation of fuzzy RDD

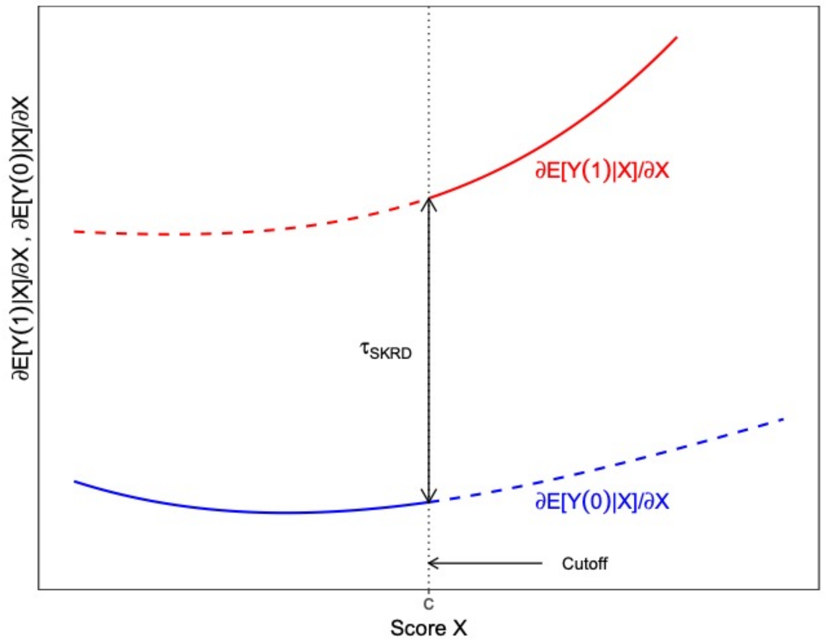
- So effectively, we do two RD estimations as explained earlier
  - the "reduced form": how y varies around the threshold x'
  - the "first stage": how the probability of treatment d varies around the threshold x'
- Intuitively, if jump in the first stage is relatively small, any jump in the reduced form will be "blown up" substantially
  - e.g. if first-stage effect is 0.1, then estimated treatment effect will be 10 times the estimated reduced form effect
  - meaning estimation results may be very sensitive to small changes
  - akin to weak instruments problem in IV
- Implementation: typically use same estimators for numerator and denominator (again can do in 'rdrobust')

Regression kink design

- Sometimes there is not a **jump** at a threshold, but a **kink**
- Generally due to a marginal change in incentives at the threshold – e.g. change in marginal tax rate



(a) Kink RD (levels)



(b) Kink RD (derivatives)

Source: https://conference.nber.org/confer/2021/SI2021/ML/RDD.pdf

Regression kink design

s:fi

- Estimation approaches from RDD generalize quite naturally to this setting

- But tend to require quite a lot of data or very clear kink – intuitively, may be hard to tell apart a kink from a change in slope occurring for other reasons

- Examples in (household) finance:
  - Scharlemann and Shore (RFS 2016)
  - Indarte (JF 2023)

Overview

- Sharp and fuzzy RDD
- Estimating RDD
- **Checking validity**

Internal validity

- As with quasi-natural experiments, there are a couple of tests that one should do to **check for internal validity**.

- Already discussed:
  - Show graphical analysis
  - Use different bandwidths, kernels, maybe polynomials

- There are some other checks one can do that are often even more important, depending on the setting

Internal validity

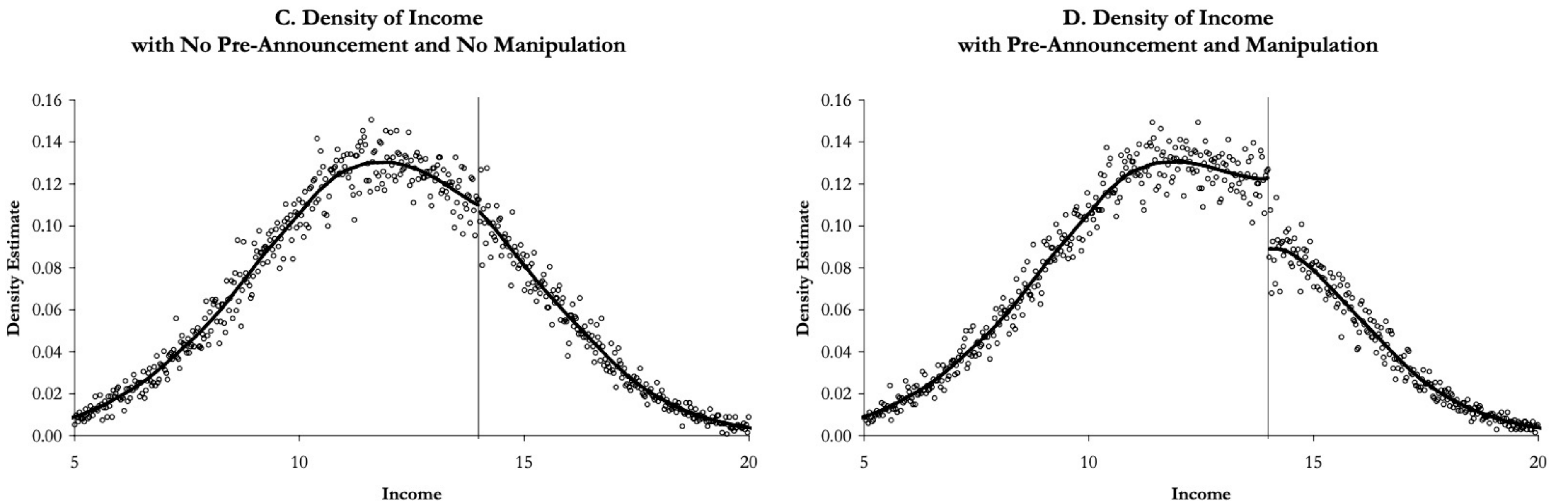
- **Local continuity:** if agents can manipulate the value of the forcing variable, or if administrators can choose the forcing variable or its threshold, then local continuity may be violated
  - With manipulation, we may observe jumps around *x'* absent treatment – always crucial to consider to what extent manipulation is possible

- But manipulation is not *necessarily* invalidating an RDD

- What is crucial is that agents cannot **precisely** manipulate the forcing variable. We will then still have randomness in treatment.
  - e.g. credit scores

Internal validity

- A common test is whether there are any discontinuities in the distribution of the forcing variable at x' -- often called the **McCrary** (2008) test. Illustration (from his paper):



- Can use https://rdpackages.github.io/rddensity/
- There are also approaches that allow for some bunching – see https://francoisgerard.github.io/rdbounds/

Internal validity

- **Balance tests:** Observations close but on different sides of the threshold should have similar potential outcomes.
- Equivalently, these agents should be comparable both in terms of observable and unobservable characteristics
- We can check this for observables – redo graphs using these variables. They should not exhibit a jump at the threshold.
  - Or formal estimation – there should be no statistical/economic significance
  - We cannot do this test for unobservables...
- Related **falsification tests:** Redo estimation around thresholds or for subsamples where you would expect no treatment effect.

Internal validity

- We can also include the other **covariates as control variables**.
- If **local continuity assumption** holds, then including covariates should only influence the precision of the estimates by absorbing residual variance.
- If they strongly affect the estimated treatment effect, we may have “bad controls”, or observations around the cutoff may not be comparable.

Conclusion on RDD

- Intuitive method with many potential applications
- Shortcoming used to be that researchers would make many discretionary choices that could strongly affect results
  - even more so in fuzzy RDD
- Recently, methodological innovations have created principled "default choices" that you should probably use (or, if you choose to not use them, have good explanations why)
- Main limitations: "extremely local" and typically need large sample sizes