Causal Inference

Bruno Arpino

Department of Statistical Sciences
University of Padua (Italy)

Course description

What is the effect of smoking on health? Does having an additional child increase the risk of poverty? Are development policies targeted on small firms effective in increasing investments?

Most studies in the social sciences are motivated by questions that are causal in nature.

However, in these areas experiments are not always possible because of ethical or practical reasons and the estimation of causal effects has often to rely on observational studies. The validity of inference will then strictly depend on the plausibility of the assumptions underlying the employed statistical techniques.

This course will cover some of the most popular techniques for estimating causal effects with observational data: propensity score matching, instrumental variable regression, fixed effects models, difference-in-differences designs. Special emphasis will be placed during the course on discussing the plausibility of the identifying assumptions, the data requirements and other practical and theoretical challenges for the implementation of each method.

This short course will offer participants theoretical and applied perspectives on the covered topics. Examples will be drawn from political science, sociology, economics, public health and policy evaluation. The implementation of the covered techniques will be demonstrated using the software Stata. References to examples in R will be also provided.

Software
Stata will be used to illustrate the implementation of the covered methods. Although familiarity with Stata is preferable, this is not strictly necessary. Easy to follow do-files and dataset to replicate the analyses discussed in class will be provided.

Prerequisites and course level
Familiarity with regression analysis and basic statistical concepts is strongly recommended. This in an intermediate level course in causal inference. The goal is to provide the fundamental tools to navigate in a growing field of methodological and empirical research, while also providing solid understanding on several specific methods.

Schedule

Day 1

- Introduction and course overview
    - Potential Outcome framework
    - Good Causal Inference is Theory + Data + Methods
    - Directed Acyclic Graphs: A useful graphical tool to visually represent causal models
    - Randomized experiments versus observational studies
    - Overview of statistical methods corresponding to different sets of assumptions
- Quick review of regression methods: pros and cons
- Propensity score matching (and similar methods):
    - Assumptions
    - Implementation
    - Examples from published research
    - Pros and cons
- Hands on!
    - Guided lab session on propensity score matching in Stata
- Summary and next class "trailer"

Day 2
- Brief resume of previous class
- Sensitivity analysis to the presence of unobserved confounders
- Instrumental Variable Regression:
    - Assumptions
    - Implementation
    - Examples from published research
    - Pros and cons
- Hands on!
    - Guided lab session on Instrumental Variable Regression in Stata
- Summary and next class "trailer"

Day 3
- Brief resume of previous class
- Fixed effects and Difference-in-Differences models
    - Assumptions
    - Implementation
    - Examples from published research
    - Pros and cons
- Hands on!
    - Guided lab session on Fixed Effects Models in Stata
- Summary
    - Key take home messages
    - Directions on current and future developments in causality for the social
      sciences

References:

Angrist, J. D., and Pischke, J. S. (2014). *Mastering' metrics: The Path from Cause to Effect. Princeton University Press.*

Arpino B., and Cannas, M. (2016) Propensity score matching with clustered data. An application to the estimation of the impact of caesarean section on the Apgar score, *Statistics in Medicine*. 35(12), 2074–2091.

Arpino B. and Aassve A. (2013) Estimation of causal effects of fertility on economic wellbeing: Data requirements, identifying assumptions and estimation methods, *Empirical Economics*, 44 (1), 355-385.

Becker, S.O. (2016). Using instrumental variables to establish causality. *IZA World of Labor*.

Cameron, A., & Trivedi, P. (2010). *Microeconometrics using Stata*. College Station: Stata Press.

Guo, S., and Fraser, M.S. (2009). *Propensity score analysis: Statistical methods and applications*. Sage Publications.

Imbens, G.W. & Rubin, D.B. (2015). *Causal inference for statistics, social, and biomedical sciences: An introduction*. Cambridge University Press.

Shahidur R. Khandker, Gayatri B. Koolwal and Hussain A. Samad (2010) *Handbook on Impact Evaluation. Quantitative Methods and Practices*, The World Bank, Washington D.C. Freely available on-line at:
http://documents.worldbank.org/curated/en/650951468335456749/pdf/520990PUB0EPI1101Official0Use0Only1.pdf

   [This handbook gives a (relatively) non-technical overview of all the topics covered
    in the course. It also offers Stata hints and exercises]

Short biography of the instructor

Bruno Arpino is a full professor in Social Statistics at the Department of Statistical Sciences, University of Padua (Italy). His previous appointments include positions at the University of Florence (Italy) and the Pomepu Fabra University (Spain) where he also co-directed the Research and Expertise Centre on Survey Methodology (RECSM). He obtained a PhD in Applied Statistics from the University of Florence in 2008 with a thesis on causal inference in multilevel observational studies that was awarded by the Italian Statistical Society as the best thesis in Applied Statistics 2007/2008.

His main research interests are in the areas of causal inference, applied statistics and social demography. From a substantive point of view, he has been studying intergenerational relationships, ageing and health, fertility and immigrants' assimilation.

He is an associate editor of the European Journal of Population and published articles in international peer-reviewed journals such as The Annals of Applied Statistics, Journal

of the Royal Statistical Society - A and C, Statistics in Medicine, Proceedings of the National Academy of Sciences (PNAS), Demography, The Journals of Gerontology: Series B, Journal of Marriage and Family, European Sociological Review.

Personal Web-page: https://sites.google.com/site/brunoarpino/
Google Scholar: https://scholar.google.com/citations?user=p6AcfaQAAAAJ&hl=en&oi=ao