Measuring Impact

Conceptual Issues in Program and Policy Evaluation

Daniel L. Millimet

Southern Methodist University

23 May 2013

Introduction

- Primary concern to economists, policymakers, organizations, etc. is to measures the *causal* effect of a program or policy on outcomes of importance
- Statistical and econometric literature analyzing causation has seen tremendous growth over the past several decades
- The more we learn, the more we realize how *complex* the world can be and how *difficult* measuring impacts can be
- Our objective is to provide
    1. A non-technical overview of issues and concepts that arise when seeking to measure *causal* impacts
    2. A brief introduction to some of the methods used by researchers
- By the end, you will hopefully agree with Mark Twain:
  "*Education: the path from cocky ignorance to miserable uncertainty*."

Outline

1. Definition of causation
2. Types of questions one may ask
3. Types of data
4. Data issues
5. Proper interpretation of statistical analyses
6. Roadmap to remaining sessions

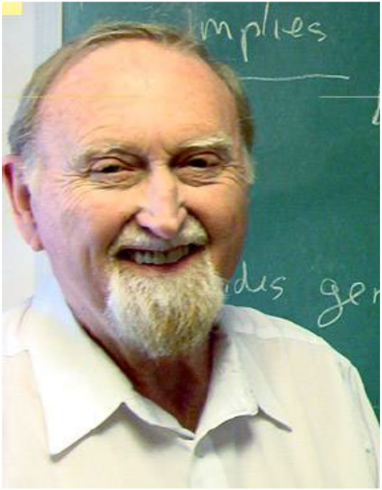
Causation

Philosophy of causality

- Rich literature in analytic philosophy on causality
- Two main approaches to defining causality:
  - **Regularity approaches**: Hume: “We may define a cause to be an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second.” (from *An Enquiry Concerning Human Understanding*, section VII)
  - **Counterfactual approaches**: Hume: “Or, in other words, where, if the first object had not been, the second never had existed.” (from *An Enquiry Concerning Human Understanding*, section VII)

Regularity approach: a minimal conjunction between two objects

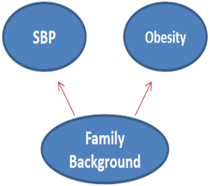
- Suppes: a probabilistic association between the two objects, which cannot be explained away by other factors
- Idea behind *Granger causality*
- Fundamental notion underlying Granger causality is whether one object helps *predict* the occurrence of another
- Prediction, association, and correlation are not what most intend when speaking of *causation*
- While predictive ability answers questions that may be of interest, it *should not* be used to make policy decisions

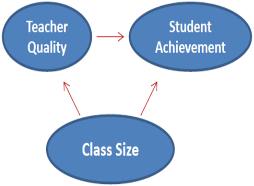


- Correlation (or predictive ability) and "causation" are not synonymous
  - Examples abound where correlation $\not\Rightarrow$ causation
    - ★ Example: SBP and childhood obesity
  - Examples abound where a *lack of* correlation $\not\Rightarrow$ a *lack of* causation
    - ★ Example: California's class size reduction policy

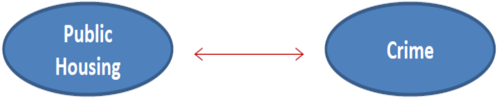
- Two reasons why associations may not be indicative of causation

  1. Confounding factors ⇒ correlation is driven by *other* factors
     - ★ SBP & childhood obesity: family background
     - ★ California's class size reduction policy: teacher quality





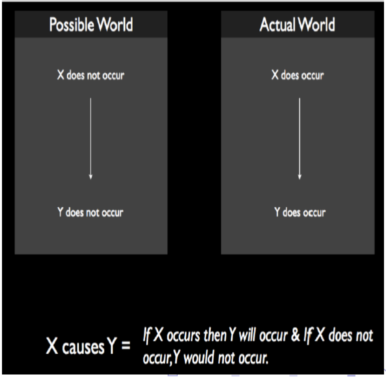
  2. Reverse causation ⇒ correlation is driven by causation in the other direction
     - ★ Public housing and crime?
     - ★ Marital wage premium?
     - ★ Children and female labor supply?



Counterfactual approach: imagining a range of possible worlds

- In my view, this view is most consistent with what most intend when speaking of *causation*
- Most relevant for evaluating programs and policies
- Basic idea is to imagine alternative worlds where one (and only one) object is changed and to assess the differences $\Rightarrow$ *Sliding Doors*
- Dominant view for measuring impacts in microeconomics today

Greiner & Rubin (2011):
"For analysts from a variety of fields, the intensely practical goal of causal inference is to discover what would happen if we changed the world in some way."



- Typically referred to as the *Rubin Causal Model* (Neyman 1923, 1935; Fisher 1935; Roy 1951; Quandt 1972, 1988; Rubin 1974)
- Crucial underpinning is the notion of *potential outcomes*

- Potential outcomes refer to the outcome that *would be realized* under different states of nature
    - Example: A sick individual may receive either Treatment 0 or 1. The outcome is either Recovery or Death. Thus, there are two possible *states of nature* (Treatment 0 or 1) and there is an outcome that would be realized in each state of nature.
- Under the counterfactual approach, the *causal effect* of Treatment 1 *relative to* Treatment 0 would be the difference in outcomes across these two states of nature

- Formally,
  - Let $D = 0, 1$ indicate the treatment received by the individual
  - Let $Y(0)$ indicate the outcome (Recovery, Death) the individual *would* experience *if* s/he receives Treatment 0 ($D = 0$)
  - Let $Y(1)$ indicate the outcome (Recovery, Death) the individual *would* experience *if* s/he receives Treatment 1 ($D = 1$)
  - $Y(0)$, $Y(1)$ are *potential outcomes* as only one will actually be realized and observed in the world
  - The causal effect of Treatment 1 relative to Treatment 0 on the individual is given by

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

    which is the difference in outcomes in two alternative, but plausible, worlds

- The **counterfactual approach** is a very *simple* yet *powerful* framework in which to think about causation
- Immediately leads to three salient points

1. Causal impacts of a treatment/intervention/program/policy are only defined with respect to a well-defined alternative
    - Typically the alternative is the 'absence of treatment'
    - Not always obvious and must be made explicit
2. Causal impacts are individual-specific
    - Each individual potentially has he or her own potential outcomes and hence treatment effect
    - Referred to as *constant* vs. *heterogeneous* treatment effects
    - Has important implications for thinking about how to interpret the results of data analyses
3. Only one state of nature is actually realized at a point in time
    - We can observe at most one potential outcome for any individual, remaining are missing
    - The causal effect of a treatment is *not observable* for any individual
    - Estimating causal effects must overcome this *missing data* problem
    - To do so, requires assumptions and these assumptions must be credible

Types of Questions One Might Ask

Granger Causality

- Granger causality answers questions concerning the ability of one object to *predict* or *forecast* another
- The underlying reason why it is a good predictor is not of (primary) importance
- However, often we do not care about the *why* and predictive ability is sufficient
- For example:
  1. State governments forecasting welfare caseloads or new applications for unemployment insurance or new applications for public housing
  2. Food banks forecasting demand for food
  3. Non-profits forecasting charitable donations
  4. Police forecasting 911 calls at different times

Types of Question One Might Ask

Causal Effects

- Counterfactual approach answers questions concerning the *expected effect* of a treatment (relative to a well-defined alternative)
- Understanding the mechanisms behind why a causal effect may arise is important, but requires additional analysis
- Examples:
  1. Expected effect of a job training program on employment and earnings
  2. Expected effect of a weekend food program on child nutrition and school achievement
  3. Expected effect of different financial literacy training schemes (class, one-on-one, online, none) on decisionmaking
  4. Expected effect of different incentive schemes on charitable donations (e.g., matching, raffle, gift)
- Examples 1, 2 have *binary* treatments $\Rightarrow$ alternative is the 'absence of treatment'
- Examples 3, 4 have *non-binary* treatments $\Rightarrow$ what is the alternative?

Fundamental Issues in Measurement of Impacts

- Estimation of causal effects is difficult due to *missing counterfactual* problem
  - With a binary treatment, either $Y(1)$ or $Y(0)$ is observed for any given individual
  - The unobserved potential outcome is the missing counterfactual
  - Holland (1986) refers to this as the "fundamental problem of causal inference"
- While the missing counterfactual makes life difficult, it is not impossible
- The implication is that we must *estimate* the missing counterfactual
- Such estimates are only as valid as the *assumptions* that underlie them and the *data* used to derive the estimates
- Discussion of the assumptions and statistical methods will be the focus of the later sections
- For the rest of this section, I focus on understanding data

Fundamental Issues in Measurement of Impacts

Types of Data

- Two types of data are employed to estimate causal effects
  1. Experimental (field or lab)
  2. Non-experimental or observational
- Experimental data refers to data collected by researcher after *randomizing* treatments across subjects (as opposed to allowing subjects to *self-select* into the treatment or control groups)
- Observational data refers to typical survey data collected by researchers after subjects self-select into different treatments
- While experimental data may be preferable – often referred to as the 'gold standard' – experiments are often difficult
  1. Expensive
  2. Unethical
  3. Politically unpopular

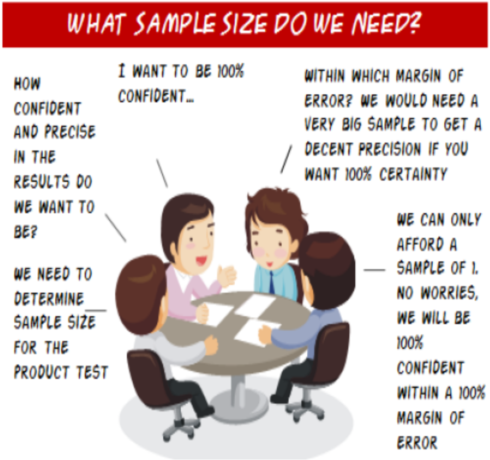
Fundamental Issues in Measurement of Impacts

Data Issues

- Regardless of the data type (experimental or observational), several issues can arise that impact data *quality* or affect the *interpretation* of statistical results
- These include
    1. Sample size
    2. Sample selection
    3. Sample attrition
    4. Measurement issues

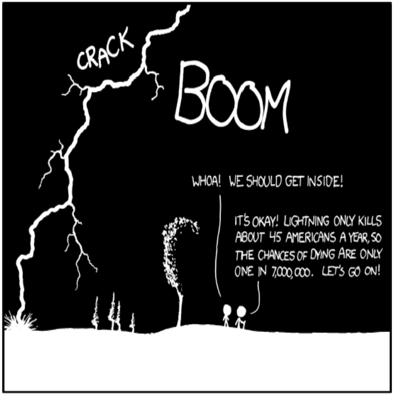
Sample size

- Refers to the number of subjects used in the analysis
- Sample size affects the *precision* of estimates of causal effects
- Precision of estimates is not always discussed outside of academic circles



Sample selection

- Refers to how the subjects used in the analysis are chosen
- Sample selection affects the *representativeness* of the sample
- Affects the interpretation of the results by affecting generalizability



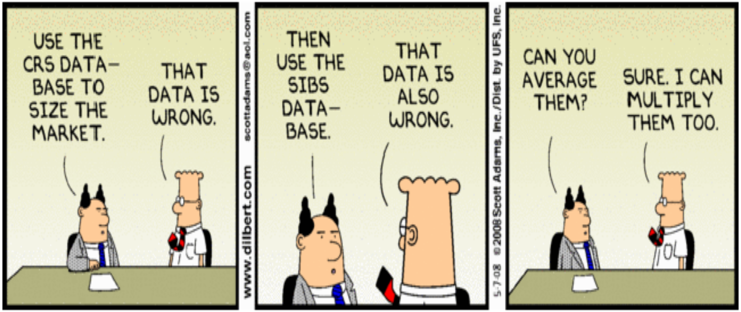
THE ANNUAL DEATH RATE AMONG PEOPLE WHO KNOW THAT STATISTIC IS ONE IN SIX.

Sample attrition

- When analysis entails following subjects over time, final sample may be *nonrepresentative* even if the initial sample is *representative*
- Need to think about whether attrition is *random* or not
- May occur for many reasons
  1. Subjects voluntarily decide to leave study, move away, etc.
  2. Subjects involuntarily decide to leave study (e.g., death, jail, etc.)

Measurement issues

- Definitions of variables, particularly outcomes
  - Short-run vs. long-run
  - Comprehensive vs. specific
- Measurement error
  - Self-reported vs. verified



Fundamental Issues in Measurement of Impacts

Interpretation

- Internal vs. external validity
  - *Internal* validity refers to whether an estimate of the causal effects is valid for the sample used to obtain the estimate
  - *External* validity refers to whether an estimate of the causal effects is generalizable to individuals outside the sample
- Treatment effect parameters
  - When causal effects are heterogeneous, several parameters may be of interest
    1. Average Treatment Effect (ATE) $\Rightarrow$ expected treatment effect for an individual chosen at random from the population
    2. Average Treatment Effect on the Treated (ATT) $\Rightarrow$ expected treatment effect for an individual chosen at random from the *treated* population
    3. Average Treatment Effect on the Untreated (ATU) $\Rightarrow$ expected treatment effect for an individual chosen at random from the *untreated* population
    4. Distributional Treatment Effects

  Note: These answer different policy questions!

Fundamental Issues in Measurement of Impacts

Interpretation

- For any treatment effect parameter, there is an important difference between *ceteris paribus* effects and *policy* effects
  - Ceteris paribus ("all else constant") effects answer questions like: "What is the causal effect of the treatment on an individual *holding all else fixed*?"
  - Policy effects answer questions like: "What is the causal effect of the treatment on an individual?"
  - Answers are not always the same, and both are important
    - Example 1: California's class size reduction policy
    - Example 2: Infra-marginal nutrition assistance

Conclusion

Critical questions to ask of causal statements

1. What is the counterfactual? What is the alternative to the treatment?
2. What is the nature of the data? Randomization or observational?
3. What is the proper interpretation of the results? To whom do the results apply? What is the outcome?
4. Are the results *internally* valid? What assumptions are made to circumvent the missing counterfactual?
5. Are the results *externally* valid? Sample selection?

Roadmap

- Assumptions used to circumvent the problem of the missing counterfactual
  1. Randomization $\Rightarrow$ Tim, Danila
  2. Observational methods $\Rightarrow$ me
- Forecasting and prediction $\Rightarrow$ Tom
  1. Data mining
  2. Big data