# Causal Inference

A. Colin Cameron
Univ. of California, Davis

These slides are part of the set of slides
A. Colin Cameron, Introduction to Causal Methods
https://cameron.econ.ucdavis.edu/causal/

March 2023

# Causal Inference

- Regression measures correlation and not necessarily causation.
- There are several methods to estimate causal effects using observational data.
- These slides emphasize analysis of data arising from so called natural experiments or quasi-experirments.
- Each method relies on its own assumptions.
- Each method has special data needs.
- In many applications we do not have the data to allow use of any of the causal methods.
- Use of any of these methods entails assumptions and subtleties
    - **so you need to read relevant references.**

# Outline

# The Causal Inference Problem

- Regression measures correlation and not necessarily causation.
- For example individual earnings are positively **correlated** with education.
- But does that imply that more education is **causing** higher earnings?
- It could be, for example, that we are just picking up that it is higher ability that causes the higher earnings (and, of course higher ability is correlated with more earnings).
- The following paths are enough to induce (noncausal) correlation between earnings and education

$$\begin{array}{ccc} \text{ability} & \longrightarrow & \text{earnings} \\ \downarrow & & \\ \text{education} & & \end{array}$$

# OLS generally gives association and not causation

- Suppose we regress individual annual earnings on years of schooling and other regressors and find that the estimated slope coefficient of years of schooling os 500.
- Then we can only say that one more year of schooling **is associated with** an average increase of $500 in annual earnings.
- In general we cannot say that one more year of schooling **causes** an average increase of $500 in annual earnings
- Note that for many purposes it is enough to measure the association.
    - but at times we want a causal estimate.
- Much of the recent empirical microeconomics research has sought to obtain causal estimates.

# Control function approach

- Causal estimates can be obtained if we make the strong assumption that the model error term is uncorrelated with the regressors.
- Since in the model $y_i = \beta_1 + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + u_i$, consistency of the OLS estimates requires that the error term $u_i$ is uncorrelated with the regressors.
- This is a very strong assumption that is unreasonable in many applications.
- Here we need to assume that the regressors in the model explain enough of annual earnings that the unexplained part of earnings (the error) is uncorrelated with education
    - this is felt to be too strong an assumption.
- Instead, other methods need to be used.

# Causal Methods

- Here are leading causal methods, summarized in other slides
  - each method has its own assumptions and data needs.
- Randomized control trial (RCT)
  - provides a useful reference point but their use is limited in economics
- Control function
- Regression discontinuity design
- Instrumental variables (and LATE)
- Fixed effects for panel data and for grouped data
- Differences in differences
- Synthetic control
- Regression adjustment, inverse-probability weighting, matching methods
  - especially to balance unbalanced RCTs.

# Graphical Presentation

- Causal relationships can be subtle and a visual map can be helpful.
- Directed acylic graphs (DAGs) present all the paths from a causal variable D to outcome Y, including the role of any intermediate variables
    - Cunningham (2021), chapter 3, provides a good introduction
    - Judea Pearl (2009), *Causality*, Cambridge University Press, is the main reference.
- DAGs are unidirectional so do not handle all cases of causality
    - they do not handle reverse causality
    - they do not handle simultaneity, such as the demand-supply model
    - but they do handle most applications in modern microeconometrics.
- For models with simultaneity (and with latent (unobservable) variables) many social sciences use path diagrams for structural equation models.

# References for causal methods


- These books are given in approximate order of increasing difficulty.
- A. Colin Cameron (2020), Analysis of Economics Data: An Introduction to Econometrics, https://cameron.econ.ucdavis.edu/aed/.
- Joshua D. Angrist and Jörn-Steffen Pischke (2015), Mastering Metrics, Princeton University Press.
- Cunningham, Scott (2021), Causal Inference: The MixTape, Yale University Press.
- A. Colin Cameron and Pravin K. Trivedi (2022), Microeconometrics using Stata: Volumes 1 and 2, Second Edition, Stata Press, especially chapters 24 and 25.
- Joshua D. Angrist and Jörn-Steffen Pischke (2009), Mostly Harmless Econometrics: An Empiricist's Companion, Princeton University Press.
- A. Colin Cameron and Pravin K. Trivedi (2005), Microeconometrics: Methods and Applications, Cambridge University Press, especially chapter 25.
- Wooldridge, Jeffrey M. (2010), Econometric Analysis of Cross Section and Panel Data, Second Edition, MIT Press, especially chapters 20 and 21.
- Guido W. Imbens and Donald B. Rubin (2015), "Causal Inference in Statistics, Social, and Biomedical Sciences," Cambridge University Press.

# References by non-economists

- These books by non-economists are similar to *Mastering Metrics* in accessibility.
- Stephen L. Morgan and Christopher Winship (2015), Counterfactuals and Causal Inference: Methods and Principles for Social Research, Second edition, Cambridge University Press.
- Richard J. Murnane and John B. Willett (2010), Methods Matter: Improving Causal Inference in Educational and Social Science Research, Oxford University Press.
- Andrew Gelman, Jennifer Hill and Aki Vehtari (2022), Regression and Other Stories, Cambridge University Press, especially chapters 18-21.