Publishing Papers in International Journals: Causality and Identification Strategies

Gene Lai
University of North Carolina at Charlotte

Presented at 2021 TRIA Seminar

Outline

- Types of Empirical Economics
- Causality
- Identification
- Channel or Mechanism

Types of Empirical Economics?

- Measurement
- Model Testing
- Model Estimation for Counterfactuals

Measurement

- How has productivity (e.g., TFP) in the U.S. auto industry changed over the last 30 years?
- What is the effect of college attendance on wage?
- What is the elasticity of aggregate demand for health insurance?

2. Model Testing

- Does Bonterra Energy Corp do well in predicting bidding at oil auctions?

Model Estimation for Counterfactuals

- Economics concern a counterfactual of the form: how would the world have been different if XXX were changed, all else not changed?
- Example: How would student outcomes differ under a different school choice mechanism?
- This is our focus

Causality

- We can never directly observe the causal effects, which is what Holland (1986) calls the “fundamental problem of causal inference.”
- Estimates of causal effects are ultimately based on comparisons of different units with different levels of the treatment.

Causality

- The gold standard for drawing inferences about the effect of a policy is a randomized controlled experiment
  - Example: Vaccine Tests
  - Treatment group vs. Control group (Placebo)
    - Compare actual out with counter facture outcome
  - Randomized

Causality

- However, in many cases, experiments remain difficult or impossible to implement, for financial, political, or ethical reasons
- For example, it unethical to prevent potential students from attending college in order to study the causal effect of college attendance on labor market experiences
- Of course, we can see whether school lunch have an impact on BMI

Causality

- A naive analysis of the observational data might compare the average employment level of states with a high minimum wage to that of states with a low minimum wage because some states have higher cost of living.
- This difference is surely *not* a credible estimate of the causal effect because of the differences in cost of living, price-insensitive consumers.
- A lot of unobservable factors cannot be observed.

Confounders

- These factors, which may be unobserved, are said to be "confounders" meaning that they induce correlation between minimum wage policies and employment that is not indicative of what would happen if the minimum wage policy changed.

Causality and Counterfactuals

- Much of empirical economics aims causal effects
- Causality is a concept defined by a counterfactual question of the form “what would happen if certain things were changed while others were held fixed (Other control variables)?
- This is true even for an RCT (Randomized Controlled Trials)

Identification and Exogenous Shock

- Exogenous variation is an important identification element
- Exogenous variation is X or our variable of interest

Difficulties of Causality

- Causal effects are simply elasticities, but they are difficult to estimate because econometricians rarely observe occasions where one variable is altered while others are held constant, that is, where there is genuine exogenous variation in a variable.
- Examples:
  - How an increase in the minimum wage affects employment
  - How training affects earnings

Identification and Causality

- In economics, researchers use a wide variety of strategies for attempting to draw causal inference from observational data.
- These strategies are often referred to as identification strategies or empirical strategies because they are strategies for identifying the causal effect.
- Angrist and Pischke (2008) present the issue of identification entirely as a search for an approximation to an ideal experiment.
- Please note: Not all interesting questions are causal in nature, and not all identification issues revolve around establishing causality
  - Example: Prediction is important

Exogenous Variation

- The government has a lever at its disposal, and the question is the outcome when the government pulls that lever.
- This simple ceteris paribus comparison surrounding a specific and limited government intervention makes identification relatively straightforward.
- If the goal of a study is simply to establish the average effect of a previous intervention, then as long as this intervention is plausibly exogenous, the causal link to a government policy has been identified.

Exogenous Variation

- But, it represents only an estimate of the average causal effect of a variable under a particular, historical intervention.
- The average treatment effect that comes from such an approach is limited in its applicability.
- For example, we want to know whether minimum wage increases hurt employment as a policy
- In other words, we are interested in causal inference

Identification

- Sometimes the model that provides identification is an economic model
- For example, one might be able to run an experiment to establish that a causal effect exists, but the experiment alone typically cannot identify the economic forces that are behind the causal effect, and these forces are usually at least as interesting as the effect itself.
- Finding exogenous variation in a variable is never sufficient for identification of an economically interesting parameter, as identification is always based on a verbal or mathematical theory.

Exogenous Variation

- Thus, identification can never be free of assumptions or even light on assumptions.
- Without economic “theory” as another foundation stone, it is impossible to make such statistical inference apply directly to the equations of economic behavior which are most relevant to analysis and to policy discussion.

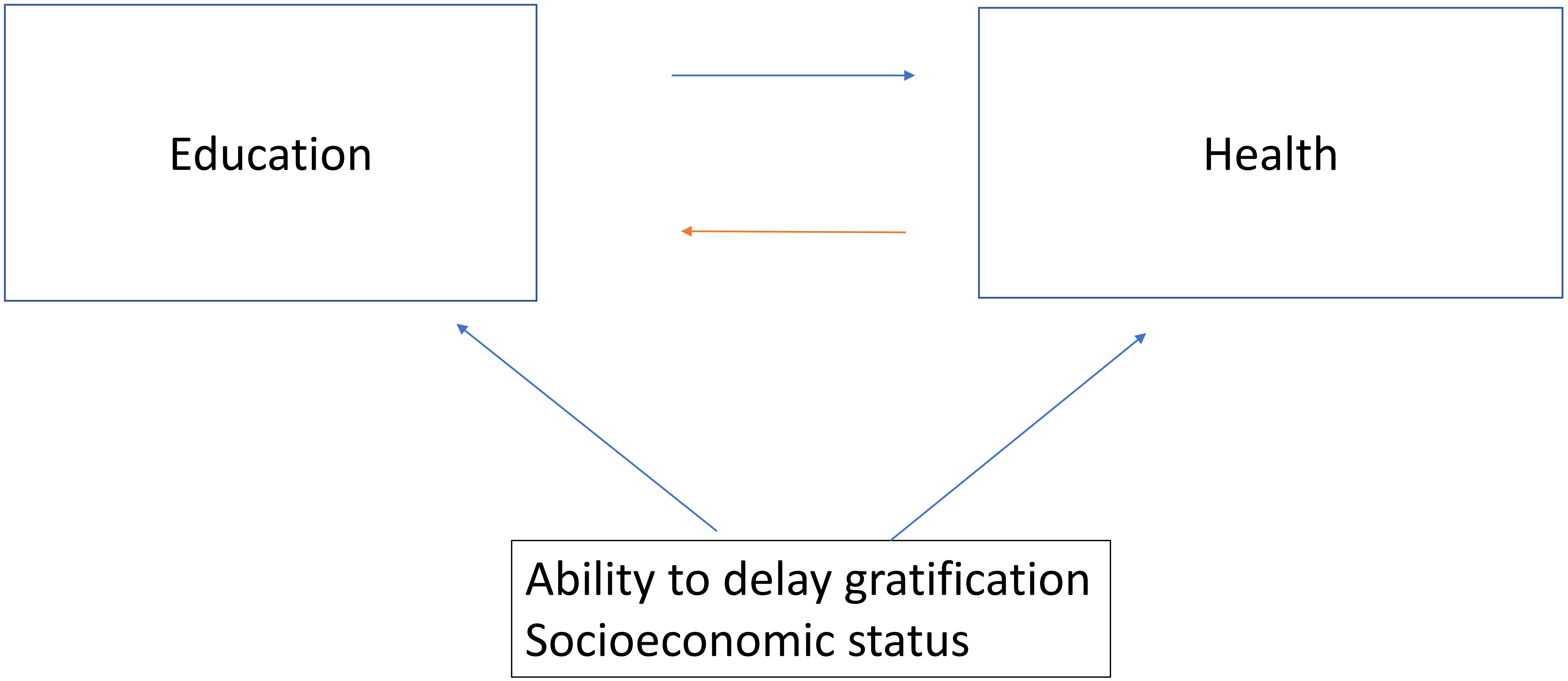
Identification with exogenous variation

- Identification with exogenous variation
  - Example:
  - Whether in-family succession of CEOs hurts performance
  - In contrast to ceteris paribus comparisons involving the impact of a government policy, here the parameter of interest is not an observable elasticity but a summary measure of an underlying agency problem.
  - To understand the magnitude of this agency problem, we want to estimate the loss in performance that is due to the choice of a family member over an outsider.
  - Parameter that measures the performance loss is of interest not because it represents an average, presumably causal, estimated effect: it is interesting because it represents a deeper agency friction.

Endogeneity

- Endogeneity is one of the major problem in identification

Endogeneity Example



Identification Strategies

- Instrumental variable
- Difference in Difference
- Regression Discontinuity
- https://www.youtube.com/watch?v=OcbvyT4danA&t=29s

Difference in Difference

- Important: Need to show NO pre-treament effect https://www.youtube.com/watch?v=J7q2H8aB8bQ&t=304s

PSM

- To conduct DID, we need to control characteristics of firms Xs
- To control Xs, we use propensity score matching (PSM).
- What are the variables needed to be included in PSM?
- The variables that are unrelated to the exposure (treatment) but related to the outcome should always be included in a PS model.
    - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1513192/

Regression Discontinuity Designs

- A regression discontinuity design enables the estimation of causal effects by exploiting discontinuities in incentives or ability to receive a discrete treatment.
- For example, school district boundaries may imply that two children whose houses are on the same street will attend different schools, or birthdate cutoffs may limit eligibility to start kindergarten between two children born only a few days apart.
- Many government programs are means-tested, meaning that eligibility depends on income falling below a threshold.
- In these settings, it is possible to estimate the causal effect of attending a particular school or receiving a government program by comparing outcomes for children who live on either side of the boundary, or by comparing individuals on either side of an eligibility threshold.

Regression Discontinuity

- The presence of an exogenous variable, referred to as the forcing variable, like the student's birthday or address, where the probability of participating in the program changes discontinuously at a threshold value of the forcing variable, or college entrance exam score.
- Assumption that the individuals close to the threshold but on different sides are otherwise comparable, so any difference in average outcomes between individuals just to one side or the other can be attributed to the treatment.

Supplementary Analyses

- Seek to shed light on the credibility of the primary analyses.
- Placebo analysis, where pseudo-causal effects are estimated that are known to be equal to zero based on a priori knowledge;
- Sensitivity and robustness analyses that assess how much estimates of the primary estimates can change if we weaken the critical assumptions underlying the primary analyses
- Identification and sensitivity analyses that highlight what features of the data identify the parameters of interest;

Placebo Analyses

- The researcher replicates the primary analysis with the outcome replaced by a pseudo-outcome that is known not to be affected by the treatment.

Channel and Mechanism

- The economic force behind the causality
- Example:
  - Difference schools produce different SAT scores
  - Why?
  - Better teachers? Better students? Better resources?

Alternative Explanations

- Try to eliminate alternative explanations

References

- Athey s. and G. W. Imbens, The State of Applied Econometrics: Causality and Policy Evaluation, Journal of Economic, Perspectives—Volume 31, Number 2—Spring 2017—Pages 3–32
- **Alan D. Crane, Sébastien Michenaud, James P.Weston, The Effect of Institutional Ownership on Payout Policy: Evidence from Index Thresholds,** *The Review of Financial Studies*, Volume 29, Issue 6, June 2016, Pages 1377–1408,