Introduction to Causal Inference

Michael Kühhirt
michael.kuehhirt@uni-koeln.de

Centre d'Estudis Demogràfics
Universitat Autònoma Barcelona

14th and 15th of September 2017
10.00am to 5:45pm

Carrer de Ca n'Altayó, Edifici E2,
08193 Bellaterra, Barcelona

More has been learned about causal inference in the last few decades than the sum total of everything that had been learned about it in all prior recorded history.
— Gary King, 2015.

DESCRIPTION

The course provides an introduction to recent advances in causal inference. After providing a formal definition of causal effects based on graphical causal models and the potential outcome framework ("Rubin-Causal-Model"), we discuss the basic logic of different research designs for causal inference and their underlying substantive assumptions. The course concludes with an overview of causal mediation analysis, that is, the investigation of the mechanisms that produce a specific causal effect. For each topic, theoretical lectures are supplemented by paper-and-pencil exercises or practical computer applications using *Stata*.

The only necessary participation prerequisite to fully benefit from this workshop is a solid understanding of basic statistics (distributions, probability, standard error, confidence intervals etc.) and common regression techniques (OLS and binary outcome regression).

SCHEDULE

Thursday 14th of September 2017

10.00am	–	11.30am	Graphical causal models, counterfactuals, and covariate adjustment
11.45am	–	13.15pm	Randomised controlled trials
2.30pm	–	4.00pm	Instrumental variables
4.15pm	–	5.45pm	Regression discontinuity designs

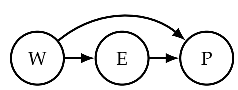
Friday 15th of September 2017

10.00am	–	11.30am	Multilevel and longitudinal designs
11.45am	–	13.15pm	Causal mediation analysis I
2.30pm	–	4.00pm	Causal mediation analysis II
4.15pm	–	5.45pm	Limitations and current frontiers, round-up

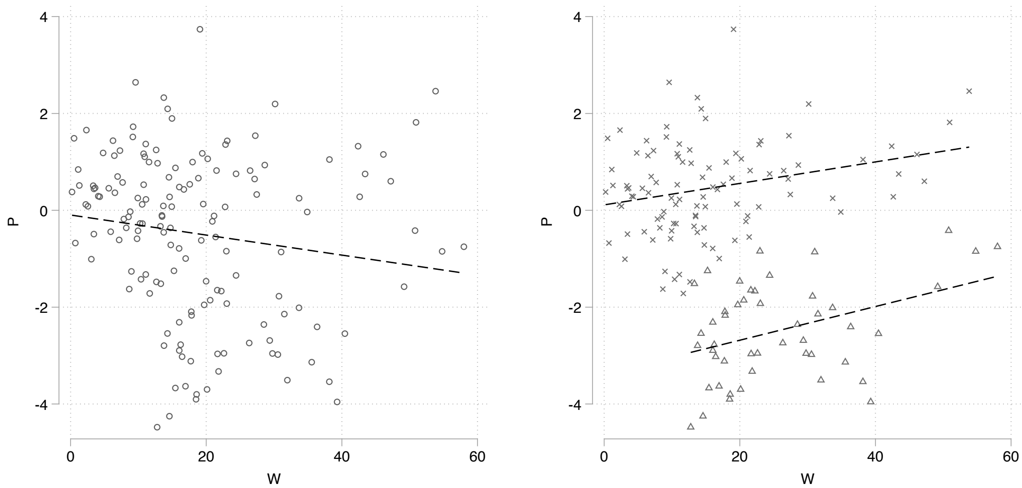
EXERCISES

1. Graphical causal models, counterfactuals, and covariate adjustment

   (a) Suppose you have data on three variables, W, P (both continuous) and E (binary) and are interested in the average total effect of W on P (which you believe to be positive). You hypothesise the following causal relations between the three variables:



   The following figures show the marginal association between W and P (left) and the conditional association given E (right). Which figure should be consulted to learn about the average total effect of W on P, given the graphical model above?



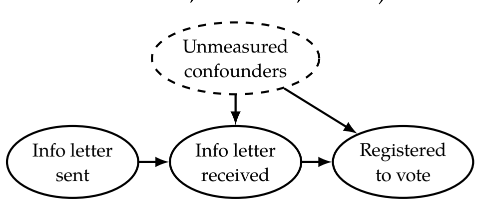
(b) Build the main model of the session (e.g, slide 4) in *DAGitty*. What are the sufficient adjustment sets for the total effect of A on E and of B on F?

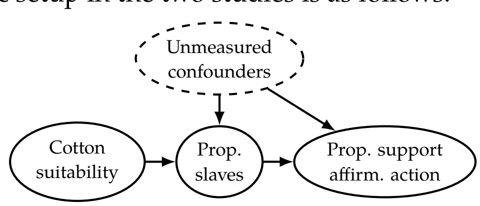
2. Randomised controlled trials

   (a) Think of your own current research. Can you think of an (ideal) RCT to answer one of the key research questions? What would the treatment values be? How would assignment be carried out? What is the target population?

   (b) Would there be reason to expect interference, noncompliance, or side effects of treatment assignment? What could be done to avoid or attenuate these problems?

3. Instrumental variables
   The basic steps of an instrumental variables analysis will be illustrated using edited versions of publicly available data from a field experiment on the political reintegration of felons (Gerber et al., 2015) and an observational study on the heritage of slavery in Southern counties of the United States (Acharya, Blackwell, and Sen, 2016b).[1] The basic setup in the two studies is as follows:





[1]The complete replication data for these articles are available at `https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/27241` and `https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/CAEEG7`.

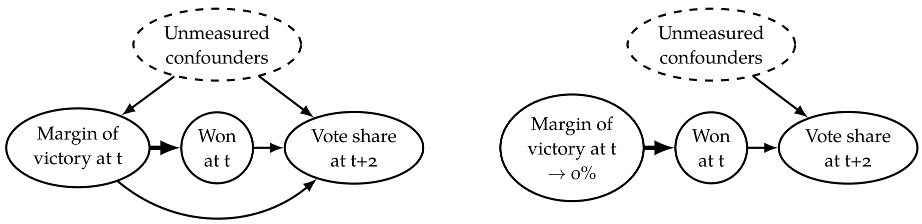
(a) Can you think of possible violations of IV assumptions ii. and iii. in each case?

(b) Open *Stata* and the dofile `03ivlab.do`, and execute the commands in the preamble (#0).

(c) Load the data `03ivlab1.dta`, successively run the commands for description, tests and analyses (#1–5), and interpret the results.

(d) Repeat with `03ivlab2.dta` (#6–10).

4. Regression discontinuity designs
Analysing an RDD will be illustrated with edited state-level data on United States Senate elections for the period 1914–2010, which can be used to investigate the incumbent party advantage (Cattaneo, Frandsen, and Titiunik, 2015).[2] The basic setup for this study looks like this:

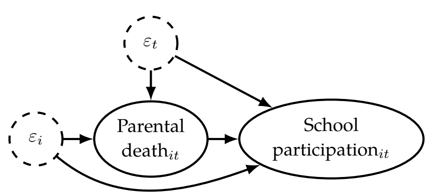


(a) Can you think of possible violations of RDD assumptions ii. and iii.?

(b) Open *Stata* and the dofile `04rdlab.do`, and execute the commands in the preamble (#0).

(c) Load the data `04rdlab.dta`, successively run the commands for description, tests and analyses (#1-6), and interpret the results.

5. Longitudinal designs
Edited replication data from a study on the effect of parental death on school participation in Kenya (Evans and Miguel, 2007) are used to demonstrate different longitudinal designs and modelling strategies.[3] The basic setup for this study looks like this:



[2]The original data are available at `https://sites.google.com/site/rdpackages/rdrobust/stata/rdrobust_senate.dta?attredirects=0`.

[3]The complete replication data are available at `https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/VMWYWJ`.

(a) Open *Stata* and the dofile `05mllab.do`, and execute the commands in the preamble (#0)

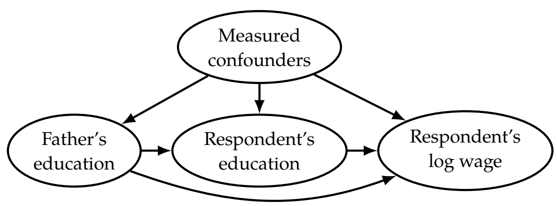
(b) Load the data `05mllab.dta`.

(c) Can you think of time-varying variables that could affect both parental death and school participation?

(d) Successively run the commands for description, tests and analyses (#1-4), and interpret the results.

(e) Given the pre-treatment development of school participation, which design/estimator should we favour?

6. Causal mediation analysis
Different methods for causal mediation analysis will be illustrated with an edited version of training data based on the NLS 1980 and provided with Wooldridge (2010).[4] With these data, we can analyse whether fathers' education affects respondents' (log) wages and whether the effect is (partly) mediated by respondents' own education (after adjusting for several measured covariates). The basic setup for the analysis looks like this:



(a) Open *Stata* and the dofile `06malab.do`, and execute the commands in the preamble (#0)

(b) Load the data `06malab.dta`.

(c) Can you think of variables not included in the data that may be confounders (violation of mediation assumptions i.-iii.)?

(d) Could some of the variables included in the data themselves be affected by father's education (violation of mediation assumption iv.)?

(e) Successively run the commands for description, tests and analyses (#1-5), and interpret the results.

[4]Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data. Second Edition*. Cambridge, MA: MIT Press. The original training data can be downloaded at `http://fmwww.bc.edu/ec-p/data/wooldridge/nls80.dta`.

SELECTED LITERATURE (* PARTICULAR RECOMMENDATION)

Causal inference and research design

Book length

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Angrist, J. D. and Pischke, J.-S. (2015). *Mastering 'Metrics. The Path from Cause to Effect*. Princeton, NJ: Princeton University Press.

*Hernán, M. A. and Robins, J. M. (2016). *Causal Inference (v. 09-11-16)*. Boca Raton, FL: Chapman & Hall/CRC. URL: http://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/.

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences: An Introduction*. New York: Cambridge University Press.

Morgan, S. L. and Winship, C. (2015). *Counterfactuals and Causal Inference: Methods and Principles for Social Research. Second Edition*. New York: Cambridge University Press.

Murnane, R. J. and Willett, J. B. (2010). *Methods Matter: Improving Causal Inference in Educational and Social Science Research*. Oxford University Press.

*Pearl, J., Glymour, M., and Jewell, N. P. (2016). *Causal Inference in Statistics: A Primer*. West Sussex, UK: Wiley.

Shadish, W., Cook, T., and Campbell, D. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Belmont, CA: Wadsworth Cengage Learning.

Article length

Gangl, M. (2010). "Causal Inference in Sociological Research". In: *Annual Review of Sociology* 36, pp. 21–47.

Imbens, G. W. and Wooldridge, J. M. (2009). "Recent Developments in the Econometrics of Program Evaluation". In: *Journal of Economic Literature* 47 (1), pp. 5–86.

Keele, L. (2015b). "The statistics of causal inference: A view from political methodology". In: *Political Analysis* 23 (3), pp. 313–335. DOI: 10.1093/pan/mpv007.

Pearl, J. (2009). "Causal inference in statistics: An overview". In: *Statistics Surveys* 3, pp. 96–146. DOI: 10.1214/09-SS057.

*Petersen, M. L. and Laan, M. J. van der (2014). "Causal models and learning from data: Integrating causal modeling and statistical estimation". In: *Epidemiology* 25 (3), pp. 418–426. DOI: 10.1097/EDE.0000000000000078.

Graphical causal models

Book length

Glymour, C. (2001). *The Mind's Arrows: Bayes Nets and Graphical Causal Models in Psychology*. Cambridge, MA: MIT Press.

Pearl, J. (2009[2000]). *Causality: Models, Reasoning, and Inference. Second Edition*. New York: Cambridge University Press.

Sloman, S. (2005). *Causal Models: How People Think About the World and its Alternatives*. Oxford, UK: Oxford University Press.

Spirtes, P., Glymour, C., and Scheines, R. (2001[1993]). *Causation, Prediction, and Search. Second Edition*. Cambridge, MA: MIT Press.

Article length

*Elwert, F. (2013). "Graphical causal models". In: *Handbook of Causal Analysis for Social Research*. Ed. by Morgan, S. L. New York: Springer, pp. 245–272. DOI: 10.1007/978-94-007-6094-3_13.

*Glymour, M. M. and Greenland, S. (2008). "Causal diagrams". In: *Modern Epidemiology. Third Edition*. Ed. by Rothman, K. J., Greenland, S., and Lash, T. L. Philadelphia, PA: Lippincott Williams & Wilkins, pp. 183–209.

Steiner, P. M. et al. (2017). "Graphical models for quasi-experimental designs". In: *Sociological Methods & Research* 46 (2), pp. 155–188. DOI: 10.1177/0049124115582272.

Randomised controlled trials

*Deaton, A. and Cartwright, N. (2016). "Understanding and misunderstanding randomized controlled trials". In: *NBER Working Paper* (22595). URL: http://www.nber.org/papers/w22595.pdf.

Jackson, M. and Cox, D. R. (2013). "The Principles of Experimental Design and Their Application in Sociology". In: *Annual Review of Sociology* 39, pp. 27–49. DOI: 10.1146/annurev-soc-071811-145443.

Sampson, R. J. (2010). "Gold standard myths: Observations on the experimental turn in quantitative criminology". In: *Journal of Quantitative Criminology* 26 (4), pp. 489–500. DOI: 10.1007/s10940-010-9117-3.

Applications

Barone, C. et al. (2017). "Information barriers, social inequality, and plans for higher education: Evidence from a field experiment". In: *European Sociological Review* 33 (1), pp. 84–96. DOI: 10.1093/esr/jcw050.

Gerber, A. S. et al. (2015). "Can incarcerated felons be (re)integrated into the political system? Results from a field experiment."". In: *American Journal of Political Science* 59 (4), pp. 912–926. DOI: 10.1111/ajps.12166.

Instrumental variables

Theory

Bollen, K. A. (2012). "Instrumental Variables in Sociology and the Social Sciences". In: *Annual Review of Sociology, Vol 38* 38, pp. 37–72. DOI: 10.1146/annurev-soc-081309-150141.

*Glymour, M. M. (2006). "Natural experiments and instrumental variable analyses in social epidemiology". In: *Methods in Social Epidemiology*. Ed. by Oakes, J. M. and Kaufman, J. S. San Francisco, CA: Jossey-Bass, pp. 393–428.

Hernán, M. A. and Robins, J. M. (2006). "Instruments for causal inference. An epidemiologist's dream?" In: *Epidemiology* 17 (4), pp. 360–372. DOI: 10.1097/01.ede.0000222409.00878.37.

Imbens, G. W. (2014). "Instrumental variables: An econometrician's perspective (with comments)". In: *Statistical Science* 29 (3), pp. 323–379. DOI: 10.1214/14-STS480.

Analysis

Glymour, M. M., Tchetgen Tchetgen, E. J., and Robins, J. M. (2012). "Credible Mendelian Randomization Studies: Approaches for Evaluating the Instrumental Variable Assumptions". In: *American Journal of Epidemiology* 175 (4), pp. 332–339. DOI: 10.1093/aje/kwr323.

Muller, C., Winship, C., and Morgan, S. L. (2015). "Instrumental variables regression". In: *The Sage Handbook of Regression Analysis and Causal Inference*. Ed. by Best, H. and Wolf, C. Los Angeles: Sage, pp. 277–299.

Palmer, T. A. et al. (2011). "Nonparametric bounds for the causal effect in a binary Instrumental-variable model". In: *Stata Journal* 11 (3), pp. 345–367.

*Swanson, S. A. and Hernán, M. A. (2013). "How to report instrumental variable analyses (suggestions welcome)". In: *Epidemiology* 24 (3), pp. 370–374. DOI: 10.1097/EDE.0b013e31828d0590.

Applications

Acharya, A., Blackwell, M., and Sen, M. (2016b). "The political legacy of American slavery". In: *Journal of Politics* 78 (3), pp. 621–641. DOI: 10.1086/686631.

Kern, H. L. and Hainmueller, J. (2009). "Opium for the Masses: How Foreign Media Can Stabilize Authoritarian Regimes". In: *Political Analysis* 17 (4), pp. 377–399. DOI: 10.1093/pan/mpp017.

Kirk, D. S. (2009). "A Natural Experiment on Residential Change and Recidivism: Lessons from Hurricane Katrina". In: *American Sociological Review* 74 (3), pp. 484–505. DOI: 10.1177/000312240907400308.

Regression discontinuity designs

Theory

Imbens, G. W. and Lemieux, T. (2008). "Regression discontinuity designs: A guide to practice". In: *Journal of Econometrics* 142, pp. 615–635. DOI: 10.1016/j.jeconom.2007.05.001.

*Lee, D. S. and Lemieux, T. (2010). "Regression discontinuity designs in economics". In: *Journal of Economic Literature* 48 (2), pp. 281–355. DOI: 10.1257/jel.48.2.281.

Analysis

Calonico, S. et al. (forthcoming). "rdrobust: Software for regression discontinuity designs". In: *Stata Journal*. URL: http://www-personal.umich.edu/~cattaneo/papers/Calonico-Cattaneo-Farrell-Titiunik_2017_Stata.pdf.

Applications

Bernardi, F. (2014). "Compensatory advantage as a mechanism of educational inequality: A regression discontinuity design based on month of birth". In: *Sociology of Education* 87 (2), pp. 74–88. DOI: 10.1177/0038040714524258.

Cattaneo, M. D., Frandsen, B. R., and Titiunik, R. (2015). "Randomization inference in the regression discontinuity design: An application to party advantages in the U.S. senate". In: *Journal of Causal Inference* 3 (1), pp. 1–24. DOI: 10.1515/jci-2013-0010.

Hainmueller, J., Hangartner, D., and Pietrantuono, G. (2017). "Catalyst or crown: Does naturalization promote the long-term social integration of immigrants?" In: *American Political Science Review*. DOI:

10.1017/S0003055416000745.

Loeffler, C. E. and Grunwald, B. (2015). "Processed as an adult: A regression discontinuity estimate of the crime effects of charging nontransfer juveniles as adults". In: *Journal of Research in Crime and Delinquency* 52 (6), pp. 890–922. DOI: 10.1177/0022427815581858.

Multilevel and longitudinal designs

Theory

*Brüderl, J. and Ludwig, V. (2015). "Fixed-effects panel regression". In: *The Sage Handbook of Regression and Causal Inference*. Ed. by Best, H. and Wolf, C. Los Angeles: Sage, pp. 327–358.

*Dafoe, A. (2016). "Nonparametric identification of causal effects under temporal dependence". In: *Sociological Methods & Research* OnlineFirst Version. DOI: 10.1177/0049124115613784.

O'Neill, S. et al. (2016). "Estimating causal effects: Considering three alternatives to difference-in-differences estimation". In: *Health Services and Outcomes Research Methodology* 16 (1), pp. 1–21. DOI: 10.1007/s10742-016-0146-8.

Rattigan, M. J. H. and Jensen, D. (2010). "Leveraging d-separation for relational data sets". In: *IEEE 10th International Conference on Data Mining*. DOI: 10.1109/ICDM.2010.142.

Rattigan, M. J. H., Maier, M., and Jensen, D. (2011). "Relational blocking for causal discovery". In: *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence*, pp. 145–151. URL: http://www.aaai.org/ocs/index.php/AAAI/AAAI11/paper/viewFile/3760/3843/.

Sjölander, A. et al. (2016). "Carryover effects in sibling comparison designs". In: *Epidemiology* 27 (6), pp. 852–858. DOI: 10.1097/eDe.0000000000000541.

*Vaisey, S. and Miles, A. (2017). "What you can—and can't—do with three-wave panel data". In: *Sociological Methods & Research* 46 (1), pp. 44–67. DOI: 10.1177/0049124114547769.

Analysis

Correia, S. (2016). "A feasible estimator for linear models with multi-way fixed effects". In: URL: http://scorreia.com/research/hdfe.pdf.

Applications

Elwert, F. and Christakis, N. A. (2008). "Wives and Ex-Wives: A New Test for Homogamy Bias in the Widowhood Effect". In: *Demography* 45 (4), pp. 851–873. ISSN: 0070-3370. DOI: 10.1353/dem.0.0029.

Evans, D. K. and Miguel, E. (2007). "Orphans and schooling in Africa: A longitudinal analysis". In: *Demography* 44 (1), pp. 35–57. DOI: 10.1353/dem.2007.0002.

Desmond, M., Papachristos, A. V., and Kirk, D. S. (2016). "Police violence and citizen crime reporting in the Black community". In: *American Sociological Review* 81 (5), pp. 857–876. DOI: 10.1177/0003122416663494.

Killewald, A. and Lundberg, I. (2017). "New evidence against a causal marriage wage premium". In: *Demography*. DOI: 10.1007/s13524-017-0566-2.

Causal mediation analysis

Theory

Hong, G. (2015). *Causality in a Social World: Moderation, Mediation, and Spill-Over*. West Sussex, UK: Wiley-Blackwell.

Keele, L. (2015a). "Causal mediation analysis: Warning! Assumptions ahead". In: *American Journal of Evaluation* 36 (4), pp. 500–513. DOI: 10.1177/1098214015594689.

*Knight, C. and Winship, C. (2013). "The causal implications of mechanistic thinking: Identification using directed acyclic graphs (DAGs)". In: *Handbook of Causal Analysis for Social Research*. Ed. by Morgan, S. L. Dordrecht u.a.: Springer, pp. 275–299. DOI: 10.1007/978-94-007-6094-3_14.

Pearl, J. (2005). "Direct and indirect effects". In: *Proceedings of the American Statistical Association Joint Statistical Meetings*, pp. 1572–1581.

Robins, J. M. and Greenland, S. (1992). "Identifiability and exchangeability for direct and indirect effects". In: *Epidemiology* 3 (2), pp. 143–155.

*VanderWeele, T. J. (2015). *Explanation in Causal Inference: Methods for Mediation and Interaction*. New York: Oxford University Press.

Analysis

Hicks, R. and Tingley, D. (2011). "Causal mediation analysis". In: *Stata Journal* 11 (4), pp. 605–619.

Liu, H. and Emsley, R. A. (forthcoming). "PARAMED: Stata module to perform causal mediation analysis using parametric regression models". In: *Stata Journal*. URL: http://econpapers.repec.org/software/bocbocode/s457581.htm.

Tingley, D. et al. (2014). "mediation: R package for causal mediation analysis". In: *Journal of Statistical Software* 59 (5), pp. 1–38. DOI: 10.18637/jss.v059.i05.

Valeri, L. and VanderWeele, T. J. (2013). "Mediation analysis allowing for exposure-mediator interactions and causal interpretation: Theoretical assumptions and implementation with SAS and SPSS Macros". In: *Psychological Methods* 18 (2), pp. 137–150. DOI: 10.1037/a0031034.

Applications

Acharya, A., Blackwell, M., and Sen, M. (2016a). "Explaining causal findings without bias: Detecting and assessing direct effects". In: *American Political Science Review* 110 (3), pp. 512–529. DOI: 10.1017/S0003055416000216.

Wodtke, G. T. and Parbst, M. (2017). "Neighborhoods, schools, and academic achievement: A formal mediation analysis of contextual effects on reading and mathematics abilities". In: *Demography*. DOI: 10.1007/s13524-017-0603-1.