gesis Leibniz Institute for the Social Sciences

GESIS Summer School in Survey Methodology 2023

Syllabus for course:
"Causal Inference Using Survey Data"

Lecturers:	Heinz Leitgöb	Tobias Wolbring
E-mail:	heinz.leitgoeb@uni-leipzig.de	tobias.wolbring@fau.de
Website:	http://sozweb.sozphil.uni-leipzig.de/de/personen.html	https://www.wirtschaftssoziologie.rw.fau.eu/person/tobias-wolbring/

Date: 07 – 11 August 2023
Time: Mo: 10:00-17:00 | Tu-Thu: 09:00 – 16:30 | Fr: 09:00 – 16:00
Venue: On-site at GESIS Cologne

About the Lecturers:

Heinz Leitgöb is an Interim Professor of Quantitative Research Methods at the Institute of Sociology, Leipzig University, Germany. He is also associated at the Goethe University of Frankfurt, Germany. He chairs the methods section of the German Sociological Association and the European Working Group on Quantitative Methods in Criminology of the European Society of Criminology. Besides causal inference, his research interests cover the topics of rare events modeling, measurement invariance, survey mode effects, digitalization, and quantitative criminology.

Tobias Wolbring is a Professor of Empirical Economic Sociology at School of Business, Economics, and Society at FAU Erlangen-Nürnberg, Germany. He is particularly focussed on applying statistical methods for panel data and on running different kinds of experiments to draw causal inference. His research covers a wide range of topics including labour market behaviour, discrimination, social inequality effects, the validity of students' evaluations of teaching, and the determinants of subjective well-being.

Selected Publications:

- Barrera, D./Gërxhani, K./Kittel, B./Miller, L./**Wolbring, T.** (forthcoming). Experimental Sociology: Outline of a Scientific Field. Cambridge: Cambridge University Press. Planned year of publication: 2023.
- Eberl, A./Collischon, M./**Wolbring, T.** (forthcoming). Subjective Well-Being Scarring through Unemployment: New Evidence from a Long-running Panel. In: Social Forces. doi: 10.1093/sf/soac022
- Friedrich, S./Antes, G./Behr, S./Binder, H./Brannath, W./Dumpert, F./Lederer, J./**Leitgöb, H.**/Ickstadt, K./Kestler, H./Pauly, M./Steland, A./Wilhelm, A./Friede, T. (2022). Is There a Role for Statistics in Artificial Intelligence? Advances in Data Analysis & Classification, 16: 823–846. doi: 10.1007/s11634-021-00455-6
- **Leitgöb, H.** (2021). Analysis of Rare Events. In Atkinson, P. A./ Delamont, S. /Cernat, A./Sakshaug, J. W./Williams, R. A. (eds.): Sage Research Methods Foundations, Vol. 1. Thousand Oaks: Sage, pp. 161–174
- **Leitgöb, H.**/Seddig, D./Asparouhov, T./Behr, D./Davidov, E./De Roover, K./Jak, S./Meitinger, K./Menold, N./Muthén, B./Rudnev, M./Schmidt, P./van de Schoot, R. (2023). Measurement Invariance in the Social Sciences: Historical Development, Methodological Challenges, State of the Art, and Future Perspectives. In: Social Science Research, 110: 102805. doi: 10.1016/j.ssresearch.2022.102805

- Leszczensky, L./**Wolbring, T.** (2022). How to Deal With Reverse Causality Using Panel Data? Recommendations for Researchers Based on a Simulation Study. Sociological Methods & Research, 51: 837–865.

Course Description:

This course will introduce participants to the concepts and methods of causal inference and causal modeling in the social sciences. It will highlight the relevance of research design, analytical methods *and* their systematic combination to optimize the validity of causal inferences drawn from empirical studies. Participants will learn the key principles and techniques of causal inference, including potential outcomes, counterfactuals, and causal graphs, and will get to know the experimental approach to causality. Building on existing knowledge concerning linear regression modelling and research design, the course will then cover key methods of causal modeling using survey data, such as fixed effects panel models, matching, difference-in-differences, regression discontinuity, and instrumental variables. Throughout the course, participants will apply these concepts and methods in hands-on sessions to real-world examples in the social sciences. The application will be conducted with the statistical software package Stata. The course will also touch upon advanced topics such as effect modification, reverse causality, measurement issues, and data quality. By the end of the course, participants will have the skills and knowledge to design, conduct, and interpret causal inference studies in the social sciences. They will be able to engage with the contemporary literature of causal inference and identify state-of-the-art methods which might be most relevant to their specific research question.

Keywords:

causal inference, observational studies, panel data, experiments, causal identification

Course Prerequisites:

- Knowledge of basic statistical concepts, including the principles of linear and binary logistic regression.
- Background in statistical software, preferably Stata.
- Basic understanding of designing quantitative studies.

Target Group:

Participants will find the course useful if they:

- have a background in the social, behavioral or economic sciences (economists, political scientists, sociologists, criminologists, psychologists, etc.)
- are interested in methods for causal inference based on experimental and/or observational data, especially panel data
- have a firm knowledge in linear regression modelling
- are motivated to apply the concepts and statistical approaches in hands-on sessions.

Course and Learning Objectives:

By the end of the course participants will:

- Have a good understanding of the potential outcome framework, causal diagrams, and the counterfactual way of thinking.
- Be capable of designing their own studies to derive causal estimates in observational settings.
- Acquire an in-depth understanding of and the skills to carry out five family of methods: fixed effects models, matching, difference-in-differences, instrumental variables, and regression discontinuity design.
- Become familiar with interdisciplinary applications of the methods covered by the course.
- Be able to engage the contemporary literature of causal inference and identify state-of-the-art methods which might be most relevant to their specific research question.

Organizational Structure of the Course:

The course will be split into a three-hour morning and a three-hour afternoon session, including coffee breaks. In order to secure a close link between the learning and the application of contents, we will switch between lecture format (~50%) and hands-on exercises, tutorials, or lab sessions (~50%) in a flexible way. In addition to shorter exercises, a selected number of more in-depth assignments will be provided which participants solve in groups of 2-3. These include the application of causal inference methods to estimate effects based on existing datasets using Stata. Lecturers will be available for individual consultations to support work on group assignments and to facilitate discussions within groups.

Software and Hardware Requirements:

Participants will need to bring a laptop computer with a recent version of Stata (13 or higher) installed to successfully participate in this course. Stata short term licenses will be provided by GESIS for the duration of the course if needed.

Day-to-day Schedule and Literature:

Day	Topic(s)
1	**Causality, Counterfactuals and Experiments** **10:00 – 13:00** ▪ General introduction to causality and causal inference (foundations in the philosophy of science, perspectives, conditioning, selection, endogeneity, …) ▪ Counterfactual model and the fundamental problem of causal inference ▪ Different types of estimands: ATE, ATT, ATC ▪ What is your estimand? **14:00 – 17:00** ▪ Directed Acyclic Graphs (DAGs) I: Foundations ▪ Different types of experiments and their link to the counterfactual model ▪ Factorial survey experiments
2	**Experiments and Cross-sectional Data** **9:00 – 12:00** ▪ Natural experiments in a survey context ▪ Stata Application: Factorial Survey Experiments & Natural Experiments **13:30 – 16:30** ▪ Regression, matching & inverse probability weighting ▪ Directed Acyclic Graphs (DAGs) II: Confounders, colliders & mediators ▪ Stata Application: regression, matching & inverse probability weighting
3	**Longitudinal Data** **9:00 – 12:00** ▪ Longitudinal (in particular panel) designs ▪ Handling panel data ▪ First-differences, random and fixed effects regression **13:30 – 16:30** ▪ Diff-in-diff-estimation ▪ Tailoring control groups (sample restrictions, matching, …) ▪ Stata Application: First-differences, random, fixed effects regression, Diff-in-diff-estimation

4	**Temporal Dynamics, Heterogeneity, and Data Quality** **9:00 – 12:00** ▪ Impact Dummies ▪ Heterogeneous treatment effects (standard tools, more advanced tools) ▪ Methods to address: reverse causality & simultaneity: Instruments & ML-SEM **13:30 – 16:30** ▪ Stata Application: Impact Dummies, Heterogeneous Effects & IV estimation ▪ Impact of data quality on causal inference I (e.g., missing values, panel mortality, reactivity)
5	**Advanced Topics** **9:00 – 12:00** ▪ Impact of data quality on causal inference II (e.g., missing values, panel mortality, reactivity) ▪ Measurement issues (graph theory; measurement invariance; changing measures) **13:00 – 16:00** ▪ Stata Application: Data Quality & Measurement Issues ▪ Further advanced topics ▪ Wrap-up & Outlook

Preparatory Reading:

For some basic background on causality:

- Elwert, F. (2013). Graphical Causal Models. In: Morgan, S.L. (ed.): Handbook of Causal Analysis for Social Research. Dordrecht: Springer, pp. 245-273.
- Gangl, M. (2010). Causal Inference in Sociological Research. Annual Review of Sociology 36: 21-47.

To review material on statistics, please consult this excellent overview:

- Gujarati, D.N. (2018). Linear Regression: A Mathematical Introduction. Thousand Oaks: Sage.

To familiarise yourself with the statistical software Stata, you can use the following textbook:

- Kohler, U./Kreuter, F. (2012). Data Analysis Using Stata. College Station: Stata Press.

Additional Recommended Literature:

- Allison, P. D. (2009). Fixed effects regression models. Thousand Oaks: Sage.
- Angrist, J.D./Pischke, J.-S. (2009). Mostly Harmless Econometrics. An Empiricist's Companion. Princeton: Princeton University Press.
- Angrist, J.D./Pischke, J.-S. (2015). Mastering Metrics. The Path from Cause to Effect. Princeton, NJ: Princeton University Press.
- Auspurg, K. and Hinz, T. (2015). Factorial Survey Experiments. Thousand Oaks: Sage.
- Best, H./Wolf, C. (eds.) (2014). The Sage Handbook of Regression Analysis and Causal Inference. London: Sage.
- Brüderl, J./Ludwig, V. (2014). Fixed-effects panel regression. In: Best, H./Wolf, C. (eds.): The Sage Handbook of Regression Analysis and Causal Inference. London: Sage, pp. 327–357.
- Dunning, T. (2012). Natural Experiments in the Social Sciences. A Design-Based Approach. Cambridge, MA: Cambridge University Press.
- Freedman, D.A. (1991). Statistical Models and Shoe Leather. In: Sociological Methodology, 2: 291-313.
- Gerber, A.S./Green, D.P. (2012). Field Experiments. Design, Analysis, and Interpretation. New York: W.W. Norton & Company.
- Guo, S.Y./Fraser, M.W. (2014). Propensity Score Analysis. Statistical Methods and Applications. Thousand

Oaks: Sage.
- Heckman, J.J. (2005). The Scientific Model of Causality. In: Sociological Methodology, 35: 1-97.
- Hernán, M.A., Robins, J.M. (2020). Causal Inference. Boca Raton: Chapman & Hall/CRC.
- Holland, P. (1986). Statistics and Causal Inference. In: Journal of the American Statistical Association, 81, S. 945-960.
- Imai, K. & Kim, I.S. (2019). When Should We Use Unit Fixed Effects Regression Models for Causal Inference with Longitudinal Data? In: American Journal of Political Science 63: 467-490.
- Imai, K., King, G. & Stuart, E.A. (2008). Misunderstandings Between Experimentalists and Observationalists About Causal Inference. In: Journal of the Royal Statistical Society: Series A, 171: 481-502.
- Imbens, G.W. (2015): Matching Methods in Practice: Three Examples. In: Journal of Human Resources 50, H. 2, 373-419.
- Imbens, G.W./Rubin, D.B. (2015). Causal Inference for Statistics, Social, and Biomedical Sciences. An Introduction. New York: Cambridge University Press.
- King, G./Keohane, R.O./Verba, S. (1994). Designing Social Inquiry. Scientific Inference in Qualitative Research. Princeton, NJ: Princeton University Press.
- Leitgöb, H./Seddig, D./Schmidt, P./Sosu, E./Davidov, E. (2021). Longitudinal Measurement (Non-)Invariance in Latent Constructs: Conceptual Insights, Model Specifications, and testing Strategies. In Cernat, A/Sakshaug, J. (eds.): Measurement Error in Longitudinal Data. Oxford: Oxford University Press, pp. 211-257.
- Leszczensky, L./Wolbring, T. (2022). How to Deal With Reverse Causality Using Panel Data? Recommendations for Researchers Based on a Simulation Study. In: Sociological Methods & Research, 51: 837–865.
- Lieberson, S. (1987). Making It Count. The Improvement of Social Research and Theory. Berkeley: University of California Press.
- Light, R.J./Singer, J.D./Willett, J.B. (1990). By Design. Planning Research on Higher Education. Cambridge: Harvard University Press.
- Little, R.J.A./Rubin, D.B. (2002). Statistical Analysis with Missing Data. New York: John Wiley (2nd Edition).
- Ludwig, V./Brüderl, J. (2021). What You Need to Know When Estimating Impact Functions with Panel Data for Demographic Research. In: Comparative Population Studies, 46.
- Morgan, S.L./ Winship, C. 2015. Counterfactuals and Causal Inference: Methods and Principles for Social Research. New York: Cambridge University Press (2nd Edition).
- Morgan, S.L. (ed.) (2013). Handbook of Causal Analysis for Social Research. New York: Springer.
- Murnane, R.J./Willett, J.B. (2010). Methods Matter: Improving Causal Inference in Educational and Social Science Research. Oxford: Oxford University Press.
- Mutz, D.C. (2011). Population-based Survey Experiments. Princeton, NJ/Oxford, UK: Princeton University Press.
- Pearl, J. (2009[2000]). Causality: Models, Reasoning, and Inference. Cambridge: Cambridge University Press (2nd Edition).
- Pearl, J. (2010). The Foundations of Causal Inference. In: Sociological Methodology, 40: 75-149.
- Pearl, J./Glymour, M./Jewell, N.P. (2016). Causal Inference in Statistics. A Primer. New York: Wiley.
- Rosenbaum, P.R. (2017). Observation and Experiment. An Introduction to Causal Inference. Cambridge, MA: Harvard University Press.
- Rubin, D.B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. In: Journal of Educational Psychology, 66: 688-701.
- Rubin, D.B. (2008). For Objective Causal Inference, Design Trumps Analysis. In: Annals of Applied Statistics, 2: 808-840.
- Shadish, W.R./Cook, T.D./Campbell, D.T. (2002): Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston: Houghton Mifflin Company.
- Textor, J./Hardt, J./Knüppel, S. (2011): Letter to the Editor: DAGitty: A Graphical Tool for Analyzing Causal Diagrams. In: Epidemiology, 22: 745.
- Vanderweele, T.J. (2015). Explanation in Causal Inference. Methods for Mediation and Interaction. Oxford: Oxford University Press.
- Woodward, J. (2003). Making Things Happen. A Theory of Causal Explanation. Oxford: Oxford University

Press.