Causal Inference for nonexperimental data

17-19 March 2021 (12 hours)

Bruno Arpino

Department of Statistics, Computer Science, Applications

University of Florence, Italy

Short biography of the instructor

Bruno Arpino is an associate professor at the Department of Statistics, Computer Science, Applications, University of Florence (Italy). Previously he was an Associate Professor at the Department of Political and Social Sciences, Universitat Pomepu Fabra (UPF) and co-director of the Research and Expertise Centre on Survey Methodology (RECSM, UPF). He obtained a PhD in Applied Statistics from the University of Florence in 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 has published articles in international peer-reviewed journals such as The Annals of Applied Statistics, Demography, European Sociological Review, The Journals of Gerontology: Series B, Journal of Marriage and Family, Journal of the Royal Statistical Society - A and C, Proceedings of the National Academy of Sciences (PNAS), Statistics in Medicine. Since October 2017 he is member of the Editorial Board of Statistical Methods and Applications.

Web-page: https://sites.google.com/site/brunoarpino/

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,

regression discontinuity designs and fixed effects models. 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. Lab sessions will demonstrate the implementation of the covered techniques using the software STATA.

Software
STATA will be used in the lab sessions. 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
Familiarity with regression analysis and basic statistical concepts.

Schedule

March 17, 2021

- Introduction and course overview
    - Potential Outcome framework
    - Good Causal Inference is Theory + Data + Methods
    - 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 with STATA
- Summary and next class “trailer”

March 18, 2021

- Brief resume of previous class
- How to estimate heterogeneous effects
- How to implement a 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
- Summary and next class “trailer”

March 19, 2021

- Brief resume of previous class
- Introduction to Regression Discontinuity Designs
- Fixed effects models
    - Assumptions
    - Implementation
    - Examples from published research
    - Pros and cons
- Hands on!
    - Guided lab session on Fixed Effects Models
- Nonstandard data structures and more advanced topics
- Summary
    - Key take home messages
    - Directions on current and future developments in causality for the social sciences

Background reading: While not mandatory, the following papers are useful background readings for the course:

Holland, P. (1986) "Statistics and Causal Inference", with discussion and rejoinder. *Journal of the American Statistical Association*, 81, 945-970.

Reiter, J. P. (2000) "Using statistics to determine causal relationships." *The American Mathematical Monthly*, 107, pp. 24-32.

References:

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