

SOC6708H1: CAUSAL INFERENCE IN THE SOCIAL SCIENCES

University of Toronto

Term: Fall 2017 – Lecture Date/Time: [DAY/TIME] – Location: Room 240, 725 Spadina

INSTRUCTOR

Geoff Wodtke

Office: Room 394, 725 Spadina Ave

Office Hours: [DAY/TIME] or by appointment

Email: geoffrey.wodtke@utoronto.ca

COURSE DESCRIPTION

Social scientists regularly ask causal questions. For example, does growing up in a poor neighborhood impede a child's progress in school? Does job training result in higher earnings? Do larger class sizes lead to lower student achievement? This course explores methods of causal inference in the social sciences. These methods and examples of their application will be drawn from a wide variety of disciplines, including sociology, economics, statistics, psychology, and epidemiology. Most of the emphasis in this course is on methods of causal inference from quasi-experimental and observational research designs. It draws on the counterfactual (or potential outcomes) framework, the do-calculus framework, and graphical methods to conceptualize causal questions and define causal estimands, to determine whether a given causal estimand can be identified from observed data, and to guide the selection of an appropriate estimator.

This course is divided into four sections. The first section covers foundational material that will be used throughout the remainder of the course, including the counterfactual framework, identification via do-calculus and directed acyclic graphs, and basic experimental research designs. The second section covers methods of causal inference with observational research designs that require the strong assumption of no unobserved selection into the different treatments, exposures, or stimuli whose causal effects we wish to estimate. The third section covers a set of “quasi-experimental” methods that relax this strong assumption when a set of other conditions are satisfied. The final section covers a series of special topics in causal inference that are particularly relevant to sociologists, including effect moderation and effect interaction; causal mediation and effect decomposition; and methods for estimating the causal effects of time-varying treatment trajectories.

PREREQUISITES

This is an advanced graduate-level course. The prerequisites to take this course are SOC6302H1 (Statistics for Sociologists) and SOC6707 (Intermediate Data Analysis) or an equivalent set of applied statistics courses. Students should have a solid understanding of statistical inference, linear regression models and least squares estimation, and generalized linear models and maximum likelihood estimation. Knowledge of probability theory, linear algebra, and calculus will be an asset but is not required. This course also involves several homework assignments that require proficiency in the statistical software package Stata.

TEXTS

Required

This course is divided into weekly topics. Most of these topics are covered in assigned readings from the following required course texts:

Angrist, Joshua D. and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics*. Princeton, NJ: Princeton University Press.

Morgan, Stephen L. and Christopher Winship. 2015. *Counterfactuals and Causal Inference (2nd Edition)*. New York: Cambridge University Press.

VanderWeele, Tyler J. 2015. *Explanation in Causal Inference*. New York: Oxford University Press.

Additional readings will be made available on the course website as needed.

Supplementary

Although not required, there are several other texts that you may find helpful in this course and/or in your own research. Two supplementary texts that I recommend are:

Cameron, A. Colin and Pravin K. Trivedi. 2010. *Microeconometrics Using Stata (Revised Edition)*. College Station, TX: Stata Press.

Long, Scott and Jeremy Freese. 2014. *Regression Models for Categorical Dependent Variables Using Stata (3rd Edition)*. College Station, TX: Stata Press.

SOFTWARE

All statistical computing for this course will be done using Stata (<http://www.stata.com>). Stata is command driven—that is, computations are executed from a set of typed commands in a text file. For a brief introduction to Stata's command language, see <http://data.princeton.edu/stata/> or <http://www.ats.ucla.edu/stat/stata/modules/default.htm>. Microsoft Word's Equation Editor Tool (http://word.tips.net/T001419_Adding_an_Equation_Editor_Tool.html) or the LaTeX typesetting system (<http://www.latex-project.org/>) may also prove helpful for your assignments.

EVALUATION

Homework Assignments

Students will be asked to complete three homework assignments throughout the semester. These assignments will generally have about 5 to 10 questions or problems. Some questions will ask you to prove or demonstrate certain properties of a method, while others will ask you to apply a method using Stata with real or simulated data and interpret the results. You are expected to complete these assignments individually, although some consultation among classmates is normal and expected. Your assignments should be double-spaced, in 12-point font, and where appropriate, accompanied by a clearly demarcated Stata log file containing the necessary statistical output. Electronic copies of

the assignments will not be accepted—students must hand in a hard copy. These assignments will be due in class, and altogether they will count for 50 percent of your final grade. After each individual assignment is marked and returned, students will have one or two weeks to re-write incorrect answers for partial credit if they so choose. Late assignments will not be accepted.

Final Research Project

Students will be asked to complete a final research project. The project may consist of an empirical study with a causal research question that uses one of the methods covered in this course, or it can cover a conceptual or methodological issue related to causal inference. Students are strongly encouraged to develop an existing project using methods covered in this course or to use this assignment as an opportunity to launch a new empirical project. The paper should include a brief description of the background and motivation for the study, a well-defined causal question, and a detailed description of the methods and data. The final written report should be 10 to 20 pages in length, double-spaced, and in 12-point font. Students are also required to submit a full set of replication files for the results reported in their papers (i.e., the original data together with a set of programs or executable scripts that generate the reported findings). A one-page proposal for the final research project is due on [DATE]. The final report is due on [DATE] and will count for 50 percent of your final grade.

Grades

Grades for this course will be assigned as follows: homework assignments (50 percent) and final research project (50 percent).

SCHEDULE

Part I: Defining and Identifying Causal Effects

Week 1: Counterfactuals, Potential Outcomes, and Ignorability

Reading:

Holland, P. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association*, 81: 945-960.

Morgan, S. L. and C. Winship. 2015. *Counterfactuals and Causal Inference (2nd Edition)*. New York: Cambridge University Press. Read pg. 37-76.

Week 2: Do-calculus, D-separation, and Graphical Identification

Reading:

Morgan, S. L. and C. Winship. 2015. *Counterfactuals and Causal Inference (2nd Edition)*. New York: Cambridge University Press. Read pg. 77-130.

Elwert, F. 2013. "Graphical Causal Models." Pp. 245-73 in *Handbook of Causal Analysis for Social Research* edited by S. L. Morgan. New York: Springer.

New assignments: Homework 1 (Defining and Identifying Causal Effects)

Week 3: Randomized Experiments

Reading:

Rubin, D. B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology*, 66:688-701.

Heckman, J, and J. A. Smith. 1995. "Assessing the Case for Social Experiments." *Journal of Economic Perspectives* 9:85-110.

Sanbonmatsu, L., J. R. Kling, G. J. Duncan and J. Brooks-Gunn. 2006. "Neighborhoods and Academic Achievement: Results from the Moving to Opportunity Experiment." *Journal of Human Resources* 41:649-91.

Part II: Estimating Causal Effects when the Treatment Selection Process is Observed

Week 4: Matching Estimators

Reading:

Morgan, S. L. and C. Winship. 2015. *Counterfactuals and Causal Inference (2nd Edition)*. New York: Cambridge University Press. Read pg. 140-150, 158-187.

Haskins, Anna R. 2014. "Unintended Consequences: Effects of Paternal Incarceration on Child School Readiness and Later Special Education Placement." *Sociological Science* 1:141-158.

Assignments due: Homework 1 (Defining and Identifying Causal Effects)

Week 5: Weighting Estimators

Reading:

Morgan, S. L. and C. Winship. 2015. *Counterfactuals and Causal Inference (2nd Edition)*. New York: Cambridge University Press. Read pg. 150-158.

Lunceford, J. K. and M. Davidian. 2004. "Stratification and Weighting via the Propensity Score in Estimation of Causal Treatment Effects: A Comparative Study." *Statistics in Medicine* 23:2937-60.

Desmond, M. and R. T. Kimbro. 2015. "Eviction's Fallout: Housing, Hardship, and Health." *Social Forces* 94:295-324.

Week 6: Regression Estimators

Reading:

Angrist, J. D. and J. Pischke. 2009. *Mostly Harmless Econometrics*. Princeton, NJ: Princeton University Press. Read pg. 27-91.

Morgan, S. L. and C. Winship. 2015. *Counterfactuals and Causal Inference (2nd Edition)*. New York: Cambridge University Press. Read pg. 188-225.

Snowden, J. M., S. Rose, and K. M. Mortimer. 2011. "Implementation of G-computation on a Simulated Dataset: Demonstration of a Causal Inference Technique." *American Journal of Epidemiology* 173:731-738.

New assignments: Homework 2 (Matching, Weighting, and Regression)

Part III: Estimating Causal Effects when the Treatment Selection Process is Unobserved

Week 7: Instrumental Variable Estimators

Reading:

Angrist, J. D. and J. Pischke. 2009. *Mostly Harmless Econometrics*. Princeton, NJ: Princeton University Press. Read pg. 113-218.

Ananat, E. O. 2011. "The Wrong Side(s) of the Tracks: The Causal Effects of Racial Segregation on Urban Poverty and Inequality." *American Economic Journal: Applied Economics* 3: 34-66.

Week 8: Fixed-effects and Difference-in-difference Estimators

Reading:

Angrist, J. D. and J. Pischke. 2009. *Mostly Harmless Econometrics*. Princeton, NJ: Princeton University Press. Read pg. 221-247.

Halaby, C. N. 2004. "Panel Models in Sociological Research: Theory into Practice." *Annual Review of Sociology* 30:507-44.

Budig M. J. and P. England. 2001. "The Wage Penalty for Motherhood." *American Sociological Review* 66:204-225.

Assignments due: Homework 2 (Matching, Weighting, and Regression)

Week 9: Regression Discontinuity Estimators

Reading:

Angrist, J. D. and J. Pischke. 2009. *Mostly Harmless Econometrics*. Princeton, NJ: Princeton University Press. Read pg. 251-267.

Angrist, J. D. and V. Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics* 114:533-602.

Assignments due: Proposal for Final Research Project

New assignments: Homework 3 (Instrumental Variables and Extensions)

Part IV: Special Topics

Week 10: Effect Moderation and Effect Interaction

Reading:

VanderWeele, T. J. 2009. "On the Distinction Between Interaction and Effect Modification." *Epidemiology* 20:863-871.

VanderWeele, T. J. 2015. *Explanation in Causal Inference*. New York: Oxford University Press. Read pg. 249-285.

Pager, D. 2003. "The Mark of a Criminal Record." *American Journal of Sociology* 108:937-975.

Week 11: Causal Mediation and Effect Decomposition

Reading:

VanderWeele, T. J. 2015. *Explanation in Causal Inference*. New York: Oxford University Press. Read pg. 20-65, 113-152, 459-467.

Wodtke, G. T. and M. Parbst. 2017. "Neighborhoods, Schools, and Academic Achievement: A Formal Mediation Analysis of Contextual Effects on Reading and Mathematics Abilities." *Demography*.

Assignments due: Homework 3 (Instrumental Variables and Extensions)

Week 12: Estimating the Causal Effects of Time-varying Treatment Trajectories

Reading:

Robins, J. M., M. A. Hernan, and B. Brumback. 2000. "Marginal Structural Models and Causal Inference in Epidemiology." *Epidemiology* 11:550-560.

Wodtke, G. T., D. J. Harding, and F. Elwert. 2011. "Neighborhood Effects in Temporal Perspective." *American Sociological Review* 76:713-736.

Wodtke, G. T. 2017. "Regression-based Adjustment for Time-varying Confounders." University of Toronto, Department of Sociology, Working Paper No. 2017-01.

December 18, 2017: *FINAL RESEARCH PROJECTS DUE*