Lecture 17

Regression Discontinuity Design (RDD) II

Regression Discontinuity

Outline

- RD (review)
- RD implementation
- Fuzzy RD

RD identification assumption

- The purpose of using RD is to mimic randomization
- The identification assumption is that around the cutoff, treatment is randomly assigned
- In other words, around the cutoff, we assume treatment is not tainted by omitted variables; individuals just to the left and right are like “apples to apples”

Sharp vs. fuzzy RD

Two types of RD:

- Sharp RD:
  - We know the exact relationship between running variable and treatment status-treatment status is completely determined by the running variable
- Fuzzy RD:
  - There is not an exact relationship- treatment status is partially determined by the running variable
  - The discontinuity strongly influences treatment status, but there is “slippage” or “non-compliance”
  - In this case, we use the discontinuity as an **instrument** for treatment status

Example: Austrian unemployment benefit- Lalive 2008

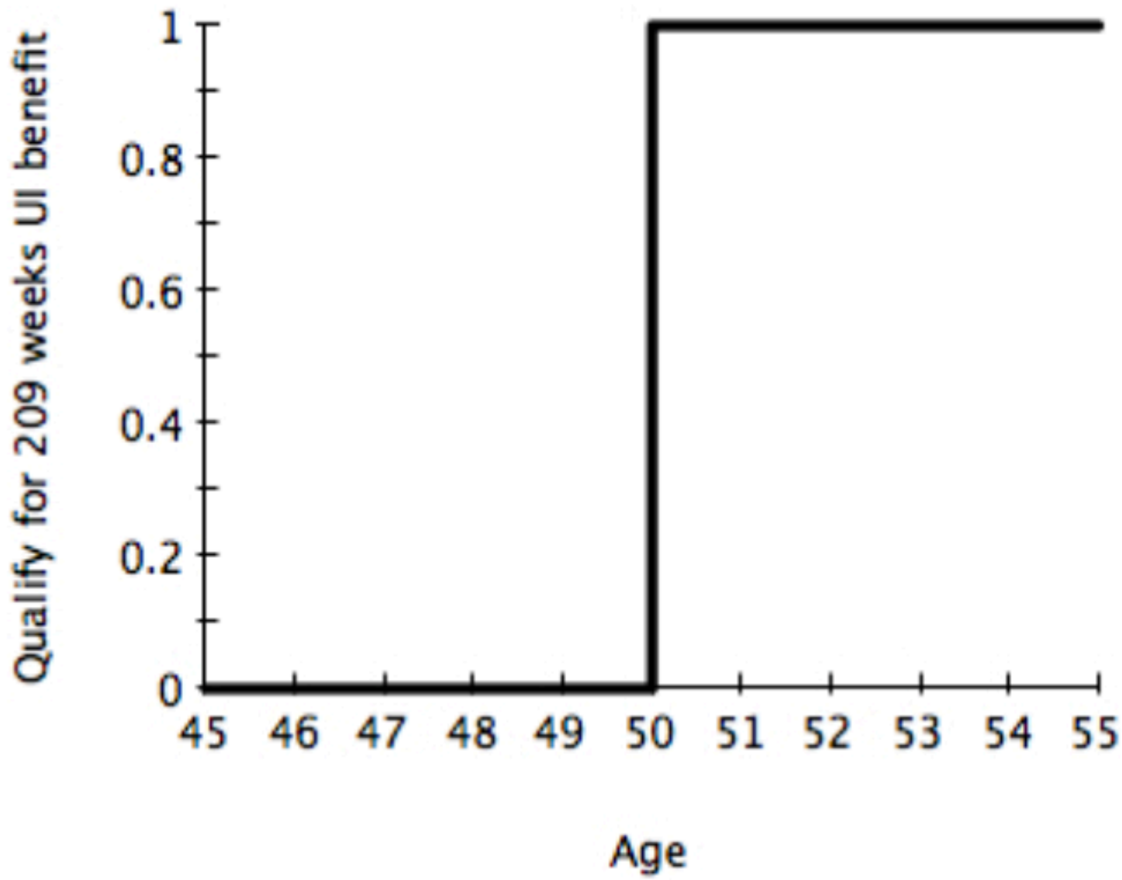
- Austrian government in 1988 extended unemployment insurance (UI) benefits for people over 50 living in certain regions of Austria from 30 weeks to 209 weeks.
- Causal question: Does the availability of 209 weeks of benefits increase unemployment duration?
- Question: Can you just compare people over 50 before and after the policy change?
- What can be some omitted variables?

Sharp RD

- Workers who are 49 only qualify for 30 weeks of unemployment insurance. Workers who are 50 qualify for 209 weeks unemployment insurance
- What is the running variable and cutoff?
- What is the mean comparison?
- Access identification assumption:
  - Are 49 years old and 50 years old different in any observed or unobserved way?
  - What are some stories that can violate the identification assumption?
- Since age perfectly predicts the maximum number of weeks unemployment, this is a sharp RD
- The sharpness can be seen in a picture

Sharp RD -“first stage”

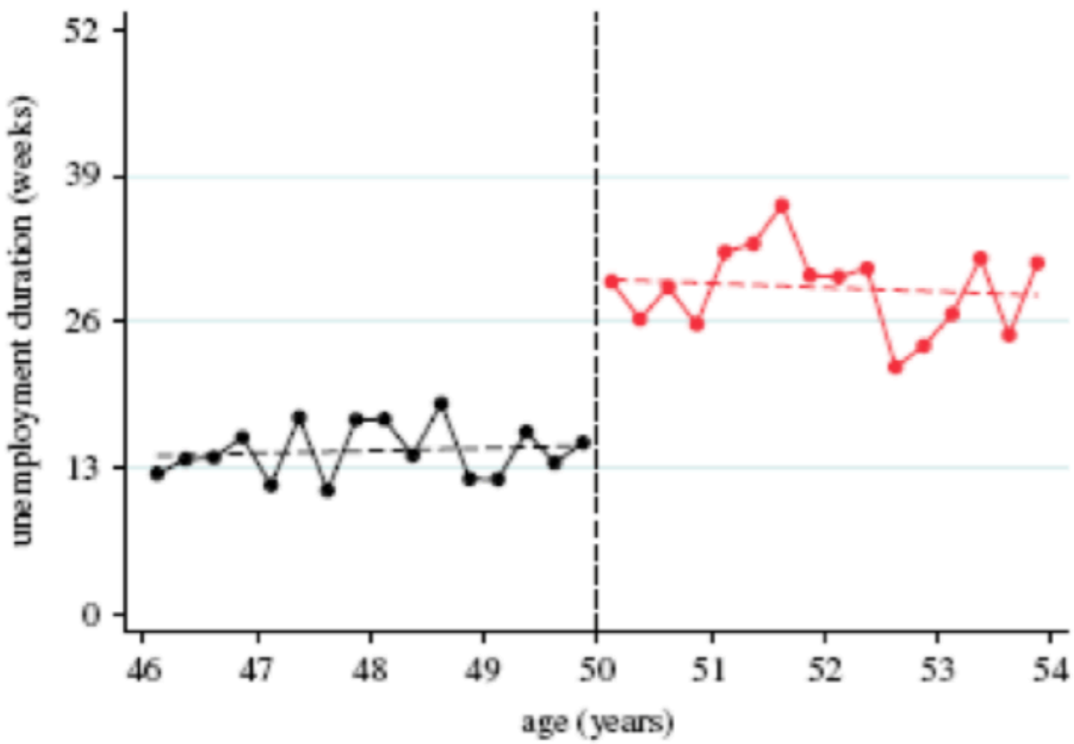
In Sharp RD, “first stage” usually is not shown



Sharp RD-"reduced form"

- Comparing unemployment duration for people aged 49 vs 50 tells us the impact on duration when age switched from 49 to 50
- It is also the impact of unemployment benefits on duration-our causal question.
- This is because we have "perfect compliance"
- All we need to assume is that 49 and 50 year olds are otherwise similar-will discuss how to explicitly test this identification assumption in the next lecture
- Results: average duration for 49 year olds was 13 weeks. For 50 year olds, average duration was 27 weeks.
- To make this even more convincing, we can show a picture

Sharp RD-”reduced form”



Discontinuity at threshold = 14.798; with std. err. = 1.928.

- Compelling evidence that the policy impacted people's unemployment duration
- The RD estimate is the vertical distance between the outcomes right at the cutoff

Linear estimation-separate left and right limit

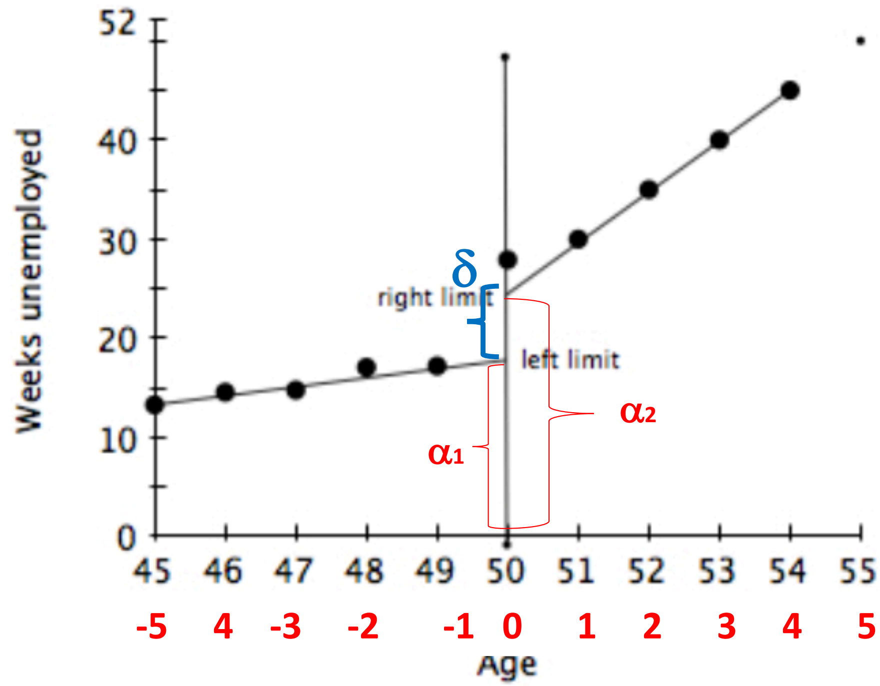
- Estimate the left and right limit by linearly extrapolating the left and right trend separately

Step1: subtract the cutoff score from running variable to center it at zero
Define S=x-c (running var.-cutoff)

Step2: $Y=\alpha_2+\gamma_2 S+\varepsilon$ for $S\geq 0$

Step3: $Y=\alpha_1+\gamma_1 S+\varepsilon$ for $S\leq 0$

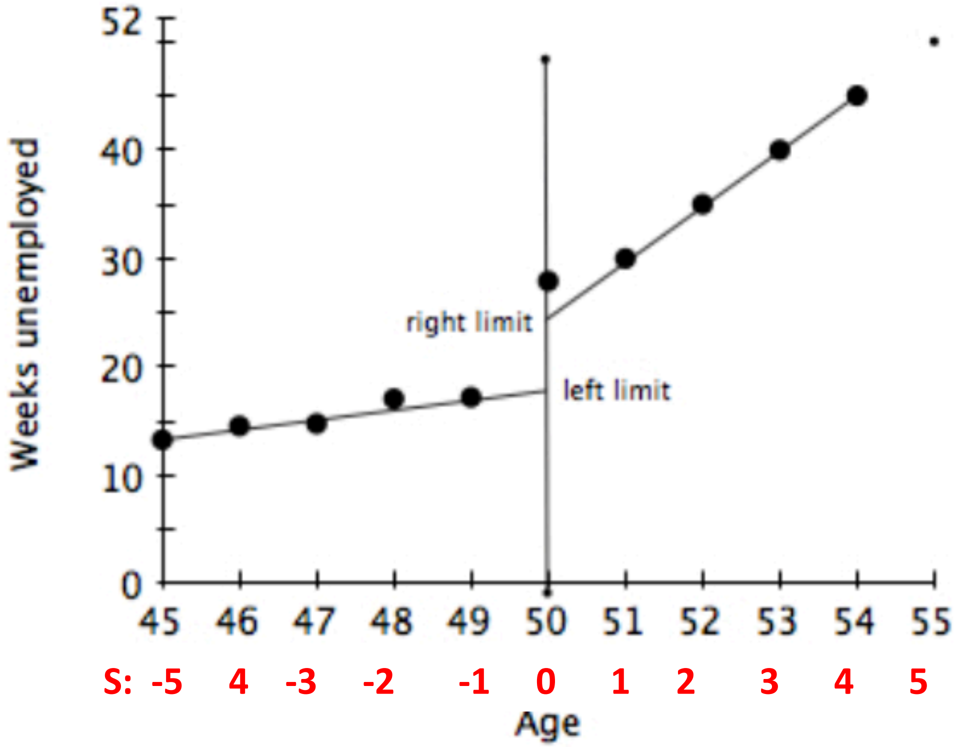
Step4:causal impact $\delta=\alpha_2-\alpha_1$



- What is the drawback of calculating δ manually?

Linear estimation-combine left and right limit

- Step 1: Create an indicator for Above=1 if S $\geq$0, Above=0 if S<0.
- Step 2: $Y = \alpha + \delta Above + \gamma S + \eta(S * Above) + \varepsilon$



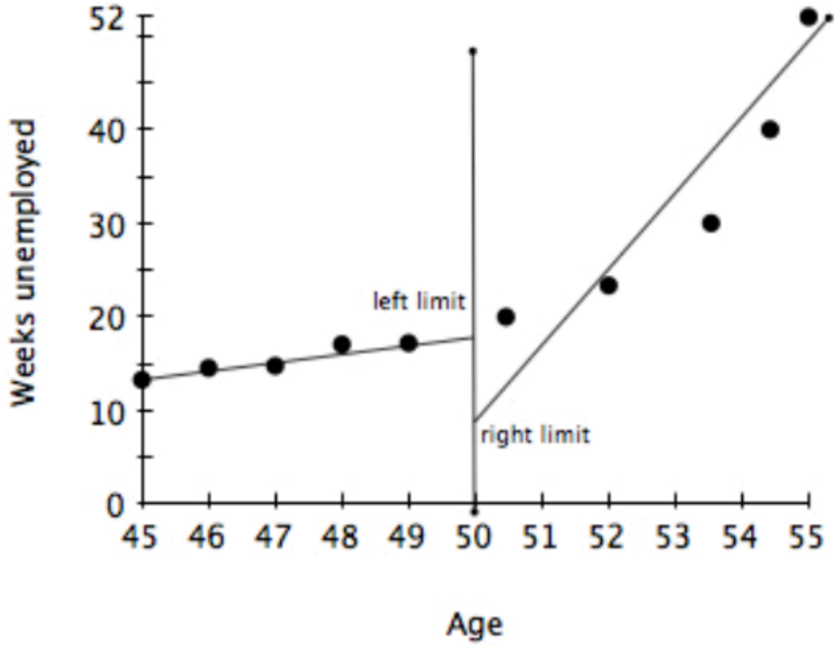
- When S is just below zero (-0.001), we have $\hat{Y} \approx \alpha$
- When S is just above zero(0.001), we have $\hat{Y} \approx \alpha + \delta$
- The estimate of the discontinuity is simply δ
- The regression will provide standard error as well

Advanced:

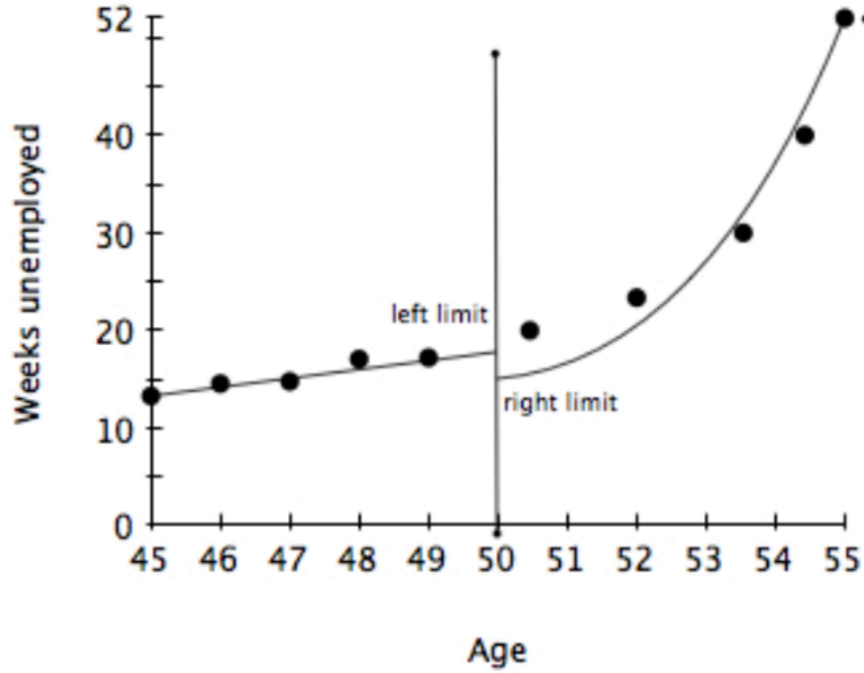
- What does each parameter mean in the regression?
- α: left limit (intercept of left regression)
- δ: treatment effect ($\alpha + \delta$ is the intercept of right reg)
- γ: slope of left regression
- η: additional slope of right regression

What if linearity is not a good assumption?

Wrong model: linear



Right model: quadratic



Fuzzy RD example: Health care at birth and academic achievement

- Research Question: does health care at birth increase one's academic achievement?
- Why is OLS regression biased?
- Bhatawagi (2013 AER) notes that doctors use a rule of thumb to decide whether to provide certain infant healthcare

1. Infants born below 1500g should get intervention
2. The actual decision is not entirely based on this rule, doctors also use their own judgments

- What is the running variable?
- Why is it a Fuzzy RD?

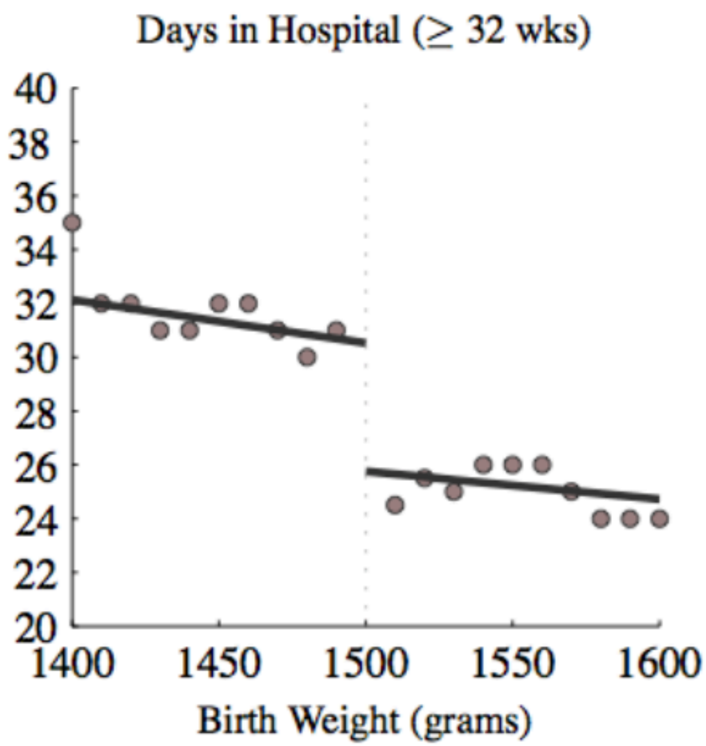
Fuzzy RD Implementation

- A fuzzy RD is just like random assignment with imperfect compliance.
- There two stages:
  - First stage: RD analysis of how running variable impacts D (treatment variable)
  - D$= \alpha + \delta_1 Above + \gamma_1 S + \eta_1 (S * Above) + \varepsilon_1$
  - Reduced form: RD analysis of how running variable impacts Y (outcome variable)
  - $Y = \alpha + \delta Above + \gamma S + \eta (S * Above) + \varepsilon$

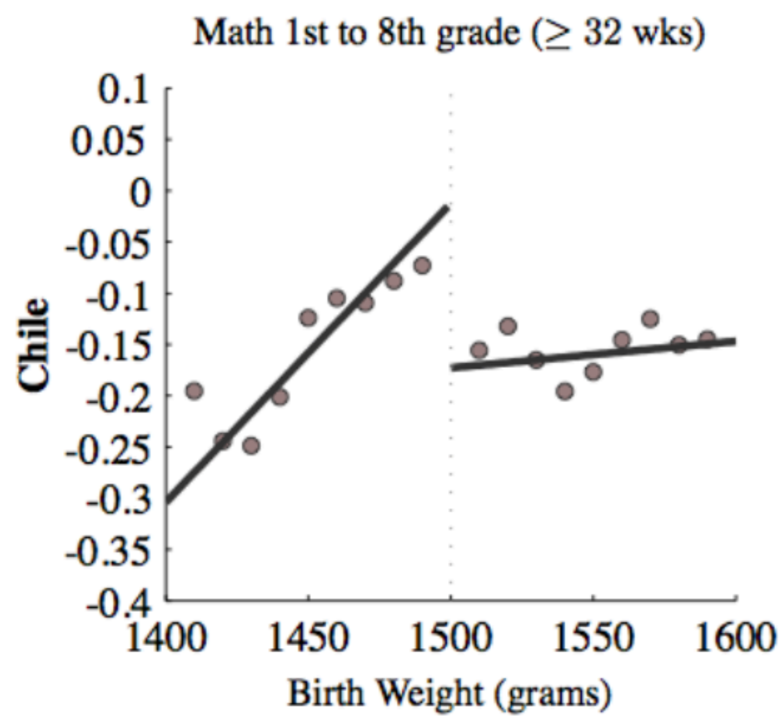
- Treatment effect (RD estimate/LATE) is $\frac{Reduced\ form}{First\ stage} = \frac{\delta}{\delta_1}$
- X is the running variable, s is the re-centered running variable

The best way to show RD estimate-graphs

First Stage



Reduced Form



Failure of the identification assumption

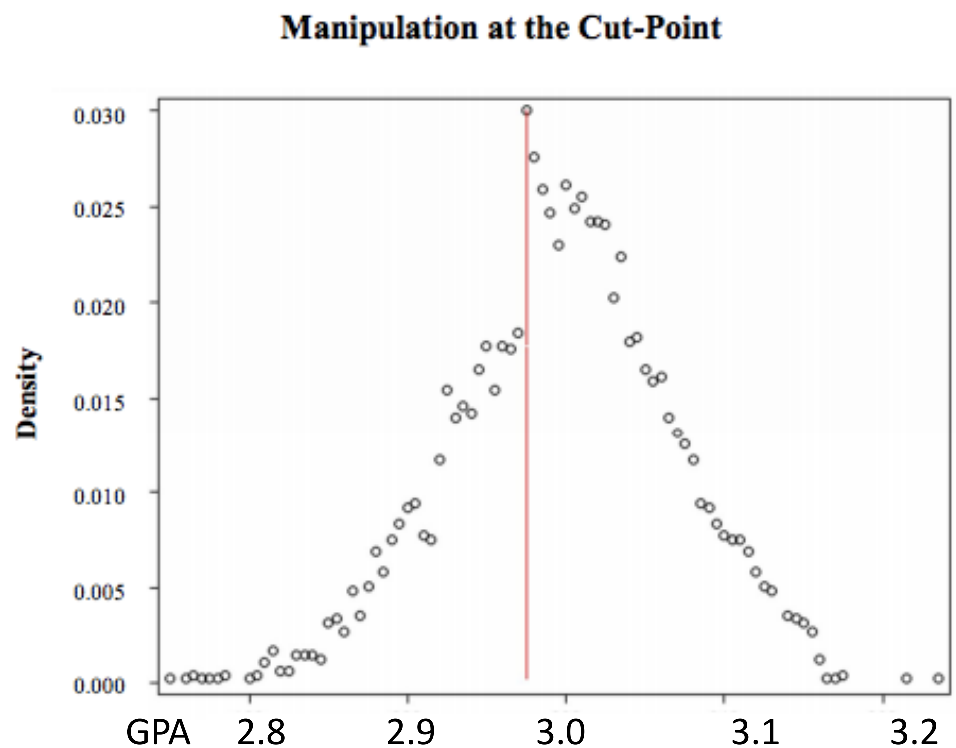
- The key assumption in RD is that individuals on either side of the cutoff are similar. More formally, this is an assumption about **continuity** of potential outcomes at the cutoff.
- The causal interpretation of RD fails if individuals on one side of the cutoff are systematically different than individuals on the other side.

RD empirical testing

- One of the big appeals to RD is that there are great tests to explore the potential threats to identification
- Two empirical tests:
  1. Examine the histogram smoothness of the running variable around the discontinuity. (to test manipulation)
  2. Examine the smoothness of many other covariates around the discontinuity. (to test manipulation and possible of contemporaneous policy changes)

Histogram test

- Draw a histogram of the running variable, showing distribution of observations (X-axis: GPA; Y-axis: frequency (density))
- If there is no manipulation of the running variable for the GPA example, what would the histogram of students look like?
  - If there is no manipulation, the histogram should be fairly smooth
- Can you draw a histogram when there IS manipulation?



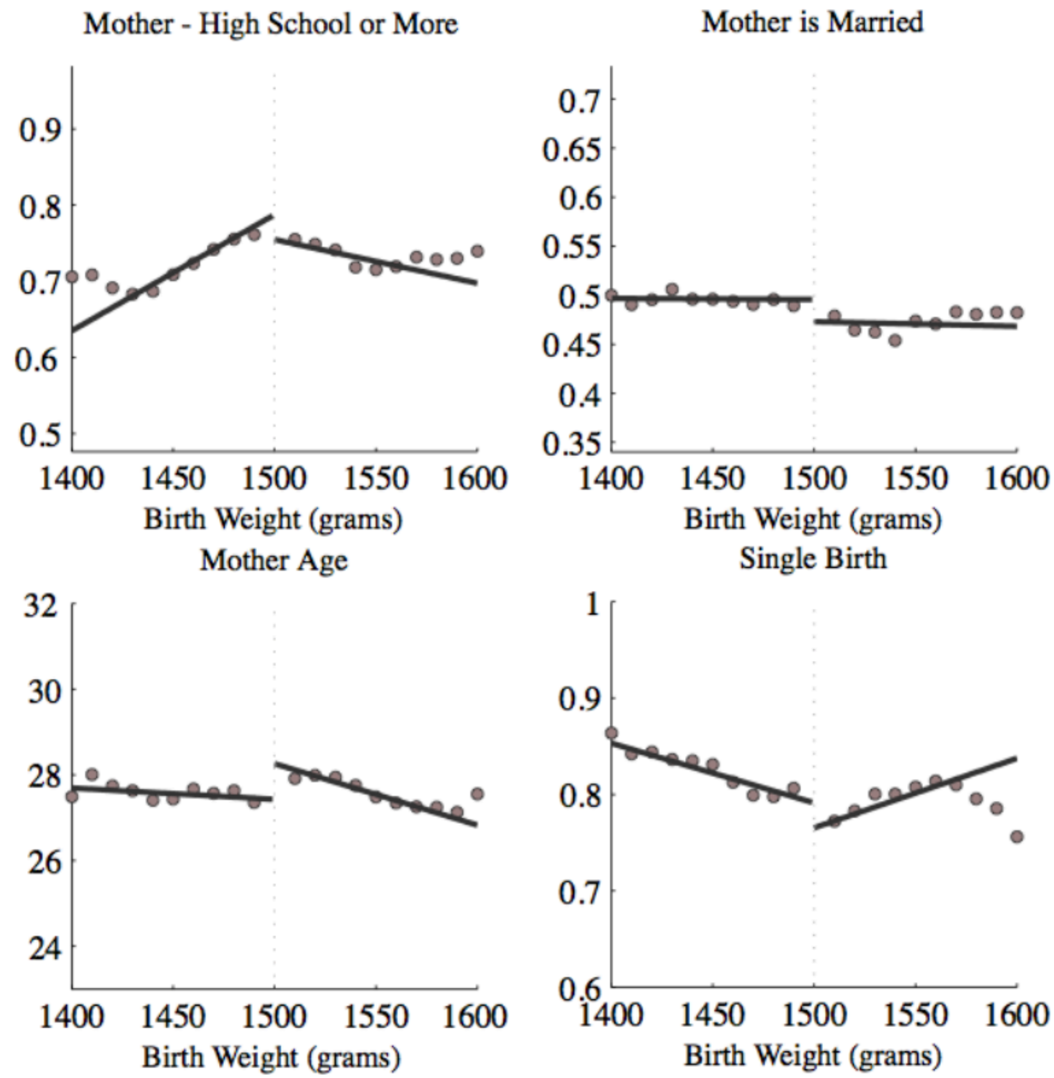
Covariate smoothness test

- Showing covariate (controls) smoothness around the discontinuity makes the RD design very convincing
- Example from the study of the impact of hospital interventions on academic performance
- To argue that children on either side of 1500g threshold are similar, we look at their covariates on either side

Covariate smoothness test-example

- Based on the covariates test, do you worry about manipulation?
- The idea is similar with the balanced table in OLS
  1. If covariates are smooth, we imply unobservables are smooth
  2. So the difference in Y is not caused by differences in observables and unobservables. **It's entirely due to the treatment status**

Figure 6: Baseline Covariates Around 1500 grams - Chile



Exercise: test RD identification assumption

- Recall the research question is whether class size has any causal impact on test score
- In Israel, public school class size is limited to 40 students
  - If there are 40 students in the grade, class size is 40
  - If there are 41 students in the grade, a second class is added and the average class size will be 20.5
- Question#1: Draw a histogram showing the case when there is manipulation (assume schools tend to have bigger class size to save money)
- Question #2: Draw a covariate test showing the case when there is manipulation (use family income, assume richer family deliberately send their kids to schools with more than 40 students, so that their kids have smaller class sizes)

RD estimates a LATE

- RD estimates a local average treatment effect (LATE). Why is that?
  - Comparing people just to the right with those just to the left-it's local
  - Fuzzy RD-compliers around the cutoff-more local

RD is not a before and after comparison

- Suppose a policy to reduce pollution is implemented in California. We cannot do regression discontinuity using just before the policy vs. just after the policy
- The spirit of RD must theoretically lim X$\rightarrow c$. In this case, comparing one second before the policy to one second after the policy makes no sense
- If we expand the time frame to look 1 year before vs. 1 year after, we no longer are sufficiently local to be able to assume nothing else changed
- If you are in this context, a Differences-in-Differences (DID) design is more appropriate

RD issues-failure of identification assumption

- The key assumption in RD is that individuals on either side of the cutoff are similar. More formally, this is an assumption about **continuity** of potential outcomes at the cutoff.
- The causal interpretation of RD fails if individuals on one side of the cutoff are systematically different than individuals on the other side.

Review for exams

- For RD implementation, understand the following formula,

$$Y = \alpha + \delta Above + \gamma S + \eta(S * Above) + \varepsilon$$

being able to explain the key parameters (α and β)

- Understand the first stage and reduced form of fuzzy RD
- Understand Why RD might fail