Measuring Impact

Program and Policy Evaluation with Observational Data

Daniel L. Millimet

Southern Methodist University

23 May 2013

Introduction

- Measuring causal impacts is best understood using the Rubin Causal Model and the idea of potential outcomes
- Recall, the following notation:
  - $D = 0, 1$ denotes treatment assignment for a given subject
  - $Y(0)$ denotes the outcome that would arise for the subject if $D = 0$
  - $Y(1)$ denotes the outcome that would arise for the subject if $D = 1$
  - The subject-specific treatment is given by

$$\tau = Y(1) - Y(0)$$

- This causal effect for a given subject is *never* observed because either $Y(0)$ or $Y(1)$ is always missing

- Due to the problem of the missing counterfactual, *any* statement concerning the causal effect of a treatment, D, *must* rely on a set of assumptions in order to estimate the missing counterfactual
- Randomization is one method to estimate the missing counterfactual
  - Under the assumption of random assignment:
    - The average outcome of the control group is a valid (unbiased) estimate of the missing $Y(0)$ for subjects in the treatment group
    - The average outcome of the treatment group is a valid (unbiased) estimate of the missing $Y(1)$ for subjects in the control group
  - The average difference in outcome across the treatment and control groups is a valid (unbiased) estimate of the average treatment effect
- Thus, randomization allows us to obtain valid estimates of the missing counterfactual by using *other* subjects assigned to a *different* treatment assignment
- This succeeds because randomization ensures that the treatment and control groups are *identical in expectation* in all respects *except for the treatment*

- As mentioned earlier, randomization is not always feasible
- In such cases, researchers utilize survey data in an attempt to measure causal impacts of treatments
- Researchers may collect the data themselves or utilize data collected by others
  - Census
  - Current Population Survey (CPS)
  - Survey of Income and Program Participation (SIPP)
  - Panel Study of Income Dynamics (PSID)
  - National Longitudinal Surveys (NLS)
  - National Center for Education Statistics (NCES)
- Survey data is referred to as *observational data* to denote the fact there is no randomization; subjects are simply observed in the real world

- Question turns to how to obtain 'valid' estimates of the missing counterfactual with observational data
- With observational data, there are two options for estimating the missing counterfactual

1. **Cross-Sectional Approach:** Use outcomes of *other* subjects with a *different* treatment assignment observed *at the same point in time*
   - ★ Example: End-of-year survey of students in DFW with data on participation in SBP and school performance
2. **Longitudinal Approach:** Use outcomes of *same* subject with a *different* treatment assignment observed *at a different point in time*
   - ★ Example: Repeated end-of-year surveys of students in DFW with data on participation in SBP and school performance

- Two approaches have different data requirements
  - First approach requires *cross-sectional* data
    - ★ Data on numerous subjects collected at a single point in time
  - Second approach requires *longitudinal* or *panel* data
    - ★ Data on numerous subjects collected at at least two points in time

- Main difficulty *under either approach* is that participation in the treatment is not randomized
- Instead subjects *self-select* into the treatment and control groups
  - Cross-sectional approach
    - ★ Outcomes of *other* subjects with a *different* treatment assignment may be a poor approximation of the outcome that a given subject would have experienced due to *differences between subjects in attributes* that lead some to self-select into the treatment and others to self-select into the control
  - Longitudinal approach
    - ★ Outcomes of the *same* subject at a *different point in time* under a different treatment assignment may be a poor approximation of the outcome a given subject would have experienced *today* due to *differences over time in attributes* that lead the subject to self-select into the treatment at one time but not another

Cross-Sectional Approaches

Introduction

- Estimation of causal effects based on cross-sectional approaches uses outcomes of *other* subjects to estimate missing counterfactuals
- Many statistical techniques are available based on different assumptions concerning how subjects self-select into treatment assignment
  1. Random selection
     - ★ Subjects *do not* self-select on the basis of any individual attributes correlated with potential outcomes
  2. Selection on observed attributes only
     - ★ Subjects *do not* self-select on the basis of any individual attributes correlated with potential outcomes but *unobserved* by the researcher
     - ★ Subjects *may* self-select on the basis of individual attributes correlated with potential outcomes and *observed* by the researcher
  3. Selection on unobserved attributes
     - ★ Subjects *may* self-select on the basis of individual attributes correlated with potential outcomes and *unobserved* by the researcher

Example:

- Question: What is the causal effect of SBP participation on child obesity?
  - Random selection
    - ★ Children decide to participate based *only* on convenience (and factors affecting 'convenience' are unrelated to child health)
  - Selection on observed attributes only
    - ★ Children decide to participate based *only* on convenience, family SES status, and other attributes included in the survey
  - Selection on unobserved attributes
    - ★ Children decide to participate based on convenience, family SES status, other attributes included in the survey, *and* nutritional knowledge and/or other attributes not measured in the survey

Cross-Sectional Approaches

Estimation

Random Selection:

- Analogous to random experiments
- Randomization just arises naturally $\Rightarrow$ natural experiments
- Casual effects estimated using the average difference in outcomes across subjects in the treatment and control groups

- Example: Effects of children on labor market outcomes
  - Angrist & Evans (1998) assess the causal effect of children on the labor supply of men and women
  - Twins represent a 'random' occurrence of an extra child
  - Find that an extra child reduces female labor supply, but not male labor supply

Selection on Observed Attributes Only:

- Analogous to a random experiment *conditional* on the set of observed attributes that are correlated with treatment assignment and potential outcomes
- Thus, estimation methods 'control' for this set of observables
- Often referred to as *quasi-experimental* methods
- Causal effects estimated using the average difference in outcomes after removing the *confounding* effects of the set of observables
- Observed attributes, denoted by X, may be controlled for in many ways
  1. Regression (Ordinary Least Squares)
  2. Matching
  3. Propensity Score Methods

- Matching
  - The missing counterfactual for a given subject is the average outcome of *other* subjects with *different* treatment assignment but *identical* attributes, X
  - This yields an estimated treatment effect for each subject
  - Causal effect estimated using the average subject-specific treatment effect
  - Other effects may be estimated by averaging over different subsets of subjects (e.g., race or gender groups)
  - Intuition is easy to explain to policymakers, boards, others: conditional on X, data mimics a random experiment

- Example: National Supported Work (NSW) Demonstration Project
  - Randomized job training program conducted from 1975-79
  - Program gave hard-to-employ individuals jobs in a supportive, but performance-based, work environment for 12-18 months
  - Eligibility based on ex-convicts, ex-drug addicts, long-term AFDC recipients, and HS dropouts
  - Among those eligible, subjects randomized into treatment or control (no job) groups
  - Dehejia and Wahba (1999) re-evaluate the program using matching methods by taking the original data on the treatment group (and discarding the original controls group) and using survey data to form the new control group
  - Matching is able to replicate the experimental results

Selection on Unobserved Attributes Only:

- Prior methods are not valid as conditioning on X is insufficient
- Need to control for, say, X and W, but W is not observed
- Estimation is very difficult
- Unfortunately, this situation is often confronted
  - Effects of education: self-selection based on innate ability
  - Effects of private schooling: self-selection based on ability, motivation
  - Effects of training: self-selection based on motivation, reliability, work ethic
  - Effects of SBP: self-selection based on neighborhood quality, home nutritional environment
- Estimation methods
  1. Instrumental Variables
  2. Regression Discontinuity

- Instrumental Variables
  - Most common technique used by empirical economists
  - Requires data on an attribute of subjects, denoted by Z, that
    1. Affects the decision to self-select into the treatment or control group
    2. But is uncorrelated with potential outcomes
    3. And, hence, the unobserved attributes that affect the decisions of subjects to self-select into the treatment or control group and correlated with potential outcomes

- Example: Effects of private schools
  - Evans & Schwab (1995) assess the causal effect of attending a Catholic HS on HS graduation and college enrollment
  - Control group includes students attending public schools
  - Use Catholic upbringing as an instrument
  - Find that Catholic schools cause higher graduation and college enrollment rates

- Regression Discontinuity
  - Applicable to situations where eligibility for participation is on the basis on some observed attribute: a score, S, relative to cut-off value, S^*
  - Subjects just above and below S^* are *nearly* identical in all respects except for eligibility
  - The missing counterfactual for a given subject is the average outcome of *other* subjects on the other side of S^*
  - This yields an estimated treatment effect for each subject
  - Causal effect estimated using the average subject-specific treatment effect
  - Intuition is easy to explain to policymakers, boards, others: conditional on being close to S^*, data mimics a random experiment

- Example: Effects of Medicaid
  - De La Mata (2012) assesses the causal effect of Medicaid on take-up, private HI coverage, child health
  - Use a RD method to compare households just above and below income cutoffs that determine eligibility
  - Find that Medicaid crowds out private HI, increases preventive care, and has no impact on child health (as measured by self-reported health, obesity status, and school absences)

Longitudinal Approaches

- Estimation of causal effects based on longitudinal approaches uses outcomes of the *same* subject at *different* times to estimate missing counterfactuals
- Two main approaches
  1. Differences model
  2. Differences-in-Differences model

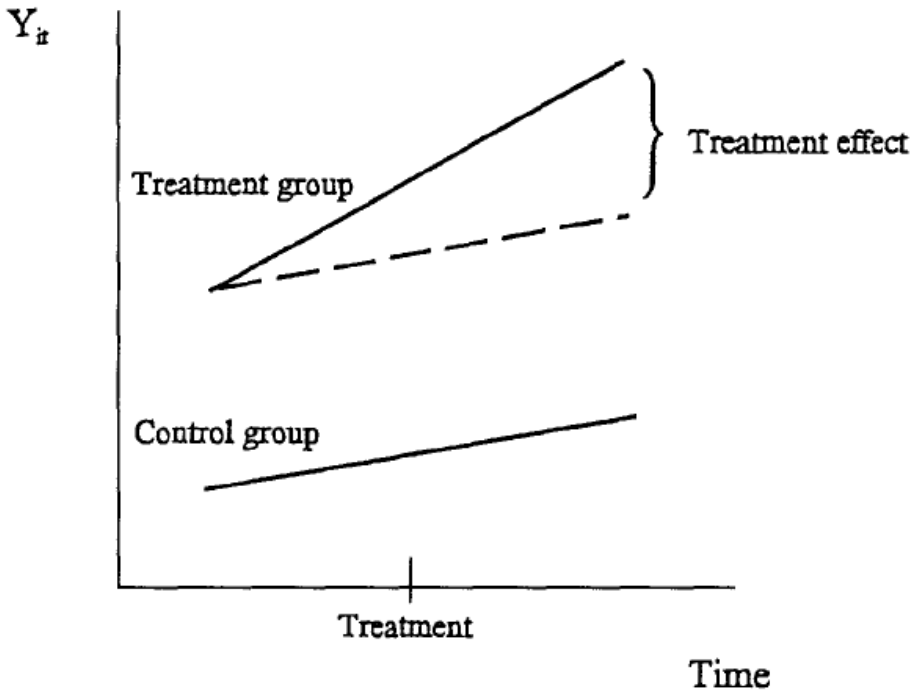
Differences Model:

- Casual effects estimated using the average difference in outcomes over time across subjects during the period of treatment and period of control
- While the same subject is used to estimate the missing counterfactual, it is still an estimate since the two outcomes are not measured at the same time
- Approximation may be poor if attributes of the subject that are correlated with potential outcomes also change over time
- If the only relevant attributes that *change over time* are observed by the researcher, then this is a case of **selection on observed attributes only**
- If at least some relevant attributes that *change over time* are unobserved by the researcher, then this is a case of **selection on unobserved attributes**

Differences-in-Differences Model:

- If at least some attributes that *change over time* and are *correlated with potential outcomes* are unobserved by the researcher, an alternative strategy is based on DiD
- Casual effects estimated using the average difference between the *average change* in outcomes over time across subjects whose treatment status changes and the *average change* across subjects who are never treated

- Requires the relevant unobserved attributes that change over time to be uncorrelated with treatment assignment



- Example: Effects of the minimum wage
  - Card & Krueger (1994) assess the causal effect of minimum wage hike in NJ in 1992
  - Compare the *change* in employment at fast food restaurants in NJ to the *change* in employment at similar restaurants just across the PA border
  - Find no adverse impact of the minimum wage hike

Conclusion

- Estimation of causal effects using observational data must rely on assumptions that are often untestable
- Questions to ask:
  1. What type of selection process is assumed by the researcher?
  2. Are there possible unobserved confounders that may invalidate the estimation procedure?
- Other issues to remember
  1. Causal effects are defined with respect to the outcome being assessed; other outcomes may yield different effects of a program or policy
  2. Many different types of causal effects can be estimated when treatment effects are heterogeneous
     - ★ Average effects for different groups
     - ★ Distributional effects
  3. With observational data, measurement error is commonplace and often consequential (Millimet 2011)