Lecture 3: Regression Discontinuity and Practical Applications

January 23, 2016

Causal Inference

- There are 5 basic empirical methods to obtain causal inference:

1. Controls (includes matching/fixed-effects) ✓
2. Randomized Experiments ✓
3. Difference-in-Differences ✓
4. Instrumental Variables ✓
5. Regression Discontinuity

Basic Intuition I

- Basic idea: In a randomized experiment, we randomized people to 'treatment' and 'control'
- What happens if a discontinuous (i.e. probability of getting treatment) assignment rule randomizes for us?
  - These are almost always institutional rules with a cutoff (i.e. get treatment if above cutoff, do not if below)
  - The intuition is similar to IVs
  - In fact, regression discontinuity can be considered as an IV

Basic Intuition II

- We see these institutional rules all the time
  - Cannot get Medicaid unless income is below 150 percent of the poverty line
  - Must have a GPA of 3.0 to get into university
  - Getting any more than 8 demerits means you fail your driving test
  - Getting 50 percent +1 vote means you win an election

- The idea to compare people that 'just' exceed the cutoff to those 'just' below the cutoff. So we compare those:
  - with an income of 150.1% of the poverty line to those at 149.9%
  - with a GPA of 2.99 to those with a GPA of 3.01
  - who got 8 demerits to those who got 9 demerits
  - who get a vote share of 50.1% to those who get a vote share of 49.9%

Basic Intuition III

- The people just below and just above the threshold should be similar as they got similar scores
- However, the individual just exceeding the threshold gets 'treated' while the one just below is 'untreated'
  - Thus, the individuals only differ in treatment status and so we can compare them to uncover the treatment effect

The Power of RD

- RD designs are considered "as good as randomized experiments"
- Biggest Reason: Every assumption is testable
- Therefore, unlike IVs we do not have to rely on 'story time' to convince someone our estimator is unbiased

Example I

- Question: Does getting an extra day in hospital after giving birth improve a child's health outcomes?
- In OLS: $Health_i = \alpha + \beta Length\ of\ Stay_i + \epsilon_i$
- What is the likely omitted variable bias in the above regression?

Example II

- In the 1990s California HMO reimbursement was such that women were granted 1 day of care of giving birth. However, a 'day' counted the number of 'midnights' in care.
- Consider two mothers giving birth a child on Monday night:
- One mother gives birth at 11:59pm: the child gets one midnight of care and they are released on Tuesday morning
- One mother givens birth at 12:01am: the child gets one midnight of care and they are released on Wednesday morning
- Thus, having a kid 2 minutes earlier gives the child another day in hospital

RD Basics I

- The (basic) math:
- What outcomes do we observe?
- What is our treatment and selection effect?

RD Basics II

- The (more complex) math:
- Just as in IVs, randomization may be imperfect
- Some newborns may stay in the hospital longer (or shorter) than one midnight
  - E.g. The newborn is sick or the mother hates hospitals
- We account for this by dividing by the probability randomization affected your treatment status:
  - If this number is 1, we call it a 'sharp RD'
  - If the number is below 1, we call it a 'fuzzy RD'

RD Terminology

- We call treatment determining variable the 'running variable' (or 'forcing variable')
  - So here our running variable is time (relative to midnight)
- 'Bunching' means that there are many more people 'just' below the discontinuity relative to 'just' above the discontinuity (or vice versa)

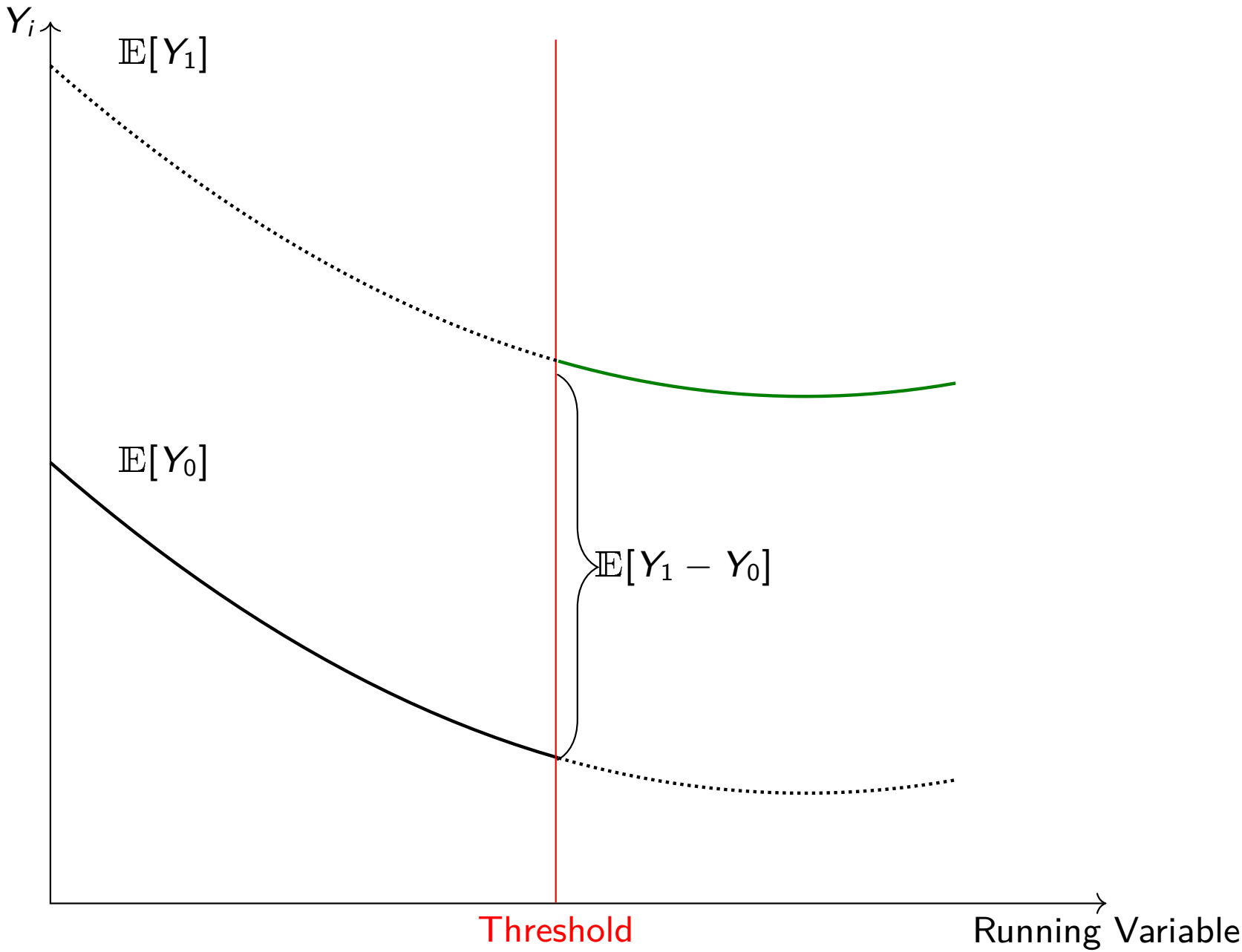
RD Assumptions

- The official RD assumption is that, in the absence of treatment, outcomes would be 'smooth' through the discontinuity
- Essentially this means that people 'just' below and 'just' above the discontinuity are similar
- We can check their similarity in two ways:
  - Check observables (as in a randomized experiment)
  - Check for bunching
- If bunching occurs, why is it likely that people on either side of the discontinuity may differ?

Adding more Data

- We would like to compare kids born at 11:59pm with those born at 12:01am
  - If we do this, however, we will have a very small sample size
- So we extend the data away from the discontinuity
  - As we get further away, precision will improve, but the bias will increase
  - How 'far away' we go is a judgement call. We call this distance the 'bandwidth'

Figure 1: Potential and Observed Outcomes



RD as a Regression

- Let the bandwidth be b.
- $Health_i = \alpha + \beta_1 AfterMidnight_i + \beta_2 BirthTime_i + \epsilon_i$ for $b \leq BirthTime_i \leq b$ (BirthTime here is relative to midnight)
  - More generally:
  - $Y_i = \alpha + \beta_1 Treat_i + \beta_2 RunningVar_i + \epsilon_i$ for $b \leq RunningVar_i \leq b$
- Why do we control for the running variable (here $BirthTime_i$)?
  - The 'correct' regression to run is actually: (why?)
  - $Y_i = \alpha + \beta_1 Treat_i + \beta_2 RunningVar_i + \beta_3 (RunningVar * Treat)_i + \epsilon_i$ for $b \leq RunningVar_i \leq b$

For the Keeners

- Just as in any regression, we do not know if the relationship between the running variable and the outcome is linear
- In this setting, failure to know this relationship is key. Visually:
- For this reason researchers should report results for three functional forms: linear, quadratic, and a triangular kernel

Another RD

- Our question is: Does being an incumbent congressman help you in future elections?
- OLS: $VoteShare_i = \alpha + \beta Incumbent_i + \epsilon_i$
- Possible failures of OLS in this instance?
- Idea: Use the fact politicians winning 50% + 1 vote become the incumbent

Framing the Research Question

- Our question is: Does being an incumbent congressman help you in future elections?

1. What is the unit of analysis?

Framing the Research Question

- Our question is: Does being an incumbent congressman help you in future elections?

1. What is the unit of analysis?
2. What is the treatment?

Framing the Research Question

- Our question is: Does being an incumbent congressman help you in future elections?

1. What is the unit of analysis?
2. What is the treatment?
3. What outcome are we interested in?

Framing the Research Question

- Our question is: Does being an incumbent congressman help you in future elections?

1. What is the unit of analysis?
2. What is the treatment?
3. What outcome are we interested in?
4. What are the counterfactual outcomes?

Framing the Research Question

- Our question is: Does being an incumbent congressman help you in future elections?

1. What is the unit of analysis?
2. What is the treatment?
3. What outcome are we interested in?
4. What are the counterfactual outcomes?
5. What is the causal link?

To STATA

- How do we mimic this?
- Open up STATA

Interpretation of RDs

- How do we interpret the RD estimate? (i.e. who complies with treatment?)
- Just like IVs, this is a *Local* Average Treatment Effect (LATE)
- This is (in my view) the biggest weakness of RDs

Internal Validity

- Check the covariates
- Check for bunching

External Validity

- Generalizability?
- Mechanisms?

Pros and Cons of RD

- Pros:
- Cons:

Causal Inference

- There are 5 basic empirical methods to obtain causal inference:

1. Controls (includes matching/fixed-effects) ✓
2. Randomized Experiments ✓
3. Difference-in-Differences ✓
4. Instrumental Variables ✓
5. Regression Discontinuity ✓

Paper 1

- Chetty, Hendren and Katz (The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment)
- Brief discussion of the paper

Framing the Question for Paper 1

- Our question is: What is the effect of moving a child to a better neighborhood?

1. What is the unit of analysis?

Framing the Question for Paper 1

- Our question is: What is the effect of moving a child to a better neighborhood?

1. What is the unit of analysis?
2. What is the treatment?

Framing the Question for Paper 1

- Our question is: What is the effect of moving a child to a better neighborhood?

1. What is the unit of analysis?
2. What is the treatment?
3. What outcome are we interested in?

Framing the Question for Paper 1

- Our question is: What is the effect of moving a child to a better neighborhood?

1. What is the unit of analysis?
2. What is the treatment?
3. What outcome are we interested in?
4. What are the counterfactual outcomes?

Framing the Question for Paper 1

- Our question is: What is the effect of moving a child to a better neighborhood?

1. What is the unit of analysis?
2. What is the treatment?
3. What outcome are we interested in?
4. What are the counterfactual outcomes?
5. What is the causal link?

Framing the Question for Paper 1

- Our question is: What is the effect of moving a child to a better neighborhood?

1. What is the unit of analysis?
2. What is the treatment?
3. What outcome are we interested in?
4. What are the counterfactual outcomes?
5. What is the causal link?
6. How do the researchers mimic this? What empirical design are they using?

Analyzing Paper 1

- Internal Validity?
- External Validity?

Paper 2

- Schlenker and Walker (Airports, Air Pollution, and Contemporaneous Health)
- Brief discussion of the paper

Framing the Question for Paper 2

- Our question is: Does increased pollution lead to negative health outcomes?

1. What is the unit of analysis?

Framing the Question for Paper 2

- Our question is: Does increased pollution lead to negative health outcomes?

1. What is the unit of analysis?
2. What is the treatment?

Framing the Question for Paper 2

- Our question is: Does increased pollution lead to negative health outcomes?

1. What is the unit of analysis?
2. What is the treatment?
3. What outcome are we interested in?

Framing the Question for Paper 2

- Our question is: Does increased pollution lead to negative health outcomes?

1. What is the unit of analysis?
2. What is the treatment?
3. What outcome are we interested in?
4. What are the counterfactual outcomes?

Framing the Question for Paper 2

- Our question is: Does increased pollution lead to negative health outcomes?

1. What is the unit of analysis?
2. What is the treatment?
3. What outcome are we interested in?
4. What are the counterfactual outcomes?
5. What is the causal link?

Framing the Question for Paper 2

- Our question is: Does increased pollution lead to negative health outcomes?

1. What is the unit of analysis?
2. What is the treatment?
3. What outcome are we interested in?
4. What are the counterfactual outcomes?
5. What is the causal link?
6. How do the researchers mimic this? What empirical design are they using?

Analyzing Paper 2

- Internal Validity?
- External Validity?

Scenario 1

- Suppose the minister of transportation for Ontario wants to raise the speed limit on the 401 to 120 km/h. He wants know whether this is a good idea (i.e. will it increase the number of fatal car crashes?). He notes that Alberta has less fatal crashes than Ontario but has a speed limit of 120 km/h and he also notes that the province of British Columbia raised their speed limits from 110 km/h to 120 km/h in certain highway sections in 2012. He asks you how you would go about answering his question?

Framing the Question

- Our question is: Will a speed limit of 120 km/h lead to an increase in fatal car crashes?

1. What is the unit of analysis?

Framing the Question

- Our question is: Will a speed limit of 120 km/h lead to an increase in fatal car crashes?

1. What is the unit of analysis?
2. What is the treatment?

Framing the Question

- Our question is: Will a speed limit of 120 km/h lead to an increase in fatal car crashes?

1. What is the unit of analysis?
2. What is the treatment?
3. What outcome are we interested in?

Framing the Question

- Our question is: Will a speed limit of 120 km/h lead to an increase in fatal car crashes?

1. What is the unit of analysis?
2. What is the treatment?
3. What outcome are we interested in?
4. What are the counterfactual outcomes?

Framing the Question

- Our question is: Will a speed limit of 120 km/h lead to an increase in fatal car crashes?

1. What is the unit of analysis?
2. What is the treatment?
3. What outcome are we interested in?
4. What are the counterfactual outcomes?
5. What is the causal link? (both pro and con)

Framing the Question

- Our question is: Will a speed limit of 120 km/h lead to an increase in fatal car crashes?

1. What is the unit of analysis?
2. What is the treatment?
3. What outcome are we interested in?
4. What are the counterfactual outcomes?
5. What is the causal link? (both pro and con)
6. How will we mimic this?

Answering the Question

- What data will we need? Is it obtainable?
- What is the actual regression you would run? How would you check its validity?

Analysis

- What will be the limitations of this study?

- Why does this study improve on the basic Alberta versus Ontario comparison? What potential biases may arise in that comparison?

Scenario 2

- Suppose the minister of education wants to introduce a similar program to 'Teach for America' in Canada. He has plenty of willing teachers and wants to evaluate the effects these teachers have in the schools they are placed. Just as in 'Teach for America' he wants to focus these teachers towards the poorest schools. Specifically, the minister wants there to be no school that is poorer than another *not* to have Teach for Canada teachers. The minister asks you how the program should be designed to allow for a rigorous evaluation.

Framing the Question

- Our question is: Do 'Teach for Canada' teachers improve student outcomes?

1. What is the unit of analysis?

Framing the Question

- Our question is: Do 'Teach for Canada' teachers improve student outcomes?

1. What is the unit of analysis?
2. What is the treatment?

Framing the Question

- Our question is: Do 'Teach for Canada' teachers improve student outcomes?

1. What is the unit of analysis?
2. What is the treatment?
3. What outcome are we interested in?

Framing the Question

- Our question is: Do 'Teach for Canada' teachers improve student outcomes?

1. What is the unit of analysis?
2. What is the treatment?
3. What outcome are we interested in?
4. What are the counterfactual outcomes?

Framing the Question

- Our question is: Do 'Teach for Canada' teachers improve student outcomes?

1. What is the unit of analysis?
2. What is the treatment?
3. What outcome are we interested in?
4. What are the counterfactual outcomes?
5. What is the causal link? (both pro and con)

Framing the Question

- Our question is: Do 'Teach for Canada' teachers improve student outcomes?

1. What is the unit of analysis?
2. What is the treatment?
3. What outcome are we interested in?
4. What are the counterfactual outcomes?
5. What is the causal link? (both pro and con)
6. How will we mimic this?

Answering the Question

- What will you need to do to ensure you can implement your empirical design?
- What data will we need? Is it obtainable?
- What is the actual regression you would run? How would you check its validity?

Analysis

- What will be the limitations of this study?
- Why does this study improve on a basic 'before and after' comparison?