Lecture 2

Regression Discontinuity Design: Theory and Applications

1

Introduction

A **Regression Discontinuity Design** (RD) is a powerful and widely applicable identification strategy.

Often access to, or incentives for participation in, a service or program is assigned based on transparent rules with criteria based on **clear cutoff values.**

Comparisons of individuals that are similar but on different sides of the cutoff point can be credible estimates of causal effects for a specific subpopulation.

Good for *internal* validity, not much *external* validity.

2

A little bit of history

- Psychology. The design was invented by Donald T. Campbell in 1958 at Northwestern University. Group of researcher in the Psychology department at Northwestern (In the 80s Trochim – a student of Campbell) explored variants of the design in terms of their validity, implementability and analysis. By 1995, no psychologists no longer work on improving the design, only to popularize its use--at which they had only modest success at best.
- Statistics. In statistics, the design has never had a high profile, perhaps because it is deemed unexciting to be able to model selection when the selection process is completely known, perhaps because the design leads to causal inferences whose generalization is limited to the cutoff point, and perhaps because its interpretation depends on functional forms that cannot be directly observed. Very empirical tool.

3

A little bit of history

- Economics. The first formal proof of unbiased causal inference resulting from the design came from unpublished papers in economics--by Goldberger in 1972. However, economists were then pursuing a broader causal agenda and interest in regression-discontinuity lapsed. It was not used in economic applications until the mid-1990s. However, it has widely caught on since them among labor economists and econometricians in both the United States and Europe.
- Finance. Earlier working paper 2003-2004, all from students coming out from either MIT or Berkeley (strongest labor economics training).
- Has life now arrived for this 50-year-old design, invented in 1958 and rarely used until the beginning of this century?

4

Introduction

Intuition:

- Participants are assigned to either the program (treatment) X_i=1or to the control group (X_i=0) on the basis of a cutoff score based on a pre-program measure, Z_i

  E.g.: giving students financial aid on the basis of their GPA grades or past achievements

Objective:

- establish whether the outcome $Y_i(1)$-$Y_i(0)$ is driven by the treatment.

Causality:

- If one could run this as a randomized experiment one would be able to establish causality, but because the treatment is administered as a function of some covariates (Z_i), establishing causality is harder

Advantage:

- The advantage over randomized experiments: one assigns treatment to those who most need or deserve it

5

Intuition

Problems in identification of causality:

- Since treatment is assigned as a **deterministic function** of an observed covariate that is also related to the outcome of interest, identifying causality poses a problem.

Solution: exploit discontinuities!

- If the assignment mechanism is discontinuous, then one can control for any smooth function of covariates and still estimate the effect of the award at the point of discontinuity.

6

Intuition

How does it work?

- The probability of receiving treatment changes **discontinuously** as a function of one or more underlying variables: the pre-program measure.

- The pre-program measure may itself be associated with the potential outcomes, but this association is assumed to be **smooth**.

→ Then, any discontinuity in the conditional distribution of the outcome, indexed by the value of this covariate at the cutoff value, is interpreted as evidence of a causal effect of the treatment.

7

External validity concerns

- A study that readily allows its findings to generalize to the population at large has high external validity.
- Because of the assumption that only at the cutoff point the two groups are equivalent, RD designs are not as statistically powerful as the Randomized Experiments.

- That is, in order to achieve the same level of statistical accuracy, and RD Design needs as much as 2.75 times the participants as a randomized experiment.

→ The attractiveness of RD Designs is ethical: one can assign treatment to the most needed.

→ And practical: often you are not the person administering the treatment (finance).

8

Examples of applications

- Campbell (1969) studied the effect of National Merit scholarships on applicants' later achievement when the scholarships were awarded on the basis of past achievement.
  - Identification of causality was done by matching discontinuities or nonlinearities in the relationship between outcomes and past achievement to discontinuities or nonlinearities in the relationship between awards and past achievement.

- Thistlehwaite and Campbell (1960) study the effects of student scholarships on career aspirations using the fact that awards are only made if a test score exceeds a threshold

9

Examples

- Van der Klaauw (1997) studies the effect of financial aid offers on students' decisions to attend a college, taking into account administration rules that set the aid amount on the basis of a discontinuous function of the students' grades

- Angrist and Lavy (1999) study the effects of class size on student test scores using the fact that at a certain number of students (40, by Maimonides rule), a class is divided into two classes.

- Remember Rauh (2006) from last lecture. The spirit of his identification is very close to RDD, but he never used a RDD. Keep it in mind for finance applications (later). We are looking for an exogenous rule. We will look at specific examples.

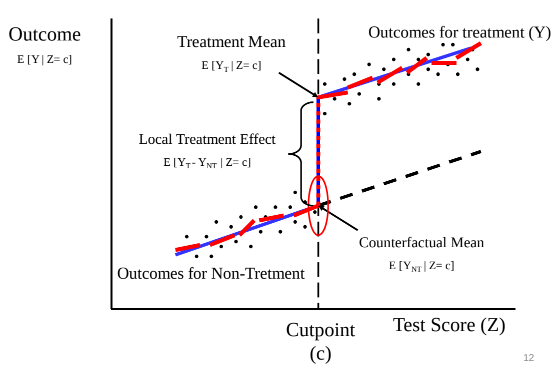
Basics

- Two potential outcomes $Y_i(0)$ and $Y_i(1)$, causal effect $Y_i(1)-Y_i(0)$, binary treatment indicator
  - $X_i=0$
  - $X_i=1$
- Covariate $Z_{i,}$ and the observed outcome equal to:

$$Y_i = Y_i(X_i) = \begin{cases} Y_i(0) & \text{if } X_i = 0 \\ Y_i(1) & \text{if } X_i = 1 \end{cases}$$

- At $Z_{i,}$ =c participation changes
- Because the treatment is determined either completely or partly by $Z_{i,}$, $Z_{i,}$ maybe associated with the potential outcome but the association is assumed to be smooth→ any discontinuity in the conditional distribution of the outcome $Y_i(1)$ is interpreted as evidence of causality.

The RD Strategy



Two cases

- Sharp Regression Discontinuity

$$X_i = 1\{Z_i \geq c\}$$

  – The sharp RD design applies when assignment of treatment is deterministic: depends on a known and specified pre-program measure and the probability of receiving treatment changes from 0 to 1 at the threshold

- Fuzzy Regression Discontinuity Design:

$$\lim_{z \downarrow c} \Pr(X_i = 1 | Z_i = z) \neq \lim_{z \uparrow c} \Pr(X_i = 1 | Z_i = z)$$

  – The fuzzy RD design applies when assignment of treatment is not deterministic, but a function of a pre-program measure with other variables that are unobserved by the econometrician. Thus, the probability of receiving treatment does not change from 0 to 1 at the threshold.

13

Sharp Regression Discontinuity

- Example: Berry and Lee (2006)
  – Does the community reinvestment act→ improve credit for minority and Low and Medium Income borrowers?
  – The regulation uses a strict cut-off of 80 percent of metropolitan area (MSA) median income to identify LMI neighborhoods and individuals, loans to which count toward a lending institution's CRA performance rating.
  – Using this RD design and a data set comprising millions of mortgage applications from 1993 to 2003, we are able to distinguish the causal effects of CRA from other (observable and unobservable) variables correlated with residential lending rates.
  – Vanishingly small effects of CRA on the reduction of loan rejection rate for minority and low-medium income.

14

Sharp Regression Discontinuity

$$X_i = 1\{Z_i \geq c\}$$

- All units with covariate value $Z_{i,}$ of at least c are in the treatment group, all the ones with value less than c are in the control group
- The discontinuity in the conditional expectations of the outcome given $Z_{i,}$ is interpreted as the causal effect of the treatment

$$\lim_{z \downarrow c} E[Y_i(1) | Z_i = z] - \lim_{z \uparrow c} E[Y(0)_i | Z_i = z] = \lim_{z \downarrow c} E[Y_i | Z_i = z] - \lim_{z \uparrow c} E[Y_i | Z_i = z]$$

Assuming continuity

15

Sharp Regression Discontinuity

- The smoothness assumption is that the conditional distribution function is smooth in $Z_{i,}$, the covariate.
- You can interpret the outcome difference at the discontinuity point as the average causal effect :

$$\tau_{SRD} = \lim_{z \downarrow c} \mathrm{E}[Y_i | Z_i = z] - \lim_{z \uparrow c} \mathrm{E}[Y_i | Z_i = z] = \mathrm{E}[Y_i(1) - Y_i(0)]$$

- Extrapolation is necessary because by design there is no units with $Z_{i,}$ =c for which we observe $Y_i(0)$

16

Fuzzy Regression Discontinuity

- Example: Vanderklaauw (2007)
  - What is the effect of financial aid on→ acceptance of college admission?
  - Z is a numerical score assigned to college applicants based on characteristics (SAT scores, grades, etc..)
  - During the initial stage of the admission process the applicants are divided in groups
  - For example, (assuming only two groups – financial aid and not):

$$G_i = \begin{cases} 1 & if\ 0 \le Z_i \prec c \\ 2 & if \quad Z_i \ge c \end{cases}$$

Financial aid

17

Fuzzy Regression Discontinuity

- Example: Vanderklaauw (2007)
- The outcome of interest is college attendance, but the statistical association between attendance and financial aid is ambiguous:
  - An aid offer makes the college more attractive (this is the causal effect of interest!)
  - A student who gets a generous aid is likely to have better outside opportunities from other colleges

18

Fuzzy Regression Discontinuity

$$\lim_{z \downarrow c} \Pr(X_i = 1 | Z_i = z) \neq \lim_{z \uparrow c} \Pr(X_i = 1 | Z_i = z)$$

- The design allows for a smaller jump in the probability of assignment to the treatment at the threshold without requiring the jump to be equal to 1.
- In this design we interpret the ratio of the jump in the regression of the outcome on the covariate to the jump in the regression of the treatment indicator on the covariate as an average causal effect
- Formally

$$\tau_{FRD} = \frac{\lim_{z \downarrow c} \mathrm{E}[Y_i | Z_i = z] - \lim_{z \uparrow c} \mathrm{E}[Y_i | Z_i = z]}{\lim_{z \downarrow c} \mathrm{E}[X_i | Z_i = z] - \lim_{z \uparrow c} \mathrm{E}[X_i | Z_i = z]}$$

19

SDR vs FDR

- Terminology is a bit misleading, as both designs require a sharp jump in the assignment of the treatment.
- Deterministic vs probabilistic assignment is a better terminology
- The special case where we know the conditional mean of treatment above and below the cutoff, as with a binary treatment is the deterministic RD design since there is "deterministic assignment" of treatment conditional on the observed assignment variable. In any other case, we have to estimate the jump in the conditional mean of treatment at the discontinuity.

20

Recap: assumptions of RDD

- The first assumption is that treatment is not randomly assigned, but is assigned based at least in part on a variable we can observe. I called this variable the covariate variable, or Z, but it is often called the assignment variable, or "*running*" or "*forcing*" variable.
- The crucial second assumption is that there is a discontinuity at some cutoff value of the assignment variable in the level of treatment. For example, in the financial aid example, students above a certain cutoff are eligible, below they are not.

21

Recap: assumptions of RDD

- The <u>third</u> crucial assumption is that individuals cannot manipulate the assignment variable (e.g. by changing Z) to affect whether or not they fall on one side of the cutoff or the other, or more strongly, observations on one side or the other are exchangeable or otherwise identical.

22

Recap: assumptions of RDD

- The <u>fourth</u> crucial assumption is that the other variables are smooth functions of the assignment variable (z) conditional on treatment (x), i.e. the only reason the outcome variable (y) should jump at the cutoff is due to the discontinuity in the level of treatment.
- The basic idea of this very important assumption is that people close to the discontinuity have very similar values of z but very different values of x. Any discontinuity in the outcome is ascribed to the discontinuity in x.
- This implicitly assumes that the outcome is continuous in z for both treatment and control groups.
- If this is not the case then the difference in outcomes at $Z= c$ could be the result of a discontinuity in Z not the discontinuity in x.
- Note this differs from IV, in that the assignment variable Z can have a direct impact on the outcome Y , not just on the treatment X, though not a discontinuous impact.

23

Applications in Economics

24

Butler, Lee and Moretti (2004)

- Two alternative hypothesis:
  - Competition for votes induces politicians to move toward the center. In this view, elections have the effect of bringing about some degree of policy compromise.
  - Voters merely *elect* policies: politicians cannot make credible promises to moderate their policies and elections are merely a means to decide which one of two opposing policy views will be implemented.

25

Butler, Lee and Moretti (2004)

- SRD design
- The effect of party affiliation of a congressman on congressional voting outcomes. The key idea is that electoral districts where the share of the vote for a Democrat in a particular election was just under 50% are on average similar in many relevant respects to districts where the share of the Democratic vote was just over 50%, but the small difference in votes leads to an immediate and big difference in the party affiliation of the elected representative.
- In this case, the party affiliation always jumps at 50%, making this an SRD design.
- They looks at the effect of being a winner to policies.
- Why this paper? Very well executed! Use it as a model!

26

I. RD design: Finance Applications

- Three applications
  - Rauh (2006)
  - Chava and Roberts (2007)
  - Black, Jang and Kim (2006)

27

Rauh (2006)

Joshua D. Rauh (2006). “Investment and Financing Constraints: Evidence from the funding of Corporate Pension Plans” (continuation)

Recall that the objective of this paper was to exploit nonlinear funding rules of defined benefit pension plans in order to identify the dependence of corporate investment on internal financial resources.

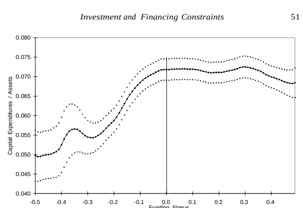
Not a RDD but similar in spirit: Identification comes from the function that relates the pension funding status to investment opportunities;

28

Rauh (2006)

Examples of non-linearities:

- investment opportunities have no reason to make a discrete jump at the level of full pension funding, whereas required pension funding jumps to 0 in that case.
- Investment opportunities have no reason to be correlated with pension funding for underfunded plans but not for overfunded plans.



29

Rauh (2006)

Data and empirical strategy:

Data: large unbalanced panel of Compustat firms that filed an IRS 5500 (which contains information on the funding status) for the period 1990 – 1998.

Model:

$$I_{it} / A_{i,t-1} = \alpha_i + \alpha_t + \beta_1 Q_{i,t-1} + \beta_2 \frac{NonPensionCashFlow_{it}}{A_{i,t-1}} + \beta_3 \frac{Z_{it}}{A_{i,t-1}} + x_{it}\gamma + \varepsilon_{it}$$

where

- Z(i,t) is the mandatory contribution and β_3 is the coefficient of interest,
- X(i,t) is a vector of controls, including in some specifications the funding status itself,
- firm and year fixed effects absorb certain sources of variation in required contributions that may be undesirable because of their correlation with investment opportunities.

30

Rauh (2006)

Results:

- When mandatory contributions alone are considered, they have a negative and significant effect on capital expenditures
- When including the funding status as a control, MC's have a slightly lower estimated effect on investment but it is still negative and significant.
- Underfunding has a negative and significant effect on investment, but overfunding does not have any effect. This suggests an effect of the cash drain from required contributions on capital investment
- When underfunding, overfunding and nonlinearity is controlled for, a significant and negative effect of required contributions on investment is still measured.

31

Rauh (2006)

Results:
When dividing the sample by different characteristics (that are proxies for financing constraints) to study the effects of mandatory contributions for capital expenditures of firms, the results are that the strength of the (negative) relationship between the two is:

- Inversely related to the quality of a firm's credit rating
- Strongest for the youngest and middle-aged firms
- Strongest for firms with low dividends-to-assets ratios
- Strongest for firms with relatively little cash-to-assets ratios

Conclusion:
Financing constraints affect investment.

32

Criticisms

- The Bakke- Whited paper mis-characterizes the paper as an RD paper and then debunks that mis-characterization. The actual discontinuity at 0% is indeed very small, so an RDD design would be difficult to claim. Does the identification, still holds? Figure 5 is certainly intriguing. So, what is the economic argument against the identification?
- Not surprising that over the range of very small shocks (0.0001% of assets what is used) firms absorb that with net working capital. Of course, very small changes are probably absorbed by changes in NWC. Figure 5 is quite flat around 0%.
- The Bakke Whited paper makes the charge that results are driven by the 15% most underfunded firms. Take ANY empirical paper, remove the 15% of the results that are treated by the effect, and find the result remains. 1 or 2% would be reasonable to challenge.
- "The effect is too large." The effects are measured with large standard errors, so the 95% confidence interval is like [e>0,1]. There is a nonzero effect, and that it is bigger for more constrained firms. Interesting objection, but not very substantive.

33

Chava and Roberts (2007)

How does financing impact investment? The role of debt covenant violations. (Chava and Roberts 2007)

This paper examines the impact of debt covenants (financing frictions) on corporate investment.

→ Upon violation of debt covenants control rights shift to the creditors who can use the threat of accelerating the loan to intervene in management.

→ The discrete nature of the covenant violation generates a potentially exogenous source of variation in the distance to the covenant threshold that can be used for the empirical strategy of the Regression Discontinuity design:

  → The borrower retains control rights as long as net worth is above a certain threshold and looses controls rights as soon as net worth falls below the threshold, regardless of the amount of net worth.

34

Chava and Roberts (2007)

Data:

Bank and non-bank loans to companies from the period 1994-2005. Quarterly data.

Empirical model:

Treatment is assigned as follows:

$$Bind_{it} = \begin{cases} 1 & if \quad z_{it} - z_{it}^0 > 0 \\ 0 & otherwise \end{cases}$$

With *i* the firm and *t* the year-quarter, z_{it} the observed current ratio (or net worth), and z_{it}^0 the threshold specified by the covenant.

The estimated equation is:

$$Investment_{it} = \alpha_0 + \beta_0 Bind_{it-1} + \beta_1 X_{it-1} + \eta_i + \upsilon_t + \varepsilon_{it}$$

β_0 is the parameter of interest and X_{it-1} a vector of covariates.

η_i and υ_t control for firm and year-quarter fixed effects.

35

Chava and Roberts (2007)

The RD design is enabled by the fact that the function mapping the distance between the underlying accounting variable and the covenant threshold into the treatment effect is discontinuous:

→ Even if ε_{it} is correlated with $z_{it} - z_{it}^0$, the estimate β_0 is unbiased as long as ε_{it} does not exhibit precisely the same discontinuity as $Bind_{it}$.

Results:

- Covenant violations are associated with a decline in investment of the order of 1.5% of capital per quarter. This translates into a relative decline of 13%
- Other specifications that include Tobin's q, firm size, and controls like debt overhang do not change results qualitatively.
- Their results highlight how state contingent allocation of control rights potentially mitigates investment distortions arising from financing frictions.

Problems??

36

Black, Jang, and Kim (2006)

- A corporate governance application
- Natural candidate, as corporate governance rules are often applied to some firms and not other depending on a cutoff (generally size).
- Question: does corporate governance rule increase market value?
- Problem: application of corporate governance rules may be endogenous and the effect on market value may be driven by selection

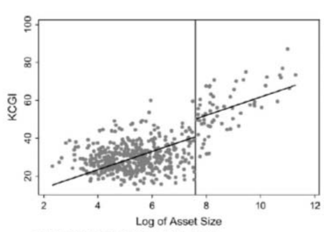
37

Black, Jang, and Kim (2006)

- Strategy:
  - Exploit regression discontinuity design to identify whether the effect is causal.
  - In Korea, only firms above 2 trillions have to follow some rules.
  - They use a smooth parametric form to capture the direct effect of firm size on Tobin's q. This lets the asset size dummy capture the discontinuous effect of size on governance at 2 trillion won. They use ln(assets) as a simple parametric form for firm size, which hopefully captures the direct connection between firm size and Tobin's q.

38

Black, Jang, and Kim (2006)



39

Black, Jang, and Kim (2006)

- They use the discontinuity as an instrument as in Angrist and Lavy (1999). IV estimates use discontinuities or nonlinearities in the relationship between corporate governance and size.

Diagnostic when using RDD

Graphical analysis – plot 1

- The first plot is a histogram-type estimate of the average value of the outcome Y by the forcing variable Z
- The question is whether around the threshold *c* there is any evidence of a jump in the conditional mean of the outcome
- Also, you want to make sure that there are no other jumps in the conditional expectations of Yi given Z_i that are comparable to the discontinuity at the cutoff value. If otherwise, the interpretation of the jump at cutoff value as casual effect would be questionable.

Plot 1

- Choose a binwidth and a number K of bins to the left & right of the cutoff
- Calculate average outcome in each bin
- Plot it against midpoint of the bins
- You want to see a jump in the conditional mean of the outcome around cutoff
- Inspect if there are other jumps comparable to the discontinuity at the cutoff (suspicious!)

43

Butler, Lee and Moretti (2004)

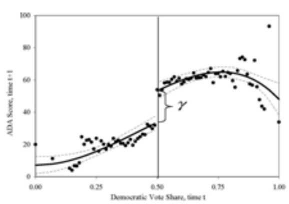


Figure I

Total Effect of Initial Win on Future ADA scores: γ

Note: This figure plots ADA scores after the election at time t+1 against the Democrat vote share, time t. Each circle is the average ADA score within 0.01 intervals of the Democrat vote share. Solid lines are fitted values from 4th order polynomial regressions on either side of the discontinuity. Dotted lines are pointwise 95% confidence intervals. The discontinuity gap estimates:

$$\gamma = \underbrace{\pi_0\left(P_{t+1}^{*D} - P_{t+1}^{*R}\right)}_{\text{"Affect"}} + \underbrace{\pi_1\left(P_{t+1}^{D} - P_{t+1}^{R}\right)}_{\text{"Elect"}}$$

44

Graphical analysis – plot 2

- The second plot is a histogram-type estimate of the average value of the outcome Y by the other covariates W_i
- The objective of this plot is to see graphically that the jump is NOT due to other covariates rather than the forcing variables.
- Choose other baseline covariates
- Do a similar plot to the one you did for Y
- You do *not* want to see a jump around cutoff

45

Graphical analysis – plot 2

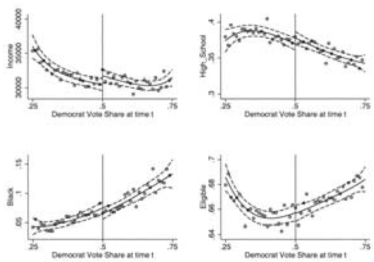


FIGURE III
Similarity of Constituents' Characteristics in Bare Democrat and Republican Districts–Part 1

Graphical analysis – plot 3

- The third plot (for FDR) is to plot the mean value of the treatment X_i to make sure that there is a jump in the probability of treatment at the cutoff point.
- (for example, in the Keys, et al.. paper showing that when the FICO score is higher than a certain threshold higher probability of treatment (securitization) – equivalent to plot 1 for SDR)

Graphical analysis –plot 4

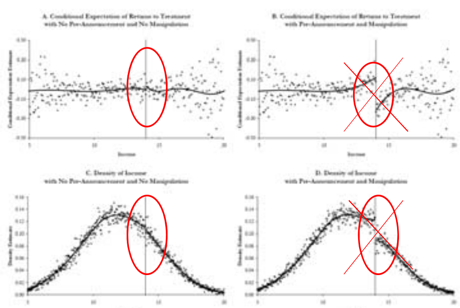
- The fourth plot should inspect if there is any discontinuity in the distribution of the forcing variable Z_i at the threshold
- If individual can manipulate the test scores by retaking the test, then individuals just below the test score may do so and invalidated the design
- → discontinuity of the conditional density of the test score at the threshold.
- It turns out that if the running variable is not entirely under agent's control (partial manipulation), identication can still be achieved (Lee 2007).

Plot 4

- Plot the # obs in each bin against the mid points of the bins
- This is done to see if the distribution of forcing variable is smooth around the cutoff
- You do *not* want to see a discontinuity in the density around cutoff, as it would be a sign of manipulation by individuals
- Example: McCrary (2008), fig. 2

Plot 4

Figure 2. Hypothetical Example:
Gaming the System with an Income-Tested Job Training Program



McCrary: Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test

Regression analysis

- Local regression analysis
- There is a great deal of art involved in the choice of some continuous function of the assignment variable Z for treatment and outcomes.
- You choose some high-order polynomial of Z to estimate separately on both sides of the discontinuity, or better, a local polynomial, local linear, or local mean smoother, where the art is in the choice of kernel and bandwidth (your smoothing parameter).
- Stata now offers lpoly which supports aweights; users of prior versions can findit locpoly and expand their data to get weighted local polynomial estimates. The default in both is local mean smoothing, which is fine, but local linear regression is usually preferred.

Choose of the bandwidth

- There are several rule-of-thumb bandwidth choosers and cross-validation techniques for automating bandwidth choice, but none is foolproof. McCrary (2007) contains a useful discussion of bandwidth choice, and claims that there is no substitute for visual inspection comparing the local polynomial smoother with the pattern in the scatterplot.
- Because different bandwidth choices can produce different estimates, the researcher should really report more than one estimate, or perhaps at least three: the preferred bandwidth estimate, and estimates using twice and half the preferred bandwidth.
- As it is, though, local polynomial regression is estimating hundreds or thousands or even hundreds of thousands of regressions. Bootstrapping these estimates requires estimating millions of regressions, or more. Still, computing time is cheap now.

52

More choices

Of somewhat less importance, but equally art over science, is the choice of kernel. Most researchers use the default Epanechnikov kernel, but the triangle kernel typically has better properties at boundaries, and it is the estimates at the boundaries that matter in this case.

- Show the Data
  - Given how much choice the researcher has over parameters in a supposedly nonparametric strategy, it is always wise to show a scatter or dotplot of the data with the local polynomial smooth superimposed, so readers may be reassured no ad hoc of picking parameters were involved.

53

Testing for the existence of a discontinuity

- The first test should be a test that the hypothesized cutoff in the assignment variable produces a jump in the level of treatment. In the class size example, this is easy: the probability of having a smaller class jumps when in number of enrolled kids in the grade reach the cutoff: in the education example, the discontinuity is far from obvious. (parallel to plot 1 and 3).
- The test for a discontinuity in treatment X is the same as a test for a discontinuity in the outcome Y . Simply estimate a local linear regression of X on Z, both above and below the cutoff, perhaps using a triangle kernel with a bandwidth that guarantees 10-20 observations are given positive weight at the boundary, approaching the cutoff from either side. This test to rule out manipulation (similar to plot 4)
- The local estimate at the cutoff for regressions constrained below the cutoff is $x-$ and the local estimate at the cutoff for regressions constrained above the cutoff is $x+$. The computation of the difference $x+ - x-$ can be wrapped in a program and bootstrapped for a test of discontinuity.

54

Testing for sorting at discontinuity

- McCrary (2007) gives a very detailed exposition of how one should test this assumption by testing the continuity of the density of the assignment variable at the cutoff. As he points out, the continuity of the density of the assignment variable is neither necessary nor sufficient for exchangeability, but it is reassuring.
- McCrary (2007) also provides tests of sorting around the discontinuity in voting in US Congressional elections, where there is no sorting, and in roll-call votes in Congress, where there is sorting.

55

Testing for Extraneous Discontinuities in Y and X

- Placebo test: no extra jumps in the levels of treatment or the outcome where no hypothesized cutoff exists.
- This is a test of the fourth crucial assumption. One can easily pick 100 random placebo cutoff points, and test the difference in X and the difference in Y (about 5 placebo cutoffs will show significant jumps, of course).

56

Conclusions on RDD

- Very simple tool
- The challenge is to understand how to apply to your question
- We have seen very different applications that combine RDD with IV, with structural estimation, with theoretical models that guide the identification strategy…

57