PMDA

Lecture 8: Regression Discontinuity Designs

Coming up

- This class: RD design
- Next class:
  - Predictive methods
  - Causal inference using predictive methods such as Lasso and Trees (forests)

Regression Discontinuity Design (RD)

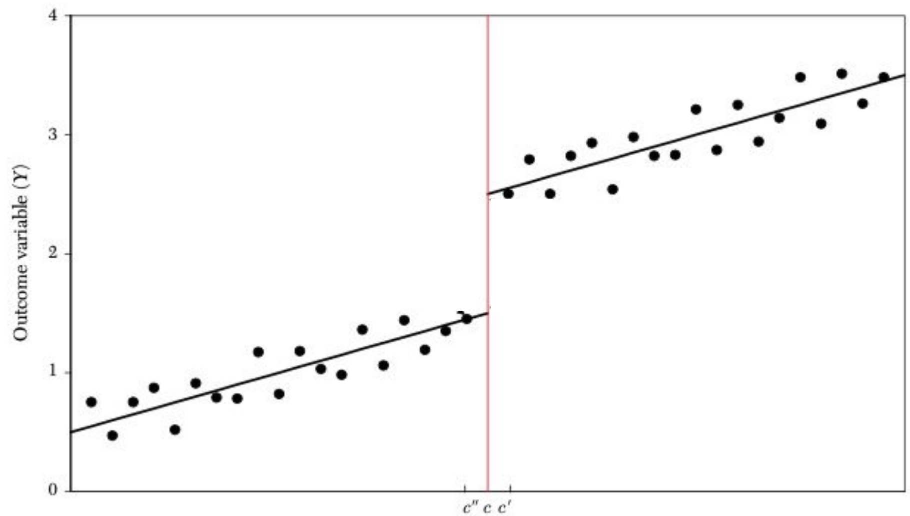
- Regression discontinuity (RD) is a way of estimating the treatment effect in non-experimental settings when the treatment is determined by whether an observed <u>“assignment” variable</u> (a..k.a “running variable”) surpasses a known <u>cutoff</u>

- We deduce the causal effect of a treatment that is determined by a mechanical rule
  - Individuals above the cutoff receive treatment
  - Individuals below the cutoff fail to receive treatment
  - If individuals just above and just below the cutoff are similar, we can compare them to calculate the treatment effect

- The latter can happen when individuals **cannot precisely control the assignment variable near the cutoff**: this leads to **randomization**

Running Example: Impact of Merit Awards

- Consider the impact of merit award on future academic outcomes:
  - Allocation of awards is based on an observed test score
  - Rule: Only students with test scores (X) above a cutoff (c) receive the award



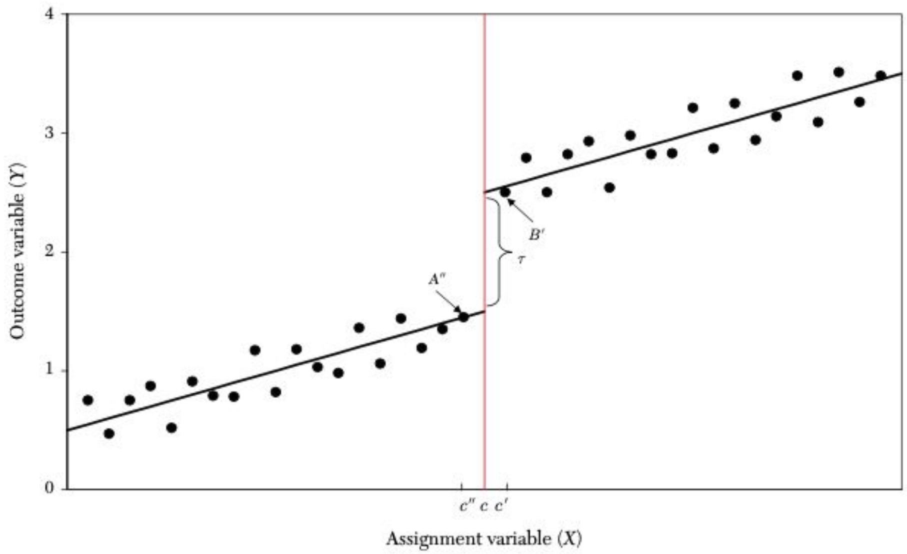
Lee and Lemieux: Regression Discontinuity Designs in Economics

- **Q:** How would you estimate the treatment effect (τ)?

A: Key idea: individuals with a test score just below the cutoff are very similar to those with a test score just above the cutoff

  - Treatment group: students with X>=c
  - Control group: students with X<c

Running Example Continued



Lee and Lemieux: Regression Discontinuity Designs in Economics

Q: Are B’ and A’’ reasonable guesses for treatment and control outcomes for someone that scores at X=c?

- **A:**
    - If all factors are evolving smoothly (will make this precise later) w.r.t to X then individuals B’ and A’’ are similar in all other respects except on the award
    - τ = B’-A’’
    - Cannot get too close to c => need to use data away from c

A Formal RD Framework: Sharp RD

- The most common approach is:

$$Y = \alpha + \tau D + \beta X + e$$

where

$$D = 1\{X \geq c\}$$

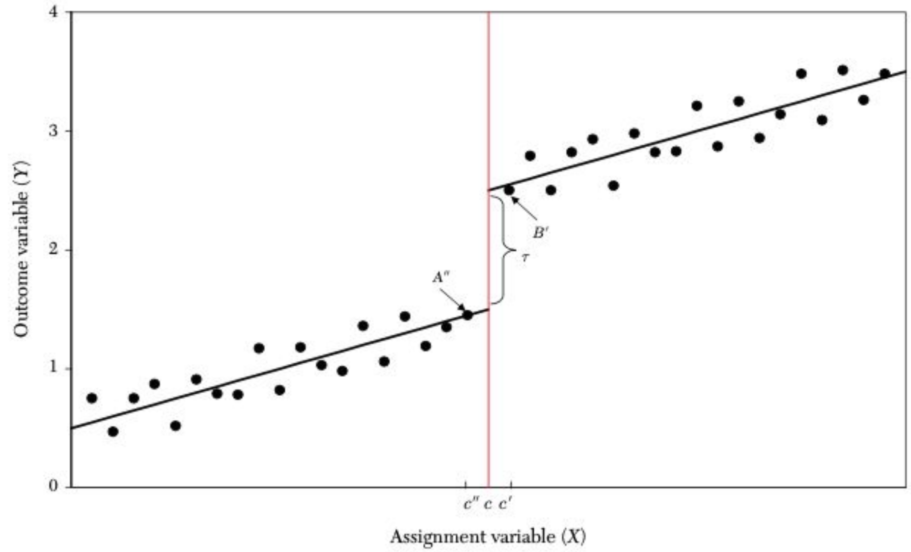
  - D = "receipt of treatment" is takes values 0 and 1 (award allocation)
  - X is the assignment variable (test score)
  - c is the cutoff (minimum score needed to receive award)
  - Y is the outcome (academic performance later on)

- when the relationship between X and Y is otherwise linear, the equation captures both lines (above and below the threshold)
- If the only reason for Y to be discontinuous is crossing the cutoff then: we attribute the jump of Y at c to the causal effect of treatment measured by τ

Q: Is there OVB in the regression above?

A: D is determined solely by X. Assuming that the effect of X on Y is captured by a linear function, we can be sure that no OVB afflicts our regression

Two Important Points

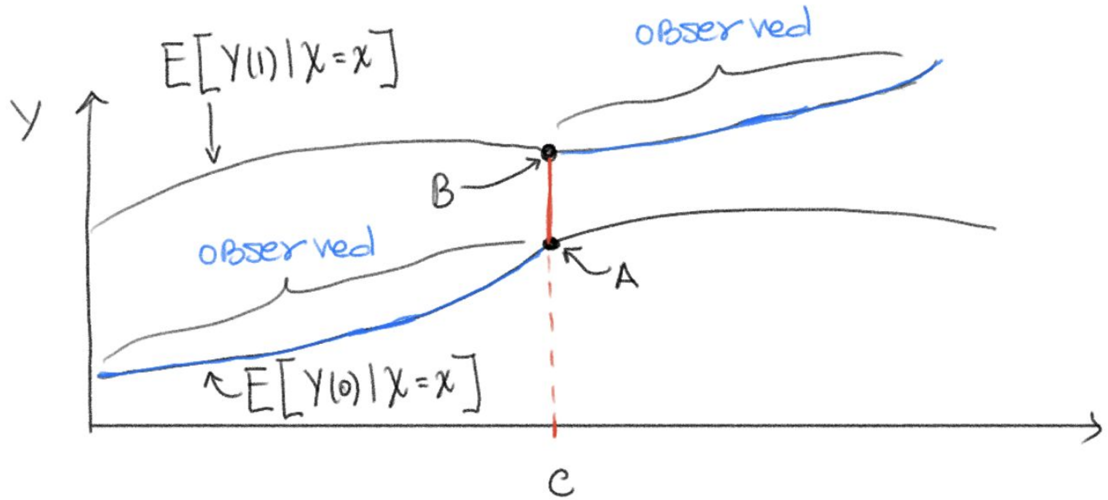


Lee and Lemieux: Regression Discontinuity Designs in Economics

1. **“All other factors” that determine Y must evolve smoothly w.r.t to X: If other variables jump at c then τ will be biased**
   - We need this because for no value of the running variable X we observe outcomes for treatment and control group (fail to satisfy the overlap assumption)
   - In RD, we extrapolate across values of the running variable in the neighborhood of c
2. Since we need data away from c, the estimate will depend on the chosen functional form:
   - e.g., if we impose β =0 then the estimate would be biased

Potential Outcomes Framework

- Y_i (1) = what would occur if exposed to treatment
- Y_i (0) = what would occur if not exposed to treatment
- Y_i (1)-Y_i (0) = treatment effect
- We cannot observe Y_i (1) and Y_i (0) simultaneously => Use ATE



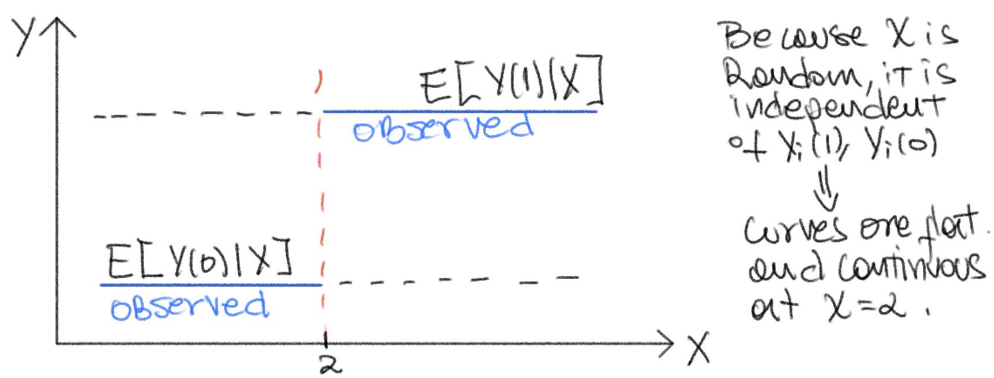
$$\lim_{\varepsilon \downarrow 0} E[Y_i|X_i = c + \varepsilon] - \lim_{\varepsilon \uparrow 0} E[Y_i|X_i = c + \varepsilon] = E[Y_i(1) - Y_i(0)|X = c]$$

- This is the average treatment effect at the cutoff c and it equals τ

Key assumption: All other factors are continuous w.r.t X

RD as a Local Randomized Experiment

- Consider a randomized experiment:
  - v ~ U[0,4]; If v >=2 then units get treated, otherwise they don't
  - This is like an RD design with X = v and c = 2
  - In this context, continuity is a consequence of randomization



RD as a Local Randomized Experiment

- Imagine, agents are compensated for a bad draw: get a multiple of 1/v in money

    - Y = found a job or not within 1 month? and D =1 if gets job search assistance

    - **Q:** What would the potential curves look like in this case?

      A: They would slope upwards: people with less compensation will work harder to get a job than people with more compensation

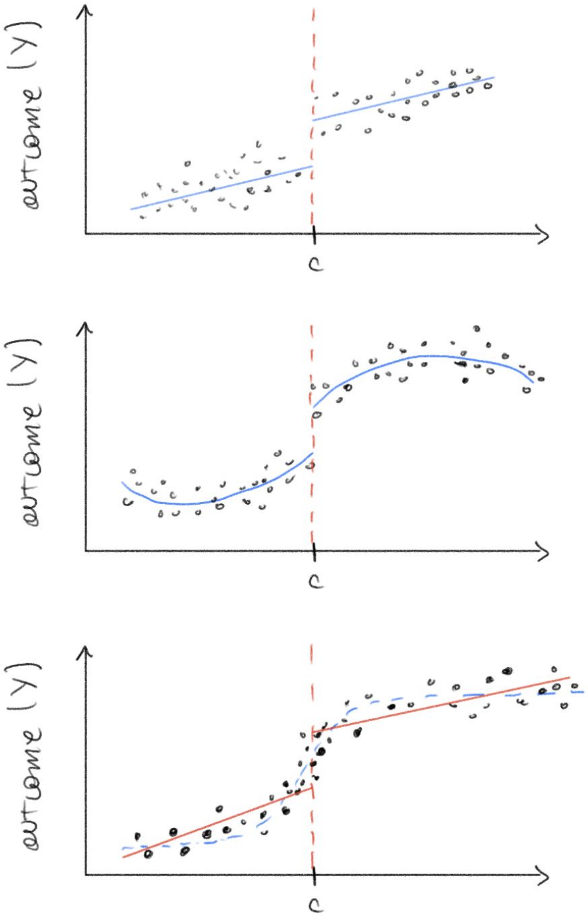
    - **Q:** Can we simply compare the means? **A:** No, bias estimator

Focus just above and just below c: gives a consistent estimator because these people receive essentially the same compensation

Can check if randomization worked by comparing the baseline covariates of the two sides of c

Functional form issues

- RD does not guarantees to produce reliable causal estimates



⇒ Model nonlinearities or focus on observations near the cutoff

Sharp RD: Non-linear case

- Typically, we model nonlinearities using polynomial functions of X
- For example, an RD with quadratic running variable control:

$$Y = \alpha + \tau D + \beta_1 X + \beta_2 X^2 + e$$

- We can also have different X coefficients to the right and left of c:
  - Center running variable: (X-c)
  - Replace X with (X-c)
  - Add an interaction term: (X-c)*D

$$Y = \alpha + \tau D + \beta(X - c) + \delta(X - c) \cdot D + e$$

Q: Why would we want to allow for different coefficient to the left and right?

A: In our running example:

- To the left of c: students don't receive the award. If we extrapolate what would happen for this group, we may see that that higher scores lead to better outcome at a diminishing rate
- To the right of c: students do receive the award, this may benefit more (marginally) student with very good test scores than students just above the threshold

Sharp RD: Non-linear case

Q: What is the treatment effect for X>>c away from the cutoff?

$$\tau + \delta(X - c)$$

Be careful with extrapolating! We don't have counterfactual data on what occurs at X >>c for individuals that are not treated

Q: Rewrite a model that includes quadratic trends and changes in slope at the cutoff so that both linear and quadratic terms change at c

$$Y = \alpha + \tau D + \beta_1(X - c) + \delta_1(X - c) \cdot D + \beta_2(X - c)^2 + \delta_2(X - c)^2 \cdot D + e$$

Sharp RD: Nonparametric approach

- For points around the cutoff, nonlinear trends are less of a concern
- ⇒ Compare average in a narrow window around c
- Tradeoff:
  - If window is too narrow, we have fewer observation and, hence, less precision
  - As we come closer to c, we can have less bias due to wrong specification
- Nonparametric RD:

$$Y = \alpha + \tau D + \beta X + e$$

$$\text{with } c - b \leq X \leq c + b$$

b = width of the window or *bandwidth*

Sharp RD: Nonparametric approach

$$Y = \alpha + \tau D + \beta X + e$$
$$\text{with } c - b \leq X \leq c + b$$

- How should we pick the bandwidth?
  - To avoid concern about polynomial choice, we would like to be close to c
  - Less data means less precision
  - ⇒ b should vary as a function of sample size: the more data around c, the narrower the bandwidth while still having precision
  - “Imbens, G. and Kalyanaraman, K., 2012. Optimal bandwidth choice for the regression discontinuity estimator. The Review of economic studies, 79(3), pp.933-959.”

Which model should you use?

- Note that both parametric and nonparametric method can be bias:
  - When the parametric functional form is incorrect
  - When the sample consists of points only around the cutoff and using a local linear regression (unless the true function is linear)
- Linear vs nonlinear:
  - Simple (linear) RD: can capture what happens at the threshold well
  - Nonlinear RD: can also capture changes in behavior beyond the threshold
- Nonparametric:
  - Good if we have enough data around cutoff
  - Try different choice of bandwidth to show results are not a fluke
- **Important.** Do not to rely on one particular method or specification: check for stability of your results across alternative and equally plausible specifications [more on this during the workshop]

Valid or Invalid RD design?

- **Key question you should ask yourself:** Are individuals able to control X? If so, what is the nature of this control?
- **Q:** Why could this be a problem?

  A: If agents have a lot of control over X and they can benefit from controlling X, then agents to the left and to the right of c maybe different!

Valid or Invalid RD design?

- **Example:** Suppose we have two type of students:

  A: have more ability and know that being above c=50% gets them the award

  B: less ability, don’t know about award

  50% of questions trivial to answer correctly

  Due to random chance students make errors in their initial answer, but can correct the error if they check their work

- **Q:** What type of students are above c, what type of students are below c?

  A: Only type B students will be below c because type A students will check their answers and correct the errors

  - ⇒Marginal failing student is not a valid counterfactual for the marginal passing student: RD is invalid

Valid or Invalid RD design?

- Suppose now that questions are not trivial
- There are no guarantee passes, no matter how many times a student checks

Q: What type of students are above c, what type of students are below c?

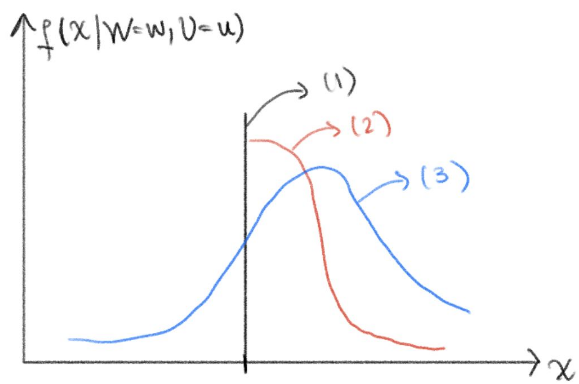
> **A:** It seems more plausible that for students around c, it is a matter of luck if they are above or below. They can exert effort, but they don't know the score they will get ⇒ RD is more reasonable here

Important:

- You need to have some **knowledge about the mechanism generating X** beyond the obvious
- **Check for precise vs imprecise manipulation**: precise sorting around the cutoff is self-selection which is exactly what we try to avoid in causal inference!

Some Control Mechanisms

- (1) Complete and exact control over X: degenerate distribution -> In repeated trials the individual would choose the same X
- (2) If some error but can still have precise control over whether they receive treatment: truncated distribution
- (3) If some error and do no have precise control over X: density of X conditional on observable (W) and unobservable (U) determinants of the outcome should be continuous



- **Imprecise control implies that variations in the treatment status will be randomized in a neighborhood of c:**

$$P[W = w, U = u | X = x] = f(x | W = w, U = u) \frac{P[W = w, U = u]}{f(x)}$$

⇒ All observed and unobserved predetermined characteristics will have identical distribution on either side of X=c in the limit: the treatment is “as good as” randomly assigned around the cutoff

When Does RD Work?

Summary of assumptions:

- The treatment is based on a cutoff or threshold value of an observable variable (the “running variable”)
- Only the treatment causes discontinuity in the dependent variable
- All other factors (observable and unobservable) are smooth at the threshold
- Units cannot precisely manipulate the assignment variable to influence whether they receive treatment or not
  - This is what gives the randomization
  - That is, randomization is a consequence of the assumption that individuals have imprecise control over the assignment variable

How do we test the RD assumptions?

- It is always good practice to check visually that there is a jump at X=c
- If the treatment around the threshold is approximately randomized then all baseline characteristics should have the same distributions just above and and below the cutoff
    - This is like checking that P(W=w|X=x) is continuous
    - We cannot check continuity of P(W=w,U=u|X=x) because we don't observed U
    - Most favorable statement: "the data failed to reject the assumption of randomization"
- If the density of X for each individual is continuous (that is, f(x|W,U) is continuous) then the marginal density of X over the population (and observable, f(x) and f(x|W)) should be continuous as well

Final thoughts: Weighted Avg. Treatment Effect

- We saw that τ measures the gap at the cutoff c

Q: Does this mean that RD delivers only a credible treatment effect for the subpopulation of individuals at c and nothing about the treatment away from c?

A: This statement is overly simplistic. The discontinuity gap in RD design can be interpreted as: "a weighted average treatment effect across all individuals"

Let τ(w,u) be the completely unrestricted heterogeneity in treatment effect

$$\lim_{\varepsilon \downarrow 0} E[Y_i|X_i = c+\varepsilon] - \lim_{\varepsilon \uparrow 0} E[Y_i|X_i = c+\varepsilon] = \sum_{w,u} \tau(w,u) P[W = w, U = u | X = c]$$

$$= \sum_{w,u} \tau(w,u) \frac{f(c|W = w, U = u)}{f(c)} P[W = w, U = u]$$

- The weights are the terms that multiply τ(w,u)
- They are proportional to the ex-ante likelihood that an agent's realization of X will be close to c
- More weights is put on individual that are more likely to be close to c

Workshop

- Go to BruinLearn and download the workshop + dataset
- The workshop is based on Chapter 4 of “Mastering Metrics”
- The data set contains data about mortality rates around age 21
- We want to see if MLDA has a causal impact on mortality rates