**SS489C: Quick Summaries of Common Empirical Strategies**

Empirical strategies typically try to answer counterfactual questions. A common counterfactual question follows the following form: what would have happened to a given person's behavior if she had been subjected to an alternative policy (e.g. would she work more if marginal taxes were lower, would she earn less if she had not gone to school, would she be more likely to be immunized if there had been an immunization center in village?). Listed below are quick summaries of some of the most common empirical strategies.

1. **Randomized Evaluation (a.k.a Randomized Control Trial or RCT)**
2. *What is it?*

* In a randomized experiment, a sample of individuals is selected from the population (note that this sample may not be random and may be selected according to observables).
* This sample is then divided randomly into two groups: the Treatment group and the Comparison group.
* *Specification:* , where is an indicator equal to 1 if individual i is in the treatment group.
* *STATA Example:* reg testscore i.treatment

1. *Identification of Average Causal Effect*

* Average causal effect is the simple difference between the treatment and comparison group averages ( so long as the groups selected randomly (i.e. no correlation between and ). Randomized evaluation is the gold standard!

1. *Common Internal Validity Threats*

* Lack of Fidelity:

An experiment is only as good as the randomness of the treatment assignment. Authors typically provide tables to convince readers that randomization was achieved in practice (e.g. the characteristics of treatment and comparison groups look the same on average).

* Non response bias:

People may move away during the experiment. If people who leave have particular characteristics systematically related to the outcome then there is attrition bias. (cf. Hausman and Wise (1979) about attrition in the NIT experiment).

* Mix up of Treatment and Controls:

Sometimes, maintaining the allocation to control and treatment to be random is almost impossible. Example: (Krueger 2000) evaluation of the Tennessee Star small class size experiment: children were moved to small classes (due to parental pressures, bad behavior, etc..). The actual class is therefore not random even though the initial assignment was random. It is then important to use the initial assignment as the treatment, because it is the only variation that was randomly assigned. It can then be used as an instrument for actual class size.

1. *Other Problems*

* Typically very time consuming and expensive.
* Often requires institutional buy-in that may be very difficult for you to get.
* Often small samples – low power.

1. **Basic OLS Regression**
2. *What is it?*

* Loosely speaking, a regression finds the difference between individuals in the treatment and comparison groups “controlling” for observable characteristics.
* More formally, a regression yields a weighted average of comparisons between people in the treatment and comparison group with the same observable characteristics.
* *Specification:* , where is an indicator equal to 1 if individual i is in the treatment group and is a vector of characteristics of individual I (“controls”).
* *STATA Example:* reg reenlist i.treatment i.age i.sex i.afqtcat

1. *Identification of Causal Effect*

* is the average causal effect when is not correlated with (conditional on ). In other words, treatment is assigned randomly conditional on .

1. *Common Internal Validity Threats*
   * Omitted variable bias (Under-controlling):

This occurs when variables that are correlated with an individual’s likelihood of treatment are not included in the regression (“controlled for”). The classic example is that of returns to education. If ability (or other factors affecting future earnings) are correlated with education choice and are not included in the regression, the OLS coefficient is biased.

* + Over-controlling:

Controlling for variables that are caused by the variable of interest will also lead to biased coefficient. For example, if wage and ability (as measured by IQ, for example), are both caused by schooling, then controlling for IQ in an OLS regression of wage on education will lead to a downward bias in the OLS coefficient of education (intuitively: the ability variable picks up some of the causal effect of education, namely the increase in wages which is due to the effect of education on ability which itself affects wages).

1. **Difference-in-Differences**
2. *What is it?*
   * Leverages a “Natural” experiment and data that tracks groups (or individuals) over time.
   * Compares the difference in outcomes before and after a policy change for a group affected by a change (Treatment Group) to a group not affected by the change (Comparison Group).
   * *Specification:* , where is an indicator equal to 1 if individual i is in the treatment group (this does not change over time) and is an indicator equal to 1 if year (or other time period) is after the policy change.
   * *STATA Example:* reg health i.medicaidexpandstate i.post2010 i.medicaidexpandstate##i.post2010
3. *Identification of Causal Effect*

* is the average causal effect so long as the treatment group would have changed in the same way as the comparison group if it had not been treated (“Common Trends” or “Parellel Trends” assumption).
* In other words, any difference between the two groups would be fixed over time if neither are treated. (Note: this is much weaker than needing the two groups to be the same).
* Another way to think of this is that is not correlated with in the specification above. If this was correlated, it would imply that there is something else changing for the treatment group in the post period besides exposure to treatment.

1. *Common Internal Validity Threats*
   * Violation of Common Trends

*Example:* If the outcome for the treatment group is always trending upward and the comparison group is always flat, a DD estimate of the average causal effect will be biased upward.

Economists employ a number of strategies to convince readers that their DD results are not due to a violation of common trends: (1) Event study - shows how the difference between treatment and comparison group evolves over time (shouldn’t see any movement prior to the actual treatment). (2) Placebo start date– Run DD with “pretend” treatment start dates during the pre-period (shouldn’t see any effects). (3) Alternative comparison groups. (4) DDD strategy takes the difference between DD of interest and a placebo DD that should yield zero if common trends assumption is true (e.g. pre vs. post, treated vs. untreated states, but for population that does not receive treatment). See Gruber (1994) for an example of DDD in action.

* + Targeting based on Differences

A pre-condition of the validity of the DD assumption is that the program is not implemented based on the pre-existing differences in outcomes.

*Example:* “Ashenfelter Dip” – Once upon a time, it was common to compare wage gains among participants and non participants in training programs to evaluate the effect of training on earnings. (Ashenfelter and Card 1985) note that training participants often experience a dip in earnings just before they enter the program (which is presumably why they did enter the program in the first place). Since wages have a natural tendency to mean reversion, this leads to an upward bias of the DD estimtate of the program effect.

1. *Other Problems*
   * *Time Horizon external/internal validity tradeoff:* Common trends is more likely to hold over a shorter time horizon (just before policy change to just after), but this means measuring only a short-term effect which might be of less interest. The short-term and longer-term effects of a policy often differ (e.g. imagine a $2/gallon tax on gasoline, the long-term effects on gas consumption is likely to be larger than the short-term effects since people can adjust their lifestyles over time).
2. **Generalized Difference-in-differences (sometimes called Fixed Effects)**
3. *What is it?*
   * Like simple DD this leverages a “Natural” experiment and data that tracks groups (or individuals) over time. It generalizes DD for more than two periods and more than 2 groups.
   * The advantage of generalized DD compared to simple DD is that it can handle lots of policy changes over time in a single regression. It can also easily adjust from a treatment indicator to a continuous measure (e.g. minimum wage, size of tax credit, etc.).
   * In place of simple “TREATED” indicator, this approach uses separate indicators for each group. These are called group “fixed effects”. These capture average differences between groups that are “time-invariate” (i.e. constant over time).
   * In place of a simple “POST” indicator, this approach uses time period “fixed effects”. These capture changes over time that are common to all groups. In other words, these essentially subtract out the average differences between years.
   * *Specification:* , where is an indicator equal to 1 if individual i is treated in period t (note: this variable now varies over time for each individual). represents time period fixed effects (a series of indicators for each value of t). represents individual fixed effects (a series of indicators for each value of i).
   * *STATA Example:* reg health i.medicaidexpanded i.year i.state, cluster(state)
   * *ALT STATA Example: xtset state year; xtreg health i.year, fe*
4. *Identification of Causal Effect*
   * This approach uses variation in outcomes (and treatment) within group over time (accounting for average differences between time periods) to identify .
   * One way to think of it is that a given group is compared to itself in a different time period (when its treatment differed), accounting for the overall differences across groups in those time periods.

* Like in the simple DD, is the average causal effect so long as there are “Common Trends” across groups. In particular, any unobserved differences between cities must be constant over time (“time invariant”).

1. *Common Internal Validity Threats / Other Problems*

* See DD

1. **Instrumental Variables**
2. *What is it?*

* Leverages partial/incomplete random assignment. Short story:
  + You want to estimate the effect of X on Y, but X is not randomly assigned.
  + There is another variable Z (instrument) that …

a) has a causal effect on X (First Stage)

b) is “as good as randomly” assigned (Independence Assumption)

c) only affects Y through its effect on X (Exclusion Restriction)

* + The part of the variation in X that is caused by Z is not endogenous (no selection bias) 🡪 use this variation to estimate the effect of X on Y
* Simple Example:
  + Want to estimate the effect of receiving writing tutoring (X) on GPA (Y).
    - will be a biased estimate of the effect of tutoring because those who seek it out tend to have a higher GPA.
  + Randomly selected cadets are given reminders that writing tutors are available to them (Z). Cadets are randomly offered the option of having a writing tutor.
    - is an unbiased estimate of the effect of getting a reminder on likelihood of receiving tutoring (First Stage).
    - is an unbiased estimate of the effect of getting a reminder on GPA (Reduced Form). This is called the *Intent-to-Treat (ITT)* effect. The estimate will be unbiased so long as Z is “as good as randomly” assigned. BUT, it isn’t an answer to the question you were really after (in this case).
    - is an IV estimate of the Local Average Treatment Effect (LATE) of tutoring on GPA (this simple version is called a Wald Estimator). LATE is the estimate of the effect on the “compliers”. In this example, the compliers are the cadets who change their behavior because of the reminders. If those cadets differ substantially from other cadets, then the LATE may be very different from the average treatment effect (ATE).
* *Specifications:*
  + *OLS:* , since X is not randomly assigned (endogenous) 🡪 🡪OLS estimate of will be biased.
  + *Reduced Form:* , this is a regression that returns the effect of the instrument (Z) on the outcome (e.g. the effect of being offered admission to a charter school – NOT the effect of actually going to a charter school).
  + *First Stage:* , this is a regression that returns the effect of the instrument (Z) on the “treatment” variable (X). must be non-zero and significant. Otherwise Z doesn’t actually move X.
  + = (this is equivalent to Wald Estimator above)
  + *Two-Stage Least Squares (2SLS)* provides an alternative way to generate an estimate of that allows for easily including other (exogenous).
    - *First Stage:*
    - Use first stage estimates to predict X:
    - *Second Stage:*
    - Note: If independence and exclusion restriction assumptions hold for Z, then .
* *STATA Example:*
  + First stage: reg tutor i.reminder satscore, robust
  + Reduced Form: reg gpa i.reminder satscore, robust
  + 2SLS: ivregress 2sls gpa (tutor=reminder) satscore, vce(robust)

1. *Identification of Causal Effect*
   * IV yields the Local Average Treatment Effect (LATE), so long as:
     1. -- the instrument is uncorrelated with the error term (Independence / Exclusion Restriction)
     2. – the instrument is correlated with the treatment variable (First Stage)
   * Difficult to test validity 🡪 want validity to be “obvious”
     1. Random encouragement designs: Probability of treatment varies randomly across people and then actual treatment status may then result from a choice (endogenous)
        1. *Vietnam era draft lottery (Angrist 1990):* high lottery number makes it more likely you are drafted (can still dodge with high or volunteer with low)
        2. *Flu vaccine reminders (Imbens, K.Hirano, D.Rubin and A.Zhou 2000):* randomly assigned letters reminding doctors to propose flu vaccine to clients (actual flue vaccine not randomly assigned).
     2. Instrument approximates random encouragement design
        1. *Judge Severity (Dobbie, Goldin, and Yang 2018):* Random assignment to judges with differing sentence severities
     3. Policy Changes: DD can be used as the first stage (instrument is Post x Treated) and then you can control for Post and Treated in second stage.
        1. *School Program implementation (Duflo 2000):* Implementation of school program in some regions but not others produces an increase in schooling (first stage), this is used to estimate the effect of schooling on wages (second stage).
        2. *Clean Air Act (Chay, Dobkin, and Greenstone 2003):* Clean air act binds for some states and not others in 1970. This leads to a reduction in air pollution in those states (first stage) which is used to estimate the effect of air pollution on mortality (second stage)
2. *Common Internal Validity Threats*
   * Bias can be magnified. If instrument is not truly exogenous, even if the bias of the reduced form is similar to OLS, the bias of the IV will be made much larger since the denominator is less than 1.
   * Randomly assigned instruments are not necessarily valid. If the instrument has a direct effect on the outcome (not through the treatment variable), then IV will be biased (Exclusion Restriction).
     1. E.g. If a low draft number led people to stay in college to evade the draft, this would be a direct effect on earnings (not through military service).

1. *Other Problems*
   * The LATE estimate is “local” in that you are only finding the treatment effect on the group that is affected by the instrument ( the “compliers”). If they look different than the rest of the population, then the effects of providing a treatment to the rest of the population may look very different from the LATE.
2. **Regression Discontinuity**
3. *What is it?*
   * Leverages a situation where exposure to treatment varies ***discontinuously*** at a cutoff or threshold value in a continuous assignment variable (sometimes called running variable). Examples include:
     1. (e.g. student test score, % free or reduced lunch eligible students at a school, percent voting for a given party, etc.).
   * Compares individuals (or other units like schools or firms) just above and just below the cutoff in the assignment variable.
     1. These groups