Final Topic List

Sunday, May 1, 2022 3:47 PM

1. Basic Understanding of Linear Regression
    a. Understand the output of a linear regression, including being able to interpret meaning of coefficients
        i. Be able to work with logarithm transformation of DVs (if you need to interpret these coefficients, you can write exp(coeffecient)
    b. Be able to interpret interaction effects.
    c. Different types of treatment effects

2. Basic Understanding of Causal Inference
    a. Know and understand the common threats to causal inference, namely omitted variable bias (confounding), simultaneity, measurement error, and sample selection.
    b. Requirements for establishing causality.

3. Randomized Experiments
    a. Basics: what is randomization, what types of endogeneity does it solve, and how?
    b. Sample size calculation: which information is needed; what is statistical significance (alpha), power, Type I vs. Type II error, sample size too small/large
    c. Randomization / balance checks
    d. How to analyze experimental data: t-tests, linear regression, full factorial design
    e. Threats of validity of experiments (interference, social desirability etc.)
    f. Block randomization vs. on the fly randomization

4. Matching Methods
    a. Assumptions behind matching, why does it work?
    b. Procedure of propensity score matching and coarsened exact matching;
    c. Sensitivity analyses of matching results

5. Panel Data Models
    a. What is panel data;
    b. What types of confounding variables can be addressed with panel data models;
    c. How does a fixed-effect regression work; different ways to estimate fixed-effect regressions;
    d. Underlying assumptions behind fixed-effect and random-effect regressions; how to choose between the two (specific details of Hausman test are not tested).

6. Difference in Differences
    a. What types of confounding variables can be addressed with DiD;
    b. Assumptions behind DiD model;
    c. What does a DiD model look like;
    d. Dynamic DiD model: why do we use it, what does it tell us;
    e. Placebo test.

7. Synthetic Control
    a. When does synthetic control make sense to use?
    b. Conceptually, how do we implement a Synthetic Control estimation?

8. Instrumental Variables
    a. What is an instrumental variable?

b. What are the criteria for a valid instrumental variable?
c. How to run an instrumental variable regression;
d. How to evaluate the quality of instrumental variables
e. What does it mean for an IV regression to be under, over or just-identified?

9. Regression Discontinuity
    a. When does it make sense to use regression discontinuity, what is the intuition?
    b. What assumptions does RDD require?
    c. How do you implement a regression discontinuity design?
    d. What is bandwidth?
    e. What are the limitations of RDD?

Week1

Sunday, May 1, 2022 3:53 PM

1. Remember that we are interested in relationships in this class!
    a. We are interested determining the value of $\beta 1$
    b. And if this is significantly different from 0
    c. R2 measures how well the model fit the data

2. Type 1 Type 2 error
    a. Type 1
        i. Acutally not biased, but because of the error in data, we can make wrong conclusion
        ii. Or we have not enough evidence (p>0.05), but reject H0
        iii. It is more dangerous than type 2
        iv. p-value = probability of type 1. So small p-value -> less chance of type 1
    b. Type 2
        i. We have enough evidence (P<0.05) but accept H0
        ii. Too rigorous but less dangerous

Type 1 and Type 2 errors

		State of the World	
		H0 (Actual Fair Coin)	H1 (Actual Biased Coin)
Decision	H0 (Inferred Fair Coin)	Correct Acceptance	Type II Error (b)
	H1 (Inferred Biased Coin)	Type 1 Error (a)	Correct Rejection

Week2

Sunday, May 1, 2022 4:37 PM

Threats to Causal Inference

1. Omitted Variables
   a. Unobserved Confounders for X

Intuition: Omitted Variable Bias

- You'd like to estimate…

$$y_i = \alpha + \beta x_i + \gamma z_i + u_i$$

b.
- But you forget about z, and instead estimate…

$$y_i = \alpha + \beta x_i + \varepsilon_i \quad \text{(where } \varepsilon_i = \gamma z_i + u_i\text{)}$$

- If x and z are correlated, and z influences y, then the effect of z will be attributed to x in your regression, inflating or deflating the estimate.
  - If x and z are uncorrelated – no omitted variable bias present!

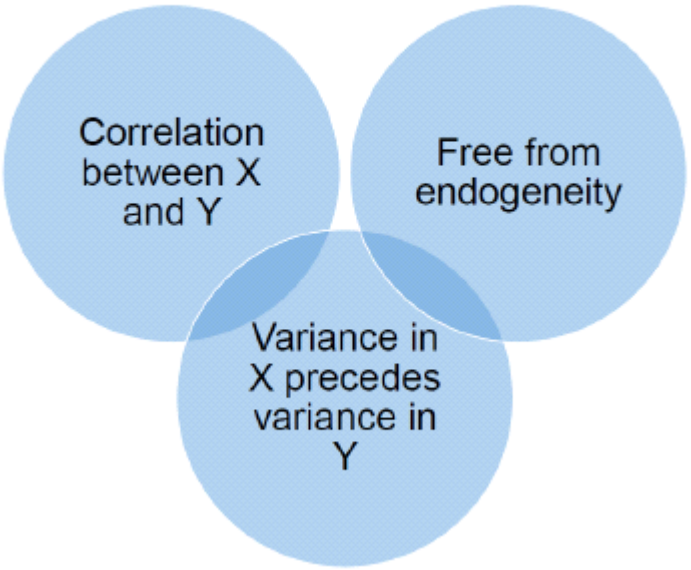
2. Selection Bias
   a. Whether we Observe a Representative Sample. Sample statistic simply does not reflect the population
   b. You only observe y for some values of x
   c. You can't recover a generalizable estimate of x's effect on y.
   d. your data is no longer representative

3. Simultaneity
   a. X and Y Cause Each Other = Matthew Effect
   b. When two variables are co-determined, you have a feedback loop.
   c. There's no way to isolate what portion of the observed association is due to X's effect on Y (and not the other way around).

4. Mis-Measurement
   a. Our Measure ofX is Imperfect.
   b. Because your measure ofx is 'noisy', your estimates ofx's effect on y understate the actual effect (if measurement error is random)
      i. Systematic measurement error can lead to bias
      ii. However, even random measurement error in independent variables can bias estimates

5. Endogeneity
   a. Omitted variables, simultaneity, and measurement error are formally referred to as sources of endogeneity.
      i. Formally: x is correlated with the regression error term.
      ii. Informally: x is associated with relevant, yet unmodeled factors in the regression.

Recap: Establishing Causality



Week3

Wednesday, March 9, 2022 7:47 AM

Statistical power

1. Overview
    a. From 2 samples, we want to prove that they are from the different distribution
    b. If distributions are so a apart and no overlap, mean sample will have big difference
    c. Then we can reject the null hypothesis(no difference, both samples from the same distribution) with a great confidence
2. What is power?
    a. Power is the probability that we will correctly reject the null hypothesis
    b. In other words, power is the prob that we will correctly get a small p value
    c. small p value, we have a large amount of power
3. Low power
    a. If 2 distributions are overlap a lot but still we think they came from the different distribution
    b. Most of time, mean sample difference will have a high p value, so we have low power
    c. But, if we increase the # of samples, we can increase the power
    d. That's why we do 'Power analysis' to check how many samples we need to prove

Power analysis

1. Overview
    a. Power analysis determines what sample size will ensure a high probability that we correctly reject the null hypothesis that there is no difference between the two groups
    b. In other words, if we use the sample size recommend by the power analysis, we will know that, regardless of the p-value, we used enough data to make a good decision
    c. If we have power =0.8, then we can correctly reject the null hypothesis by 80% of chance
2. Sample size
    a. When we conduct experiment, we don't compare individual observations.
    b. We compare the aggregations, usually MEAN
    c. sample mean changes by the sample size!
        i. If we sample 1 observation, sample mean has variation and makes it hard to estimate the population mean
        ii. So for 2 distributions, we will not correctly reject the null hypothesis with the mean from 1 single point
    d. But if we have more sample, extreme case will be balanced out, and the sample mean will be closer to the population mean as the # of sample increases
3. Analysis
    a. Commonly we use power = 0.8
    b. alpha = 0.05
    c. effect size: it might differ but we may guess from the pilot test

Week4

Sunday, May 1, 2022 5:03 PM

1. Requirements for Causality
    a. Correlation: X must be associated with Y.
    b. Temporal Precedence: Variation in X must precede the variation in Y (this is actually part of thethird point too).
    c. No Endogeneity: No unobserved confounds, no measurement error, no simultaneity, no sample selection
    d. Step 1: Make sure the two groups are similar
        i. We use T-test for similarities
    e. Step 2: Test if there is a difference in outcomes:
        i. Did the group that got the pill have weight

1. Randomized Experiments
    a. Randomized Experiments (RCTs) are the Gold Standard for causal inference.
    b. The researcher intervenes to manipulate x, at random (e.g., coin toss), such that it is nearly guaranteed not to correlate with e.
        i. This gives you **temporal precedence** (hit X, then observe changes in Y).
    c. By randomly assigning x, you are eliminating any possibility of an **unobserved confound and any possibility of simultaneity.**
        i. This may or **may not solve measurement error**!
        ii. Sample selection may still be a problem (if treatment causes attrition bias, for example).

2. How Do We Know Cov(X,e) = 0? how do we know the randomization worked?
    a. As your sample gets larger, the probability of a random correlation between X and e goes to 0 (if your randomization isn't broken)
    b. You might also provide some suggestive evidence.
        i. What difference would you expect to see in the gender proportion between treatment and control groups? What about the difference in average height?
    c. Randomization Check
        i. Conduct t-test or regression between treatment and independent variables
        ii. t-test for discrete variable and regression for continuous variable
        iii. p-value should be > 0.05 to prove that there is no difference (H0)

```
> t.test(likes~treated,data=MyData)

        Welch Two Sample t-test

data:  likes by treated
t = 0.060292, df = 233.42, p-value = 0.952
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -465366.0  494747.5
sample estimates:
mean in group 0 mean in group 1
        2343120         2328429
```

```
> summary(lm(disc~log(likes),data=MyData))

Call:
lm(formula = disc ~ log(likes), data = MyData)

Residuals:
   Min     1Q Median     3Q    Max
-27.73 -26.82 -17.32  25.80  94.23

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  15.4814    15.5443   0.996    0.320
log(likes)    0.7839     1.1019   0.711    0.478

Residual standard error: 31.93 on 234 degrees of freedom
Multiple R-squared:  0.002158,  Adjusted R-squared:  -0.002106
F-statistic: 0.5061 on 1 and 234 DF,  p-value: 0.4775
```

4. Interpretation

Effect of Magic Pill

VARIABLES	Dependent Variable log(BMI)	
	Magic Pill	With Height and Gender
Magic Pill	-0.1584	-0.1609***
	(0.0370)	(0.0364)
Log(height)		-3.1989
		(1.0218)
gender		-0.2426
		(0.1147)
Constant		
	3.556***	16.60631***
	(0.0251)	(4.1847)
Observations	400	
R2	0.04387	0.08348
Adjusted R2	0.04147	0.07654

   a. Adding log(height) and gender didn't have much effect on the coefficient of magic pill
      i. No confounding effect due to the good randomization that we can check from the above t test

b. How would you interpret this coefficient?
   i. Exp(-0.158) -1 = -0.146
   ii. That is, the magic pill reduces BMI by about 14.6%.

	Dependent variable:	
	leases	
	Binary	Continuous
	(1)	(2)
treated	0.220*	
	(0.118)	
log(disc + 1)		0.055*
		(0.030)
Constant	0.508***	0.512***
	(0.084)	(0.083)
Observations	236	236
R2	0.015	0.014
Adjusted R2	0.010	0.010
Residual Std. Error (df = 234)	0.908	0.908
F Statistic (df = 1; 234)	3.476*	3.401*
Note:	*p<0.1; **p<0.05; ***p<0.01	

a. How do you interpret the coefficient?
   i. log(disc+1) coeff: 0.055. So if discount is increased by 100%(1), leases increase by exp(0.055) -1 = 0.055
   ii. A 10% increase in discount --> 0.005 increase in rentals.

	Dependent variable:
	leases
	Like Moderator De-Meaned
treated	0.221*
	(0.117)
likes_demean	0.00000**
	(0.00000)
treatedTRUE:likes_demean	-0.00000*
	(0.00000)
Constant	0.508***
	(0.083)
Observations	236
R2	0.042
Adjusted R2	0.029
Residual Std. Error	0.899 (df = 232)
F Statistic	3.356** (df = 3; 232)
Note:	*p<0.1; **p<0.05; ***p<0.01

a. How Do We Explore Heterogeneity?
    a. Add interaction term
    b. We do find some evidence that likes attenuates the treatment

5. How Many Subjects Do We Need?
    a. What does it mean for an experiment to be underpowered?
        i. This refers to having a sample that is too small!
        ii. If your sample lacks power, it means it cannot reliably detect effects.

    b. Does this mean that a small sample leads to a “conservative” evaluation of the treatment effect?
        i. No! This is a common misconception.
        ii. If your sample gets really small, your chance of discovering effects that aren’t really there goes up!
        iii. Intuition: random noise in the data leads to more chance differences in small samples.
        iv. If you toss coins only 5 times and decide if the coin is biased or not -> you can get

Type1 or Type2 error easily. But not being convervative

c. When do we need more data
   i. High variance in the data, increase sample size
      1) $\sigma\uparrow \Rightarrow N\uparrow$
   ii. Want higher accuracy that we are picking up the true effect
      1) $(\alpha, 1-B\uparrow \Rightarrow N\uparrow)$
   iii. If we expect a higher difference between the two groups
   iv. $(\text{delta}\downarrow \Rightarrow N\uparrow)$

a. To Determine Sample Size Requirements
   i. We need a significance threshold, alpha (usually 0.05).
   ii. We need a null hypothesis (often equivalence: Yt – Yc = 0).
   iii. We need an assumed power level, beta (usually 0.80).
   iv. We need a test statistic, e.g., a t-stat.
   v. We need a sample size, N.

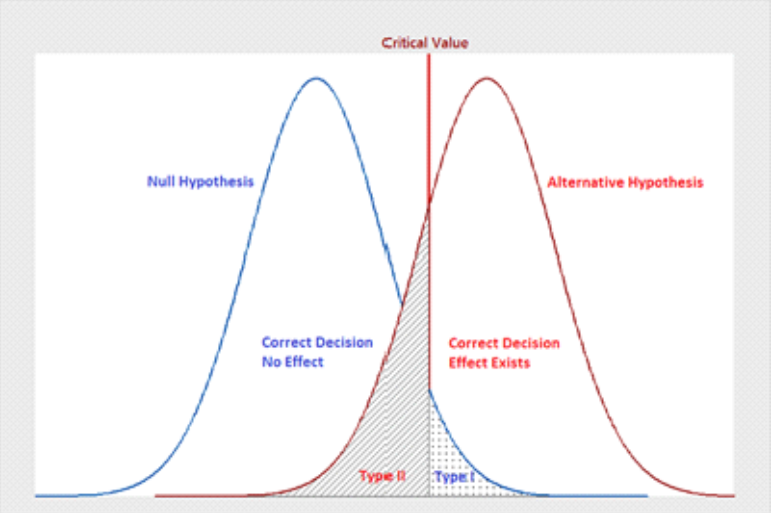
b. What is the Power Level?
   i. Type II Error: failing to reject the null when you should.
   ii. Power is 1 – Pr(Type II Error). That is, when you have more power, your probability of failing to detect a true effect is smaller.
   iii. How to calculate?
      1) If Management wants to detect a 20% increase over the average when price goes up 50 cents with 90% confidence.
      2) That means we need to detect a difference of 0.1 (0.5 * 0.2) = 0.1 alpha of 0.1.
      3) Let's assume a power of 0.8, as we typically due.
      4)

```
> power_t_test(delta=0.1,sig.level = 0.1, power=0.80, type=c("two.
sample"),alternative=c("two.sided"),n=NULL)

     Two-sample t test power calculation

              n = 1237.188
          delta = 0.1
             sd = 1
      sig.level = 0.1
          power = 0.8
    alternative = two.sided

NOTE: n is number in *each* group
```



e. Relationship Between Alpha, Beta and N?
   i. Type I Errors Decline.
   ii. Type II Errors Decline.
   iii. You can detect smaller effects. (delta)

f. Other Concerns
   i. Interference
      1) If there is interference, it means there are treatment spillovers between subjects in the two groups. It might change subject behavior. This kills your ability to recover accurate causal estimates
   ii. Valid Controls
      1) 1st one there was a additional step and second there was no step BEFORE the check out
      2) So we can't compare both group. they should have put fake step instead and remove the last step
   iii. Response Biases
      1) Are subjects aware of the experiment
         a) Social Desirability Bias : They may conceal behaviors they think are undesirable, or lie about behavior.
         b) Demand Effects Doing what they think the experimenter would want them to do.

g. Full-Factorial Design
   i. If we have two alternative treatments, a full factorial design considers the 2x2 matrix of all permutations.
   ii. This allows you to understand not just the independent effects of these treatments, but how they combine with one another.
   iii. That is, does having one treatment amplify the effect of the other, or attenuate it?
   iv. This is Called an Interaction Effect

h. On the Fly vs. Block Randomization
   i. If we are randomizing in advance, and not on the fly, we typically first “block” on features that would typically be considered confounds
   ii. Alternatively, we can randomize and check balance on those features. Rerandomize until the treatment is orthogonal.

Week5

Sunday, May 1, 2022 7:01 PM

Interpreting Logistic Regression Coefficients

1. What is the logistic regression
    a. LHS is essentially the log odds.
    b. $$log\frac{p}{1-p} = \beta_0 + \beta_1 x \ \ldots$$
2. How to interpret
    a. For discrete IV what is $\beta 1$?
        i. how much more likely are is a student to be in the honor class if you are female?
        ii. If B1 = 0.59, then exp(0.59) = 1.8, so odds is 80% higher
        iii. Odds for women are ~ 80% higher than for men
        iv. using p/(1-p), we can convert it to probability
    b. For continuous IV, what is B1?
        i. if B1 = 0.156, exp(0.156) = 1.169,
        ii. Each additional point increase in math increases odds of honor class by 17%
        iii. can't convert to probability

Week6

Monday, May 2, 2022 8:46 AM

How can we infer causality when we can't run experiments?

1. Why we can't control all confounders?
    a. The number of confounds might also be large relative to the number of observations we have (= low statistical power).
    b. This also assumes that income and GPA impact years in school, as well as Scholarship, linearly. What if they do not? Perhaps there is an income cutoff, for example (model dependency). so model not be linear relationship

2. PSM
    a. We want to mimic the experiment in observation data by creating group that similar but different in treatment
    b. Intuition: Find matched pairs that are as similar on observed variables, so that the treatment can
    be assumed to be random
    c. If no match is found, throw away the data

3. Specific Matching Methods
    a. Exact Matching: Typically not feasible as you need to match on many variables
        i. treatment and control group that are identical (except for that one got the scholarship)
    b. Propensity score matching: match based on how likely subjects were to receive treatment
    c. Coarsened exact matching: Match based on discretized observed variables

4. PSM Step-by-Step
    a. Model treatment as a function of observed variables
        i. (Use Logit/Probit to do this)
    b. Get predicted probabilities of treatment from the model
    c. For each treated observation, find control observation with closest propensity score (up to a specified threshold aka caliper)
    d. If no match found, remove observation from analysis

5. Benefits of Matching
    a. Forces Attention to Covariate Imbalance
        i. Forces researcher to evaluate covariate balance before estimating effects.
        ii. Conceptually same as "randomization checks" in experiments
    b. Provides a Way to Recover "As-Good-As" Randomization, Post Hoc
        i. If my randomization breaks (or when using an observational dataset) I can construct an experimental setup after the fact.
        ii. Recreate 2 groups which is as similar as each other

6. Some Issues of PSM
    a. Omitted Variable Bias
        i. Still don't know what we don't know. PSM assumes that treatment is determined by observed variables
        ii. Solutions: there is no way to evaluate OVB
    b. Lack of Support
        i. There may not be any good matches. Choice of caliper is somewhat arbitrary
        ii. Solutions: Evaluate Quality of Matches – e.g., use a caliper, visually check matching quality
    c. Propensity Function
        i. Logit or Probit is just as arbitrary as conditioning on the controls linearly! There might be hidden confunding factors
        ii. Solutions: Use Something More Flexible – Nothing prevents you from implementing the best flexible predictive model you can come up with (e.g., a deep net). Alternatively, other matching algorithms e.g., genetic matching, CEM (more on this one shortly).

7. Coarsened Exact Matching (CEM)
    a. Motivation: collapsing all variables into a single propensity score seems too arbitrary, but directly matching on all variables is too hard

b. Instead, we match on discretized versions of all variables
    i. Intuition: Discretize each variable into several categories. Then for each treatment observation, find control variables that match exactly on all discretized variables
    ii. Theory of CEM shows that it can clearly improve covariate balance
c. Roughly speaking, coarsened exact matching works by rst discretizing the variables on which matching is performed, i.e., fX1; : : : ; Xkg. For each Xi that is not already categorical, it is discretized using user-defined cutpoints (or \bins"). Then, a treated unit is matched with control units having **exact** values on the coarsened matching variables. The coarsening step is crucial for this method to be feasible, as it is generally hard to nd exact matches, particularly when variables are continuous

8. Sensitivity Analysis
    a. Is it generally a good idea to explore robustness of the result to alternate matching techniques (PSM & CEM)
        i. Try different propensity functions (logit vs. probit vs. something else)
        ii. Try different parameters for PSM
        iii. Different values of caliper
        iv. Number of nearest neighbor
        v. Matching with replacement vs. without replacement.
        vi. If results and conclusions are consistent across different specifications, they are robust and trustworthy
    b. Matching Can be Combined with Other Methods
        i. pairing it with difference-in-differences, for example

Keep in mind that these are valid under assumption only!

Assumptions of Methods

IMPORTANT: Observational methods bear stronger assumptions, which often cannot be tested. As such, the assumptions should always be stated explicitly!

Design	When to use	Advantages	Disadvantages
Randomized Experiment	Whenever possible	Gold standard, most powerful.	Not always feasible, ethical or legal.
Matching Techniques	When all other methods are not possible (sometimes in tandem with other methods).	Resolves confounded differences between treatment and control (aside from treatment), post-hoc. Avoids overfitting, model dependence.	Assumes no unobserved differences (often implausible assumption, and it's not testable).
Regression Discontinuity	If an intervention has a clear, sharp (arbitrary) assignment rule	Usually easy to argue and show descriptive evidence that the discontinuity is there and having an effect	Can only look at a local sub-group of the overall sample
Difference-in-Differences	If two groups are growing at similar rates before one is "treated" Baseline and follow-up data are available	Eliminates any fixed differences not related to the treatment	Can be biased if trends are not parallel before the "treatment" Ideally have 2 pre-intervention periods of data
Synthetic Control Method	If you have multiple groups observed over time, and one group is treated while others are not. If DiD parallel trend assumptions are violated.	Does not require the strict parallel trends assumption; just requires that treated group's pre-trend can be modeled well as a weighted average of control pre-trends.	Only recovers Average Treatment on the Treated (not ATE); estimates may not generalize to other units. Statistical significance hard to evaluate.
Instrumental Variables	If you can identify an exogenous variable that strongly influences the RHS variable of interest, and does not impact the final outcome of interest (except via the RHS variable).	Resolves omitted variables, simultaneity and measurement error all at once.	Good instruments are hard to find. While "strong influence" can be evaluated, the "backdoor criterion" can not be tested empirically. Has to be argued.

Week7

Monday, May 2, 2022 1:00 PM

1. Cross Section vs. Panel Data
    a. One observation per unit vs. Repeated observation per unit
    b. Typically you are observing a person or some other entity over time, t, at fixed intervals.
        i. However, this might not be the case, e.g., you might be observing a company's performance across locations.
    c. You have panel data when there are multiple observations for each entity you care about
    d. 1 observation vs Multiple observation per time

2. Time Invariant Confound
    a. Endogeneity in Panel Data
        i. Endogeneity problems exist for panel data
        ii. If we don't include confounders, we have an omitted variable bias
        iii. It turns out we can deal with time-invariant endogeneity problem (even if we don't observe it) via fixed-e ect regressions

    b. Let's examine the case of time invariant confound
        i. These can be written as confound i (as it does not vary over time)
        ii. What kind of confounds are time invariant?
        iii. Intuition: inherent / stable characteristics of the unit/individual

3. Two Solutions for Time-Invariant Confounds
    a. Fixed Effect Regression:
        i. Treat confound i as a "fixed" parameter for each individual
        ii. Intuitively: Each individual now has a unique intercept ($\beta 0+\beta 2$confound i)
            1) In FE, confound i may or may not be correlated with Xit
            2) A fixed-ed effect does not make assumption on the correlation between confoundi and Xit. they may be correlated or uncorrelated
    b. Random Effect Regression
        i. Treat confound i as a random variable
        ii. Intuitively: confound i is a part of the model error term (confound $i+\epsilon it$)
            1) confound i is assumed to be uncorrelated with Xit
            2) This is confusing! If we assume $corr(Xit,$confound $it) = 0$, does it even pose a endogeneity problem? More on this later.

4. Method of FE
    a. Method 1: Demean data to remove confound over time
        i. Also known as "within transformation"
        ii. Remember the population model : $Yit=\beta 0+\beta 1Xit+\beta 2$*confound $i+\epsilon it$
        iii. In fixed effect regression, we take unit wise means:

- In fixed effect regression, we take unit wise means:
  - $\overline{Y_i} = \beta_0 + \beta_1 \overline{X_i} + \beta_2 \overline{confound_i} + \overline{\epsilon_i}$ (1)

WITHOUT confound variable.

- And subtract it
- $Y_{it} - \overline{Y_i} = (\beta_0 - \beta_0) + \beta_1 (X_{it} - \overline{X_i}) + \beta_2 (Confound_i - \overline{confound_i}) + \epsilon_{it} - \overline{\epsilon_i}$
- $Y_{it} - \overline{Y_i} = \beta_1 (X_{it} - \overline{X_i}) + \epsilon_{it} - \overline{\epsilon_i}$

Removes all time invarying confounders

Remove OVB (due to time invariant confounde

- Now you have a new dependent and independent variables
  - Parameter (β_1) does not change

b. Method 2: Add a dummy variables
   i. Include dummy variable for each individual i.
   ii. No need to demean the equation
   iii. The dummy captures time invariant effect of each individual i.
   iv. This is mathematically equivalent to within transformation (You will get the same estimates)
   v. Note: You will need to omit one dummy as the "reference group" if your model has an intercept term
   vi. Method 1 and Method 2 has the same output of coefficient

c. Method 3: First difference
   i. Subtract the lag term for each variable from itself
   ii. $Yit - Yi(t-1) = \beta 1 (Xit - Xi(t-1)) + \epsilon' it$
   iii. This can give different results from Method 1 & 2

5. Issues with Fixed Effects Regression
   a. What if your causal effect of interest is time invariant?
      i. If you use within transformation, it will be demeaned and cannot be estimated
         1) Y-Y_bar become 0 because it does not change over time
      ii. If you use dummy variable approach, it will be "absorbed" in the coefficients of the dummy variables
      iii. Either way, you will not be able to reliably estimate the time-invariant causal effects
   b. What if you are worried about time varying confounds?
      i. Panel data regressions do not directly resolve time-varying confounders
      ii. Need other methods

6. Fixed Effect or Random Effect?
   a. It depends on assumption
      i. If the time-invariant confound is truly uncorrelated with Xit (that is the assumption of random effect regression), the random effect regression is more efficient than fixed effects
      ii. I.E. random effect regression are more precise than fixed effects regression
      iii. However, if the assumption fails, then random effects regression will give incorrect estimates
   b. How do we decide?
      i. Run both fixed effect and random effect regressions
      ii. Apply the Hausman Test. If test is rejected, use fixed effect regression

iii. Otherwise, use random effect regression
iv. It is hard to meet this assumption so RE is not used much

One final point:

- Consider the population model:
  - $Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 Confound_i + \epsilon_{it}$
  - β_1 interpretation: How much does Y change if X increases by 1 unit

- In fixed effect regression, we just take means
  - $\overline{Y_i} = \beta_0 + \beta_1 \overline{X_i} + \beta_2 \overline{confound_i} + \overline{\epsilon_i}$ (1)
  - β_1 interpretation: How much does Y change if X increases by 1 unit

- And subtract
  - $Y_{it} - \overline{Y_i} = \beta_1 (X_{it} - \overline{X_i}) + \epsilon_{it} - \overline{\epsilon_i}$
  - β_1 interpretation in (1): How much does Y_{it} change if X_{it} increases by 1 unit

What is fixed effects and random effects?

Use demeaning to delete the confounding effect

- Fixed effects:
  - $Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 confound_i + \epsilon_{it}$
  - Intuitively: Each individual now has a unique intercept $(\beta_0 + \beta_2 confound_i)$

- Random effects:
  - Treat $Confound_i$ as a random variable
  - Assumption: $Confound_i$ is assumed to be uncorrelated with X_{it}

Assumption can be selected by the result of Hausman test!

Week8

Monday, May 2, 2022 4:14 PM

DID

1. Two assumptions
    a. Parallel Trends
        i. Because, the control group needs to be a good counterfactual for the treated group!
        ii. Without the treatment, treated subjects would have continued in parallel with the control

We can generalize further..

- We have 14 weeks
- $Log(1 + viewtime_{it}) = \beta_0 + \beta_{11}Week_t^2 + \beta_{12}Week_t^3 + \cdots + \beta_{113}Week_t^{14}$

$$+\beta_2 Premium_i +$$

$$\beta_{31}Premium_i * week_t^2 + \beta_{32}Premium_i * week_t^3 + \cdots + \beta_{313}Premium_i * week_t^{14}$$

        i. Premium and the control groups should have similar trends up to week 7 (parallel trends assumption)
            1) Look at coefficients capturing the difference between treatment and control groups before week 7
            2) Before the treatment, which is week7, coefficient of premium should be 0 After week7

    a. NO INTERFERENCE!
        i. A firm in the control does not use the app
        ii. This is the same assumption we made in experiments!
        iii. This is also known as the Stable Unit Treatment Value Assumption (SUTVA)

Suppose we have data for 3 weeks in the TSTC data:

- $Log(1 + viewtime_{it}) = \beta_0 + \beta_{11}Week_t^2 + \beta_{12}Week_t^3$

$$+\beta_2 Premium_i +$$

$$\beta_{31}Premium_i * week_t^2 + \beta_{32}Premium_i * week_t^3$$

Changes in TV viewing in week 2 & 3
For premium customers

- *Why have we omitted week 1?*
    - *Similar to the previous slide, it serves as a comparison group*

DID example

1. A DiD model can be estimated with OLS
   a. Create $After\ t$ (binary) that takes value 1 if the observation is after the treatment
   b. Create Treat i (binary) that takes value 1 if the observation belongs to the treatment group
   c. Create an interaction term $After\ t$* Treat i that takes value 1 if the observation is in the treatment group AND if the observation is after the time of the treatment
   d. Estimate the following regression: Employmentit=B0+B1*$After\ t$+B2*Treat+B3* $After\ t$* Treat+ϵit
   e. Set $After = 0$ and $\text{Treat} = 0$. Using the model above, what does that represent (NJ vs. PA/pre vs post treatment)?
   f. Does the result in the OLS line up with the averages you calculated? Cycle through $A\ \ t\ \ r = 1$ and Treat = 0 and so on…
      i. How do you interpret B3? It is the DID.
2. Isn't $Prmium\ i$ and $After$ just fixed effects for group and time?
   a. Yes, that is one way to look at
   b. The Premium dummy absorbs everything that is systematically different about the treatment group
   c. The After dummy absorbs everything that is systematically different about the post-period
   d. In fact, we can replace Premium with a set of customer dummies and After with a set of week dummies (and run a fixed effects regression)

3. How Do We Evaluate Parallel Trends?
   a. Run a "Dynamic" DiD model:
   b. Interact the "treatment" variable, i.e., Premium, with week dummies.
   c. Yit=B0+B1$Premium\ i$+$\alpha Week\ t$+$\lambda Premium\ i$*$Week\ t$+ϵit
      i. λ will tell how us the treatment varies over different week (relative to the omitted week)
      ii. Intuition: we are estimating many difference-in-differences regressions all at once! The omitted week is our "pre" period now, and we estimate diff-in-diff relative to every other period, comparing treatment group with control
   d. Instead of a single Aftert variable, we use week dummies. Note that we omit week 1 to avoid dummy variable trap

The model to be estimated is:

$$Y_{it} = \beta_0 + \beta_1 Treat_i + \sum_{t=2}^{T} \alpha_t week_t + \sum_{t=2}^{T} \lambda_t Treat_i \times week_t + \Gamma X_{it} + \varepsilon_{it}$$

To understand this model, suppose you look at week 2 (i.e., $week_2 = 1$ and $\forall t \neq 2$, $week_t = 0$), the regression becomes:

e.

$$Y_{it} = \beta_0 + \beta_1 Treat_i + \alpha_2 week_2 + \lambda_2 Treat_i \times week_2 + \Gamma X_{it} + \varepsilon_{it}$$

which is just a standard DiD model, with dummy variable $week_2$ taking the role of $After_t$. As a result, λ_2 estimates the difference-in-difference effect of the treatment, **during week 2**. In general, λ_t estimates the difference-in-difference effect of the treatment, **during week** t. By looking at the trend of λ_t before the treatment date, we can empirically test the parallel trend assumption. By looking at the trend of λ_t after the treatment date, we can understand how treatment changes over time.

Issue of DID

1. Anticipation Effects
    a. Sometimes People See the Natural Experiment Coming
    b. What are some examples of where this could happen? A new feature roll-out – users may login more often to see if it's available yet, for example.

Robustness Checks

1. Placebo test
    a. Artificially move the treatment a few periods earlier than it really was.
    b. Re-estimate the DiD regression using only pre-period data.
    c. What would you expect the result to be?
        i. Nothing! Insignificance…

Week9

Monday, May 2, 2022 5:35 PM

1. Key point
    a. Overview of the Synthetic Control Method
    b. When it makes sense to use, what assumptions it makes.
    c. Pros and cons relative to Difference-in-Differences
    d. Generalizing from Single Treatment to Many Treatments

2. General Problem
    a. Pre-treatment Trends are Violated
    b. Our set of controls is not useful for DiD, so we cannot just trust the DiD estimate of the treatment effect.
    c. Matching away pre-treatment differences, and then estimating DiD?
        i. Don’t do this; it leads to a biased estimate (Chabe-Ferret 2017)
    d. Synthetic control is variation of DID. It is used when parallel trend assumption is violated(it does not need parallel trend assumption)

3. Synthesize from a combination of TS(Time series) and CS(Cross section) data.
    a. Fit a predictive model between treated Y and control Y’s in the pre period.
    b. Assuming we do a good job in pre, we can carry that model forward in time!
    c. The inputs to the predictions will still embed shocks common to the market in the post period
    d. This is about synthesize the counterfactual well. we want to see the difference between the treatment but what would you do if we can't find the counterfactual?
        i. We make the counterfactual

4. Assumptions
    a. What We Need to be True
        i. Need Repeated Observations of Treated and Control, in the Presence and Absence of Treatment.
            1) This is the same as with Difference-in-Differences.
        ii. This might mean pre-period time series data. Or, I might synthesize a counterfactual in some treated locations, using data from untreated locations.
    b. SUTVA Still Applies: No Interference
        i. We must not have any interference between treated and control units, else the counterfactual will be wrong.
        ii. That is, the predictive model will cease to be meaningful in the post period.

5. How to synthesize?
    a. Using the pretreatment data and find the weight to make the parallel line of synthetic group
    b. Apply weight to post treatment of synthetic group
    c. find the difference between synthetic group vs California
    d. pretreatment is training data and post treatment is test data

6. ATE ATT

a. What did we compare CA's cigarette consumption with?
    i. CA's expected cigarette consumption in the absence of the policy
b. This is known as Average Treatment Effect on the Treated (ATT)
    i. In Case-Control Policy Analysis, we are the government of California, and we made a policy change, and we want to quantify the effect of that change on Californians.
    ii. So do not care about making a generalizable inference

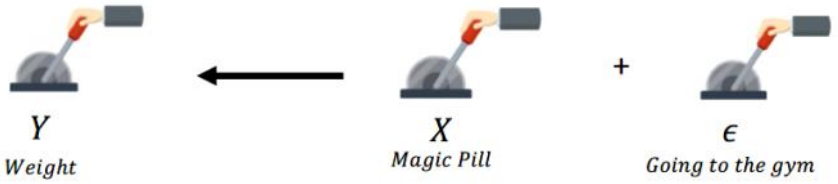
c. Unlike the wage difference between NJ vs PA = ATE, in synthetic control is called ATT

Week10

Monday, May 2, 2022 6:12 PM

An Analogy

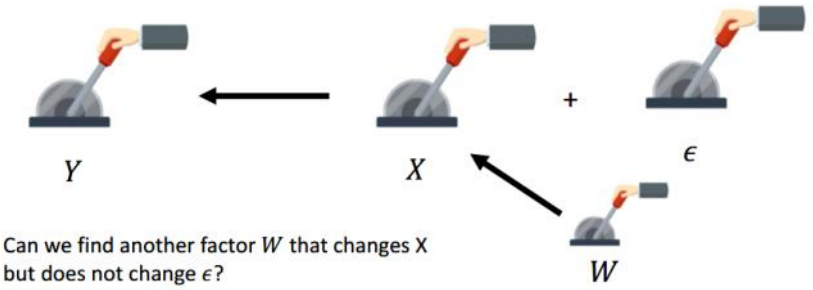
What we have been doing so far:



Example: Going to the gym is an unobserved factor (and hence captured in ϵ)

1. What if X & ϵ are correlated?
    a. In this case, when X moves ϵ will move as well
    b. We will have omitted variable bias if X and ϵ are correlated
    c. Classic case of endogenous factor (& omitted variable bias)
    d. Ideally, we want to move X without ϵ changing at all

Concept behind Instrumental Variable



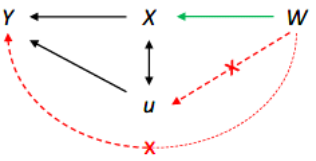
2. Two requirements
    a. W and X should be correlated
        i. The instrument should be correlated with the independent variable
        ii. This is known as the relevance criterion
    b. W and e should not be correlated
        i. This is known as the exclusion criterion
        ii. While the relevance criterion can be tested, the exclusion criterion **cannot** be tested
    c. Note: IV overcomes correlation from other sources of endogeneity in X
        i. Simultaneity
        ii. Measurement error
        iii. Omitted variable bias)

3. Mathematically
    a. You want to estimate the effect of X on Y (equation).
    b. But, you believe X is also correlated with the error term i.e., it has unobserved confounders, is mis-measured, or y also influences X

$$Y = \beta_0 + \beta_1 X + \epsilon$$

c. If we have a W that is correlated with X but NOT u or Y,
   i. We can regress X onto W
   ii. Then use predicted values ofX in our final regression.

$$Y = \beta_0 + \beta_1 X + \epsilon$$



   i. In Angrist (1990)
      1) W is the draft pick number
      2) X is ifthe individual served or not
      3) Y is annual income
      4) u are the unobserved factors (omitted variables)

4. Some examples:
   a. Taxes and Smoking
      i. We want to estimate the extent to which mother's smoking status impacts birth weight
         1) Potential Omitted Variables such as Mother's income perhaps
      ii. Use the tax on cigarettes as an instrument
         1) Tax should be correlated with $M \quad h \quad rS \quad k \quad ri$ (Relevance)
         2) Tax should not be correlated with omitted variables (exclusion)

   b. Rainfall and Supply
      i. We want to estimate the extent to which coffee supply impacts price
      ii. Potential simultaneity
         1) Supply -> Price but Price also -> Supply
      iii. What are the factors that impact the quantity ofcoffee produced?
         1) Perhaps rainfall in the growing areas -> Quantity ofcoffee grown (Relevance)
         2) Rainfall in the growing areas should not impact the price that is paid (Exclusion)

   c. Education on individual earnings
      i. Cannot regression Earnings ~ Education
         1) Omitted variables (personal attributes) could be correlated with Education and earnings
         2) The quarter the child was born is correlated with the amount of schooling
            a) Uncorrelated with other factors
            b) Use Quarter of Birth as an instrument for years of schooling

5. Methodology: 2SLS: Two-Stage Least Squares
   a. Regress X2 onto its instrument(s), W, and recover predicted values of X2 from the first stage
      i. i.e., decompose X2 into part explained by W, and "the rest," which includes the bad part.
   b. Regress Y onto predicted values of X2. (X1 is just another variable)

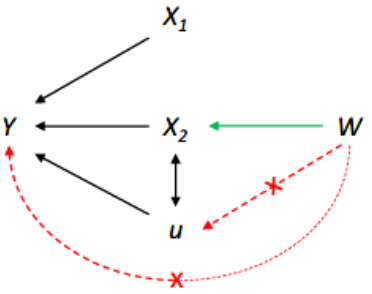
which includes the bad part.

2. Regress Y onto predicted values of X_2.

$$X_2 = \pi_0 + \pi_1 \cdot W + \epsilon$$

$$Y = \alpha + \beta_1 \cdot X_1 + \beta_2 \cdot \hat{X}_2 + u$$

- Note: this gives you a consistent estimate



c. Some Terminology
   i. The effect ofyour treatment is "just-identified" ifyou have one instrument for it. Generally, ifthe number ofinstruments = number of endogenous regressors, the equation is just identified. A just-identified 2SLS is also called the IV estimator.
   ii. A 2SLS regression is "over-identified" ifyou have more instruments than endogenous regressors.
   iii. The regression is called "under-identified" ifyou do not have sufficient instruments for the number of endogenous regressors.
   iv. General Truths: more instruments is better (up to a point), stronger instruments are better.

6. Evaluating Instrument Relevance
   a. Relevance
      i. i.e., corr (Wi , Xi) ≠ 0
      ii. Relevant instruments are highly correlated with the endogenous regressors even after controlling for the exogenous regressors. This requirement **can be empirically tested** using first stage regression model fit statistics.
         1) $Military\ i$ is endogenous
            a) Omitted variables are correlated with $Military$ and $Income\ i$
            b) Perhaps individual is from an area where wages are low and lots of people join the military
      iii. Stock and Watson's rule of thumb: The first-stage F-statistic testing the hypothesis that the coefficients on the instruments are jointly zero should be at least 10 (for a single endogenous regressor).
   b. Exclusion
      i. i.e., corr (Wi , ei) = 0
      ii. Excluded instruments are uncorrelated with the error term. This requirement needs a strong theoretical argument and can, in general, **not be tested empirically**. The theoretical argument has to be convincing.
         1) First, describe how the instrument conceptually influences the endogenous regressor.
         2) Second, rule out any direct effect ofthe instrument on the dependent variable or any effect on confounders.
         3) Rule out any reverse effect ofthe dependent variable on the instrument, i.e., Y cannot influence W.

7. Test statistics
   a. Weak IV Test: This is typically based on first-stage F-stat. Significance implies we reject the null of weak instruments.
      i. Null: Weak IV
   b. Wu-Hausman Test: The null hypothesis is that IV yields equivalent estimates to OLS (and thus we should not use IV because we are losing power by doing so, and getting wider standard errors than necessary). Significance implies we reject the null of equivalence (and thus we should use IV).
      i. Null: IV has no difference with OLS
   c. Sargan Test: If we have more instruments than endogenous variables, the Sargan test evaluates equivalence of estimates between using all vs. a subset. Significance implies we reject the null of equivalence (and thus at least one instrument is invalid).
      i. The test does not tell us which instrument is the problem, however.

8. Does the 2SLS method give us ATE or ATT?
    a. Turns out it gives us LATE (local average treatment effect)
        i. LATE is the treatment effect for a subpopulation
    b. Think back to our example on serving in the military:
        i. 1st stage: X(Serving in the military) ~ Z(Draft Number)
        ii. 2nd stage: Y(Income) ~β_hat*X(Serving in the military due to draft number)
        iii. β_hat is the ATE **for the people serving in the military due to their draft number only**
            1) IE it is the effect for the subpopulation (therefore LATE)

Week 11

Tuesday, May 3, 2022 10:41 AM

Regression Discontinuity

1. Overview
   a. When it makes sense to use, what assumptions it makes.
   b. Pros and cons relative to other methods.
   c. Evaluating assumptions.
      i. RDD: Effect ofincumbency on winning a US election.
      ii. FRDD: Simulated.
   d. Robustness Checks
      i. Descriptive plots, bandwidth sensitivity

2. Thought Experiment
   a. Does getting a scholarship lead students to go to graduate school?
      i. Using observed data:
         1) Gradschool i=B0+B1Scholarship i+ϵi
         2) Aptitude could be an omitted variable
      ii. Ideal experiment: Randomly give scholarships to some students
         1) In experiment: correlation(Aptitude, Scholarship) = 0
   b. But in real world, you can't give scholarship randomly.
      i. So we need to use other methods we don't want other OVB to affect on X

3. How to remove endogeneity?
   a. Score on PSAT determines if you obtain scholarship
      i. National Merit Scholarship: Students who got above 219 got the scholarship
         1) Let's think about a student who got 218 and one that got 220
         2) Just a difference of2 points
            a) Likely to have similar aptitude
            b) IE corr(aptitude, scholarship) = 0
         3) In this narrow window, obtaining scholarship is a random assignment (similar to an experiment)
      ii. By choosing the narrow window, we can safely assume that omitted variables has low effect on X and Y
   b. What have we done here?
      i. We have a strong cut off(PSAT Score = 219)
      ii. Right around the threshold, students are pretty much the same
         1) But due to a strict cut off, some students get scholarship
         2) Other students do not get scholarship
      iii. Around the threshold:
         1) We can assume random assignment of treatment
         2) Treatment is uncorrelated with unobserved factors

Representation of PSAT Score and Graduation Rates

it doesn't need to be linear only

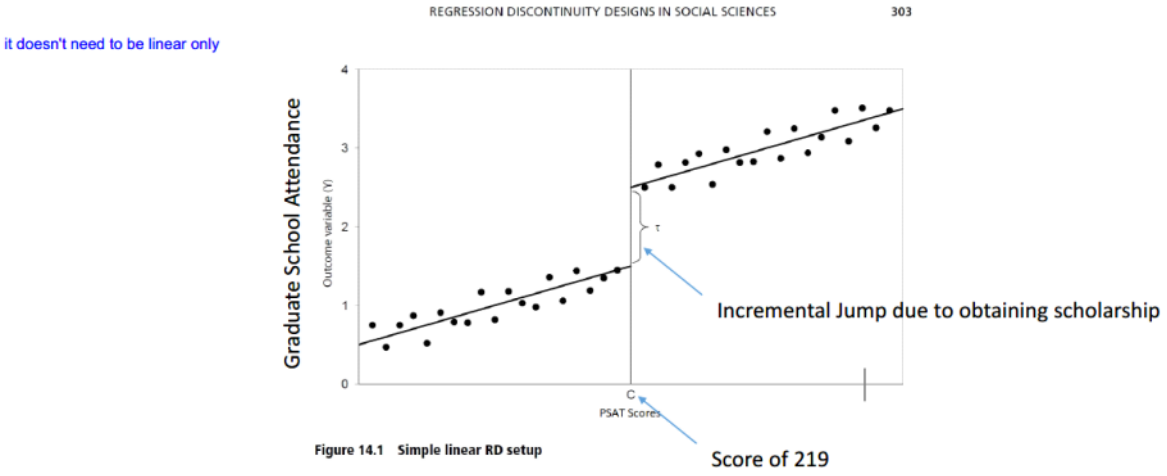


Figure 14.1 Simple linear RD setup

4. Regression Discontinuity Designs (RDDs)
    a. Exploit policies based on arbitrary cutoffs
        i. Utilize an assignment or forcing variable (test score)
        ii. Threshold defined on the assignment variable

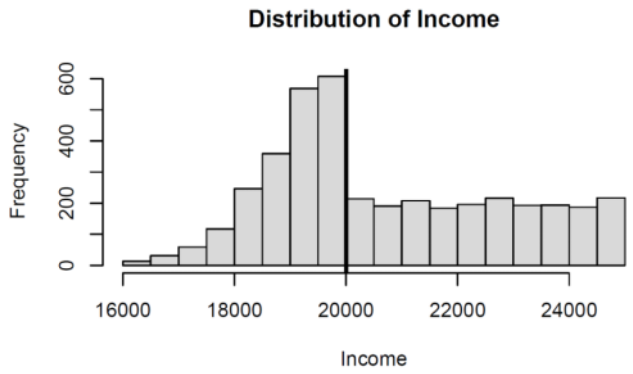
    b. Take advantage of natural "jumps" (discontinuities)
        i. Jumps result in assignment to a treatment
        ii. Provides 'local' randomness in treatment assignment
            1) When you compare units (people, cities, firms, countries, ...) that barely qualify against units that barely do not.
        iii. In absence of treatment at the cut off
            1) $Yi=\alpha+B*Xi+\epsilon i$
            2) Y linearly increasing in X
        iv. With treatment
            1) $Yi=\alpha+B1*Zi+B2*Xi+\epsilon i$
            2) Y linearly increasing in X
            3) For items above cut off, additional lift due to treatment(captured by Zi)

    c. We can generalize this to:$Yi=\alpha+B1Zi+B2(Xi-Xc)+\epsilon i$
        i. Model the Dummy Z (Above Threshold), as well as Distance from Threshold.
            1) Here, Z is the threshold dummy (1 ifabove, 0 ifbelow), X is the assignment variable, and Xc is the cutoff value of X that defines the threshold (Note: Xi – Xc captures 'distance' from the cutoff).
            2) Interpretation: Coefficient on Z captures mean difference in Y between treated and untreated when Xi - Xc = 0, i.e., right at the threshold!

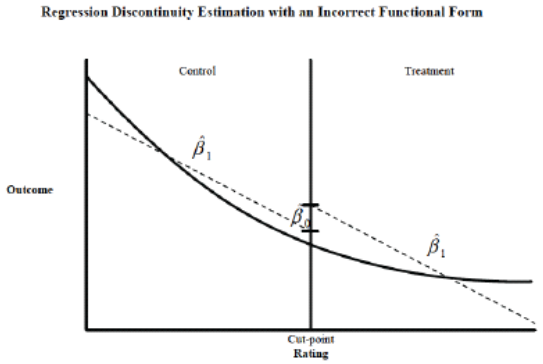
    d. Bandwidth Choice
        i. Choosing What Data to Enter Into Your Regression
        ii. We need to select a bandwidth, h, which is the range around the cutoff point of X that we consider. Lowering h yields a more trustworthy
            1) estimate of the treatment (less bias), but it comes at the cost ofpower (wider standard errors).
        iii. Good News: Algorithms proposed in recent years for optimally selecting the bandwidth (and R implements them).

5. Example
    a. US House Elections
        i. incumbency advantage: If politician wins an election, more like to stay in office
    b. Research question:
        i. If politician wins an election, are they likely to increase their vote share in the subsequent election?
        ii. $VoteShare\ t+1=B0+B1WinningElectionn\ t+\epsilon$
    c. Do we have a strong cut off?
        i. Yes, ifa politician gets more than 50% ofthe votes, (s)he wins

6. Assumptions
    a. Non-Arbitrary Assignment (Self-Selection):
        i. Suppose you want to examine the impact ofa job training program
            1) The program is offered to people making less than $20,000/year
            2) The program is advertised for a year and then offered
            3) RDD seems reasonable (arbitrary strict cut off)
        ii. People may just reduce their income to qualify for the program
        iii. They are now self selecting into the program
            1) Violates assumption that there is a random assignment oftreatment to subjects around the threshold
            2) ppl who earn 20,100 will try to reduce the income so that they can get the chance to be in job traning program.
            3) This is not random assignment
        iv. just below 20000 has higher number compare to all other bin. This means ppl self selected to get the job traning chance and can't use RDD



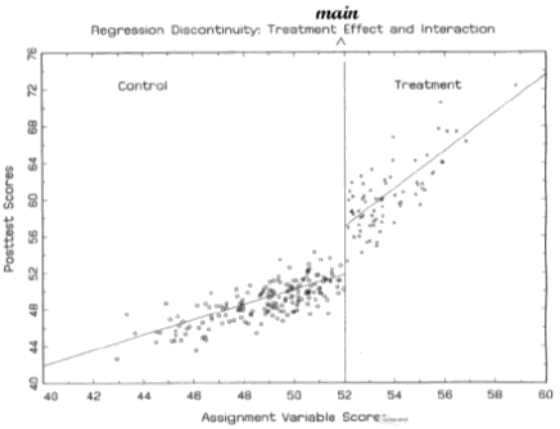
b. Model Mis-specification:
   i. What if Yi and Xi have a quadratic relationship?
      1) Even though the actual relationship has no jump, if we apply linear RDD, then it incorrectly shows jump
      2) So, it is important to identify the functional form to use RDD.
      3) Model misspecification can influence our estimate of that treatment effect!



   ii. treatment not only causes a jump, it also causes the slope associated with (Xi - Xc) to change?
      1) To account for different slopes on the two sides ofthreshold, add an interaction term between Z and $(X-Xc)$.
      2) notice that the regression line to the left of threshold has slope B2, whereas the regression line to the right of threshold has slope B2 + B3.

$$Y_i = \beta_0 + \beta_1 Z_i + \beta_2(X_i - X_c) + \beta_3 Z_i (X_i - X_c) + \varepsilon_i$$

thought the slope is different after cutoff, we can add interaction to add difference



   iii. How to Get at Model Specification?
      1) Use Data from Before Policy was Implemented!
         a) Try different regression specifications using pre-policy implementation data, relating Y to X.

b) Choose a model with best fit statistics (e.g., AIC, BIC).

2) When in Doubt: Over-specify the Model.
   a) Estimate quadratic or cubic terms, and interactions with Z.
   b) The estimates will be consistent even if these terms are unnecessary. The cost, of course, is that you are sacrificing statistical power.

7. Fuzzy RDD
   a. Sometimes RDDs are "Fuzzy" (FRDD)
      i. Cut off is not strictly applied
      ii. Slight deviation between the actual and predicted class size
   b. When Assignment Rule is Loosely Applied:
      i. Suppose we want to study impact ofclass size on test scores
      ii. If you have a class with 20 students vs. class with 50 students, does the latter have lower test scores?
      iii. if the rule is strictly applied, then the size of the class should hit 40, but even before hit 40, it is divided in to 2 classes.
         1) There is some "fuzzyness" to the discontinuity
         2) We can use an indicator for exceeding the threshold as an instrument for the treatment
      iv. So, we use instrumental variable
         1) Classsize = P0 + P1*officialRecommendedSize + E
         2) **TestScore** = B0 + B1*classsize + E
         3) Official recommended size affects only on class size, then affect on test size

1.3 Fuzzy RDD Specification

Denote the probability of receiving treatment when $X \geq X_c$ as p_r, and the probability of receiving treatment when $X < X_c$ as p_l. In sharp RDD, we simply have $p_r = 1$ and $p_l = 0$. However, in fuzzy RDD, we have $p_r, p_l \in (0, 1)$ and we require $p_l \neq p_r$ to identify the causal effect of interest.

To estimate the causal effect, we create two dummy variables:

1. Dummy variable $Z = 1$ if $X \geq X_c$ and $Z = 0$ otherwise;
2. Dummy variable $T = 1$ if treatment is received and $T = 0$ otherwise.

Note that Z and T are the same thing in sharp RDD, but they are different in fuzzy RDD. The regression we wish to estimate is

$$Y = \beta_0 + \beta_1 T + \beta_2 (X - X_c) + \varepsilon$$

However, if you follow through the same limit arguments presented in Section 1.1, you can see that β_1 no longer measures the "jump" anymore, because around the cutoff, there are treated and controlled units on both sides of the cutoff. As a result, receiving treatment is no longer random (i.e., it no longer depends just on an arbitrary cutoff point), and the above regression is endogenous. Luckily, we have a natural instrumental variable available, which is Z. Note that Z is correlated with T precisely because $p_l \neq p_r$, and Z affects Y only through its effect on T. Using Z as an instrument for T in a standard IV regression can produce reliable causal estimates.

Assumptions of Methods

IMPORTANT: Observational methods bear stronger assumptions, which often cannot be tested. As such, the assumptions should always be stated explicitly!

Design	When to use	Advantages	Disadvantages
Randomized Experiment	Whenever possible	Gold standard, most powerful.	Not always feasible, ethical or legal.
Matching Techniques	When all other methods are not possible (sometimes in tandem with other methods).	Resolves confounded differences between treatment and control (aside from treatment), post-hoc. Avoids overfitting, model dependence.	Assumes no unobserved differences (often implausible assumption, and it's not testable).
Regression Discontinuity	If an intervention has a clear, sharp (arbitrary) assignment rule	Usually easy to argue and show descriptive evidence that the discontinuity is there and having an effect	Can only look at a local sub-group of the overall sample
Difference-in-Differences	If two groups are growing at similar rates before one is "treated" Baseline and follow-up data are available	Eliminates any fixed differences not related to the treatment	Can be biased if trends are not parallel before the "treatment" Ideally have 2 pre-intervention periods of data
Synthetic Control Method	If you have multiple groups observed over time, and one group is treated while others are not. If DiD parallel trend assumptions are violated.	Does not require the strict parallel trends assumption; just requires that treated group's pre-trend can be modeled well as a weighted average of control pre-trends.	Only recovers Average Treatment on the Treated (not ATE); estimates may not generalize to other units. Statistical inferences are less straightforward / harder to evaluate.
Instrumental Variables	If you can identify an exogenous variable that strongly influences the RHS variable of interest, and does not impact the final outcome of interest (except via the RHS variable).	Resolves omitted variables, simultaneity and measurement error all at once.	Good instruments are hard to find. While "strong influence" can be evaluated, the "backdoor criterion" can not be tested empirically. Has to be argued.