Causality 2: Various Techniques

Graphs and some discussions in this handout are borrowed from Cunningham (2021) *Causal Inference.*

Today, I will discuss the following three techniques very briefly.

1. Difference-in-Difference (DID) Estimation

2. Regression Discontinuity Design, RDD

3. Propensity Score Matching, PSM – Inverse Probability Weight, IPW

Difference-in-Difference (DID) Estimation

- In essence:

    1. Two+ units will show the same change over time (common trend assumption).
    2. There is a treatment (= e.g., a policy change) in one unit (the treatment group), while no such treatment in the other units (the control group).
    3. Without a treatment, both the treatment group and the control group will show the same magnitude of changes over time (again, the common trend).
    4. With the treatment, the difference in the change over time between the treatment group and the control group is a result of the treatment (= causality).
    5. The example we discussed in the previous class, NY and NJ minimum wage, assumes that without the hike in minimum wage in NJ, the changes in employment over time will be the same for both NY and NJ. Thus, the changes in employment in NJ over time will be the same as that in NY, which becomes a counterfactual case.
    6. As shown in the graph below, the red line and the dotted blue line are the common trend without the treatment. The gap between the expected common trend (dotted blue line) and the actual result of the treatment group (solid blue line) is the DID estimate.
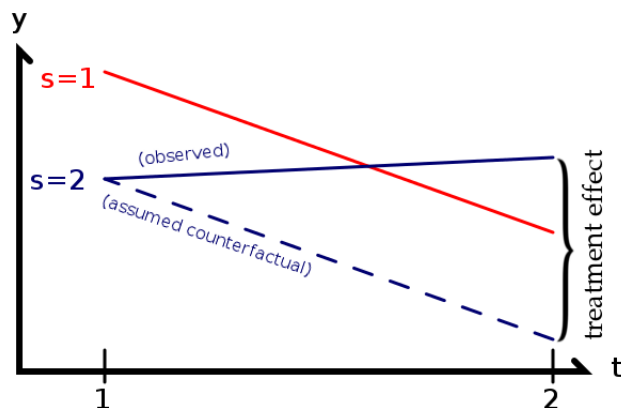
	Treatment Group	Control Group
Before (Time 0)	$\bar{Y}_b^t$	$\bar{Y}_b^c$
After (Time 1)	$\bar{Y}_a^t$	$\bar{Y}_a^c$

$$\text{DID} = (\bar{Y}_1^t - \bar{Y}_0^t) - (\bar{Y}_1^c - \bar{Y}_0^c)$$
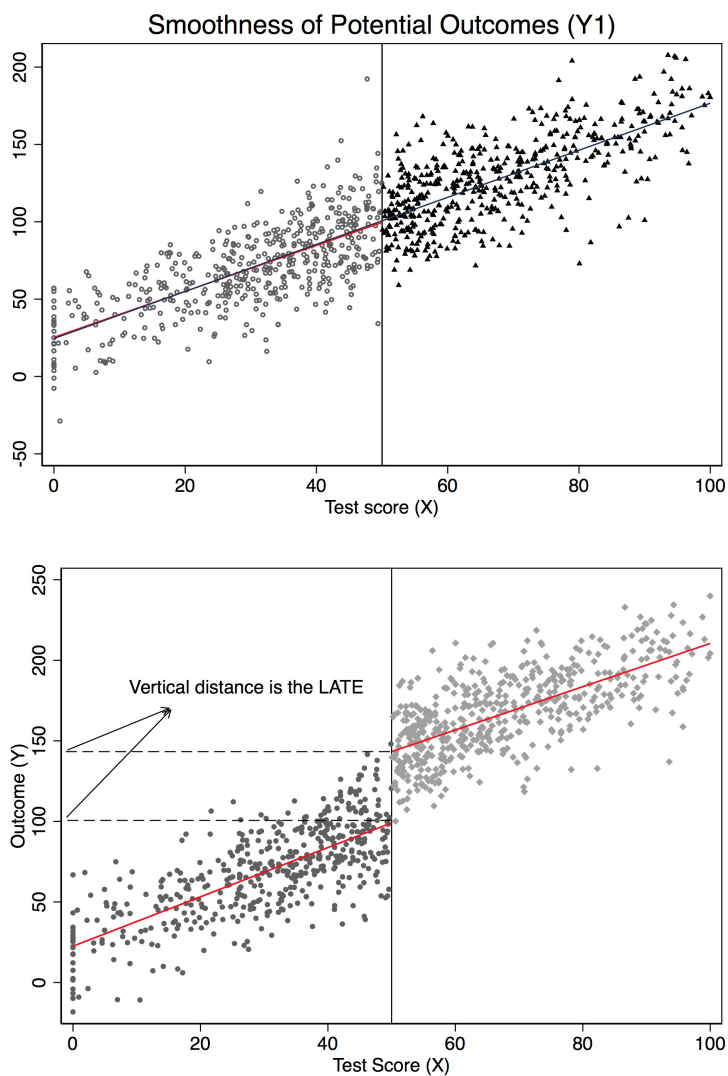
- Key Assumption: Common trend assumption

- Data requirements:

  - A treatment
  - Multiple units which share the common trend.
  - Before and after observations for both the treatment group and the control group

- DID estimation is basically a simple interaction effect. Mathematically and statistically, this is a very simple method.

$$y = \alpha + \beta T + \gamma G + \delta(T \times G) + e$$

  where T is a time variable and G is the dummy of the treated group. DID estimator is δ.

Regression Discontinuity Design, RDD

- In essence:

  - The outcome is a continuous function of an independent variable. E.g., the earnings are a function of the SAT scores.
  - Let's say there is a treatment in the independent variable. E.g., the college admission at the 1400 SAT score.
  - The gap between the group who is just above the treatment point (e.g. admitted into the college because the SAT score is 1,400) and the group who is just below the treatment point (e.g., not admitted into the college because the SAT score is 1,399) is a result of the treatment (= college admission).
  - Here the outcome is earnings; the treatment is the college admission; and the variable of the continuous function is the SAT scores.
  - The difference in the outcome between the just below and above the treatment point is the local average treatment effect, or LATE.
  - In the graphs below, the outcome variable (= y-axis) is a linear and continuous function of the independent variable (= x-axis). When there is a treatment at the test score, there is a jump between the test score right below the cut-off point and the test score right above the cut-off point.
  - the difference between two graphs is the LATE, a result of the treatment.

Smoothness of Potential Outcomes (Y1)

- Key Assumption: Continuity assumption. E.g., the earnings should be a continuous function of the SAT scores.

- Mathematically, the RDD is fairly simple.

$$y = \alpha + \beta X + \gamma D + \delta(D \times X) + \sum \eta Z + \epsilon$$

where,

- y: dependent variable
- X: key control variable
- D: a dummy variable - 1 if X is above the threshold (=treatment), 0 if X is below the threshold.
- X: other control variables.

Here, the coefficient estimated of our key interest is δ.

- In DID, the interaction effect is of main interest. In the RDD, the estimate coefficient of the dummy variable net of the interaction effects is of main interest.

- In the graphs above, the association between X and Y is linear. However, in reality, it can be curvilinear as follows:
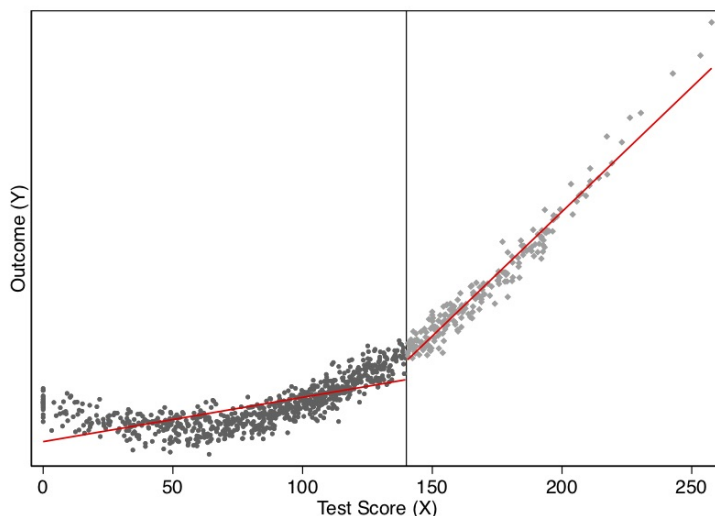
- To address the nonlinearity, add quadratic terms:

$$y = \alpha + \beta_1 X + \beta_2 X^2 + \gamma D + \delta_1 (D \times X) + \delta_2 (D \times X^2) + \sum \eta Z + \epsilon$$

- For the RDD, the best way to present your results is to draw graphs.

- Two RDD: Sharp RDD and Fuzzy RDD

   – Sharp RDD: The probability of the treatment moves from 0 to 1 before and after the threshold (= cutoff point). E.g., if your SAT is above 1,400, you're admitted. If below 1,400, you're not admitted.

   – Fuzzy RDD: The probability of the treatment in not changing from 0 to 1 before and after the cutoff point. Instead, it changes somewhat gradually.

   – In the graph below, the dotted line is the case of the sharp RDD and the solid line is the case of the fuzzy RDD.
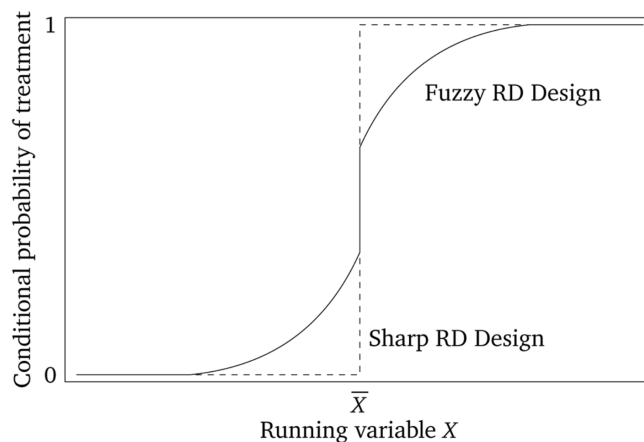
- E.g., the review process for the college admission is holistic. Some who scored 1,300-1,400 in SAT are admitted, while some who scored 1,400-1,500 in SAT are denied. In this case, we cannot create a dummy variable D.

- Fuzzy RDD: Here is how to estimate the fuzzy RDD

  1. Estimate Logit (or LPM or Probit).
     $D = b_0 + b_1 X + \sum cZ$
  2. Compute the expected probability of the treatment.
  3. Add the expected probability of the treatment in the place of D.
  4. The estimated coefficient, γ, quantifies the expected change in Y when the probability of D moves from 0 to 1 at the point of the cut-off. Note that the actual difference in the probability of the treatment before and after the cut-off point is less than 1.

- Data requirements:

  - A treatment
  - Data which fulfills the continuity assumption between X and Y.
  - Because the RDD estimate the local average treatment effect (LATE), it requires a large sample size.
  - In many cases, administrative records are utilized for the RDD

- Stata

  - Software: https://rdpackages.github.io/

Propensity Score Matching – IPW

- In essence:

  - There are various PSM techniques. Today, I will discuss the inverse probability weighting (IPW) method. In Stata, you can use `teffects ipw` option. The IPW is a technique which adjusts the probability of the treatment by weighting.

– All PSM techniques have basically the same logic. Based on the observed characteristics (X), you estimate the probability of the treatment (T), and balance the probability of the treatment between the treatment group and the control group. As a result, the treatment between the treatment group and the control group show the same distribution of X.

– Let's assume that we are interested in the effect of X on Y. But there is a confounder Z. E.g., those who are more ambitious (Z) are more likely to go to college (X). Thus, when you estimate

(1) $Y = \alpha + \beta X + \sum \gamma W + e,$

β is biased because it reflects not only X but also Z. This is a confounding problem due to selection bias.
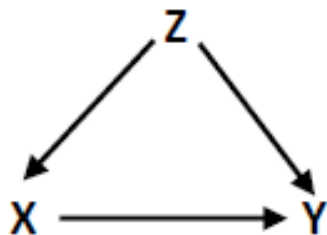
– But if the selection characteristics are observed in the data, we can condition treatment on those characteristics to get an unbiased estimate of treatment effect.

– OK, here is the process to apply the IPW technique:
  1. Estimate selection to treatment. That is, Loit, LPM, or Probit.
     $X = a + \sum bW + e$
  2. Compute the expected probability of the treatment (or selection). Let's say it is P. In Stata, you need to use the `predict` function.
  3. Assign the inverse of probability of treatment for treated individuals (i.e., $1/P$) and the inverse probability of not being treated for control individuals (i.e., $1/(1-P)$).
  4. In this stage, you create a new weighting value.
     ```
     gen     newwt = .
     replace newwt = 1/P     if treatment==1
     replace newwt = 1/(1-P) if treatment==0
     ```
  5. Re-estimated the usual

     (2) $Y = \alpha' + \beta' X + e,$

     applying the new weighting value. In this stage, no other control variables are necessary. That is, you don't need to control for W. You basically estimate the mean difference of Y between the treatment group and the control group.
  6. The gap between β' of Eq (2) and β of Eq (1) is the amount of bias caused by the selection.

– Note that when you apply the new weight, the difference in the means of the independent variables, W, between the treatment group and the control group should be insignificant. This is called 'balanced.' If there are statistically significant differences between the

treatment group and the control group for some independent variables, you need to add those independent variables as control variables.