ABSTRACTS

Conference on
Statistics and Causality

May 23rd and 24th 2014
Oskar Morgenstern Platz 1
1090 Wien

universität wien

University of Vienna
Department of Psychology
Unit of Quantitative Methods

Content

Determining Direction of Dependency in a Regression Line

Yadolah Dodge
University of Neuchâtel

Iraj Yadegari
Université de Sherbrooke

Direction of dependence in a linear regression when both variables are subject to error has been under investigation in several articles in recent years. In this article, we present different methods including a method based on higher order of their respective cumulants for determining the direction of dependency. The asymmetric dependency along with other measures such as the coefficient of variation will also be described. Empirical data from economics are used to illustrate the use of the methodology.

Bayesian Inference from Regression Discontinuity Designs

Eduardo Gutiérrez-Peña
National University Mexico

Quasi-experimental designs can be useful in establishing causal effects in settings where randomized designs are not feasible or desirable for some reason. In this talk we shall focus on the regression-discontinuity design. After a brief review, we shall discuss how data obtained from such designs can be analyzed from a Bayesian perspective.

Granger Causality for ill-posed Problems: Methods, Ideas, and Application in Life Sciences

Kateřina Hlaváčková-Schindler
Academy of Sciences of the Czech Republic

For detecting causal relationships between variables x_j, $j = 1, \ldots, p$ the concept of the so called multivariate Granger causality has been proposed. Based on the intuition that the cause should precede its effect, in Granger causality one says that a variable x_i can be potentially caused by the past versions of the involved variables x_j, $j = 1, \ldots, p$. Then, in the spirit of the statistical approach and using a linear model for the causal relationship, we consider the following approximation problem:

$$x_t^i \approx \sum_{j=1}^{p} \sum_{l=1}^{L} \beta_l^j x_{t-1}^j, t = L+1, \ldots, T. \quad (1)$$

where L is the so called maximal lag, which is the maximal number of the considered past versions of the variables. The coefficients β_l^j can be determined by the least squares method. As in the statistical approach, one can now fix the value of the threshold parameter β_{tr}> 0 and say that

x_j has a causal influence on x_i if $\sum_{l=1}^{L} \left|\beta_l^j\right| > \beta_{tr}$. (2)

It is well known that for a big number of variables p, the causality network obtained from the approximation problem (1) is not satisfactory. First of all, it cannot be guaranteed that the solution of the corresponding minimization problem is unique. Another issue is connected with the number of the causality relationships that are obtained from (1). This number is typically very big, while one expects to have a few causality relationships with a given gene. To address this issue, various variable selection procedures can be employed. In this paper we review variable selection procedures applied to multivariate Granger causality. On a practical example of a gene regulatory network we illustrate, how our method called Graphical Lasso Granger method with 2-levels-thresholding overcomes the commonly used methods.

Investigating Granger-Causality with Multidimensional Item Response Models

Ingrid Koller, Claus H. Carstensen
University of Bamberg

Wolfgang Wiedermann
University of Vienna

Alexander von Eye
Michigan State University

Investigating educational processes and trajectories are important goals in the current educational research and reflect the main aims of the national educational panel study (NEPS) in Germany. From a methodological perspective, two foci can roughly be distinguished: 1) how to measure the same competence across the life span and 2) which further competencies or context variables affect modifiability of the target competence. The latter can sufficiently be addressed by Granger causality analysis. Granger causality approaches examine whether past observations of a dimension *X* can be used to predict the last observation of dimension *Y* in a series of repeated measures. A common item response model for scaling educational tests is the unidimensional Rasch model and its extensions. However, more recently, multidimensional item response models became popular for the exploration of dimensionality and assessing the latent factor structure of multiple traits. In contrast to unidimensional models these models offer the advantage to simultaneously model effects of multiple traits on answer processes. In this study we combine the concept of Granger causality with multidimensional item response models for longitudinal studies. After a short introduction in multidimensional item response modeling in combination with Granger causality we introduce a second option for the investigation of Granger causality in the context of item response models, so-called person explanatory item response models. Both types of models will be applied, compared, and discussed by an empirical data example.

Integrating Propensity Scores with Latent Class Analysis to Identify Determinants of Class Membership

Stephanie T. Lanza & Bethany C. Bray
Pennsylvania State University

Latent class analysis (LCA) is a widely used approach for identifying population subtypes. In the behavioral sciences, LCA has been an effective technique for understanding patterns of risk exposure and for modeling complex behaviors such as substance use, sexual risk behavior, and delinquent behavior. In LCA, multiple categorical indicators are used to identify a set of mutually exclusive and exhaustive latent classes. Covariates are commonly introduced into the latent variable model to identify characteristics associated with latent class membership. Yet even when data are drawn from a longitudinal study, unless individuals are randomized to levels on the covariate, causal inferences about the effect of the covariate on class membership cannot be drawn. Modern causal inference methods can be integrated with latent class analysis (LCA) to enable important questions about the determinants of latent class membership to be addressed. Two propensity score techniques, matching and inverse propensity weighting, are demonstrated for conducting causal inference in LCA. Focus is given to the specific causal questions that can be addressed with these techniques. This approach is demonstrated using empirical data from a national sample of individuals in the United States, where the causal effect of a risk factor on subsequent behavior latent class membership is estimated. Several methodological issues, such as handling missing data on the confounders, exposure variable, and multivariate outcome, as well as the implementation of this approach in SAS and R, are discussed.

Alternative forms of Granger causality, heterogeneity and non-stationarity

Peter C. M. Molenaar
Pennsylvania State University

Alternative forms of Granger causality based on standard vector autoregressive (VAR), structural VAR and unified structural equation models are presented, including time-frequency domain extensions. The group iterative multiple model estimation (GIMME) approach is proposed as the best method to accommodate heterogeneity and avoid limitations of structural VAR modeling. State space models with time-varying parameters are illustrated with applications to non-stationary fMRI BOLD time series.

Modeling decisions and preference change for transformative experience

L. A. Paul
University North Carolina

I will discuss a problem for subjectively rational decision-making in cases where the experience is dramatically new (epistemically transformative) as well as life-changing (personally transformative). I'll start by outlining the structure of the problem and showing why it creates trouble, and then I'll argue that the problem arises for important life decisions such as choosing to get (or to avoid) a cochlear implant. I'll discuss applications involving imprecise credences, known unknowns and hierarchical Bayesian modeling.

On the importance of adjustment for latent covariates

Steffi Pohl
Freie Universität Berlin

Marie-Ann Sengewald
Friedrich-Schiller-Universität Jena

In non-randomized studies the average total effect (ATE) of a treatment is commonly estimated by adjusting for manifest covariates; either with ANCOVA or Propensity score (PS) methods. Treatment selection may, however, be based on latent constructs. Simulation studies (Steiner, Cook & Shadish, 2011, Cook, Steiner & Pohl, 2009) demonstrated that measurement error in covariates has a biasing effect on ATE estimates, when the respective covariates are relevant for adjustment. However, the simulated data does not necessarily depict real data situations. Does selection in empirical studies really depend on latent instead of manifest covariates? To what extent do other covariates cope for unreliability in one covariate? In our studies we delineated the question whether adjustment of latent instead of manifest covariates improves the accuracy of ATE estimation in empirical applications. We used data of a within-study comparison from Pohl, Steiner, Eisermann, Soellner, & Cook, (2009). In this study 202 students participated either in an English or a mathematics training. The special study design combines randomized and non-randomized assignment to the treatment groups and enables to investigate the accuracy of causal effect estimates. ANCOVA and PS methods with manifest or latent covariates were applied in the non-randomized condition for estimating the ATE of the English training. Incorporating latent covariates was realized via structural equation modeling in a generalized ANCOVA approach (Steyer & Partchev, 2008) and a recently developed approach for PS analysis (Raykov, 2012). ATE estimates of the adjusted non-randomized condition were compared to the results from the randomized arm. Including latent rather than manifest covariates increased the accuracy of ATE estimation. Thus treatment selection occurred on the latent dimension of the covariates and the importance of modeling latent covariates was demonstrated for empirical data. In further simulation studies and empirical

analyses we investigated the impact of other covariates on ATE estimation when adjusting for one fallible covariate. The results show that further covariates that are highly correlated with the fallible covariate, do limit the impact of measurement error on the ATE estimate. Limitations as well as implications for applications are discussed.

A non-Gaussian approach for estimating possible causal direction in the presence of latent confounders

Shohei Shimizu
Osaka University, Japan

In many empirical sciences, the causal mechanisms underlying various phenomena need to be studied. Structural equation modeling is a general framework used for multivariate analysis, and provides a powerful method for studying causal mechanisms. However, in many cases, classical structural equation modeling is not capable of estimating the causal directions of variables. This is because it explicitly or implicitly assumes Gaussianity of data and typically utilizes only the covariance structure of data. In many applications, however, non-Gaussian data are often obtained, which means that more information may be contained in the data distribution than the covariance matrix is capable of containing. Thus, many new methods including Dodge and Rousson (2001) have recently been proposed for utilizing the non-Gaussian structure of data and estimating the causal directions of variables. In this talk, we will first present an overview of non-Gaussian causal discovery methods and then go to some recent advances. In particular, we consider estimating the possible causal direction of two observed variables in the presence of latent confounding variables. Several existing methods have been shown to consistently estimate causal direction assuming linear or some type of nonlinear relationship and no latent confounders. However, the estimation results could be distorted if either assumption is actually violated. We first propose a linear non-Gaussian acyclic structural equation model with individual-specific effects that allows latent confounders to be considered. We then propose an empirical Bayesian approach for estimating possible causal direction using the new model. We demonstrate the effectiveness of our method using artificial and real-world data.

Propensity Score Designs for Causal Inference: Covariate Selection Issues

Peter Steiner
University of Wisconsin

The popularity of propensity score (PS) methods for estimating causal treatment effects from observational studies has been strongly increasing during the past decade. However, the success of PS designs in removing selection bias rests on strong assumptions for identifying and estimating causal effects—particularly the strong ignorability assumption that requires that all confounding covariates are reliably measured. After a brief introduction to the design and analysis of different PS methods (matching, stratification, weighting, and regression), the talk mainly focuses on two practical challenges when implementing a PS design: (i) The selection of baseline covariates for removing confounding bias, and (ii) the influence of covariate measurement error on bias reduction. All these issues will be illustrated and discussed using evidence from simulation studies, meta-analyses and within-study comparisons from the social and behavioral sciences.

Direction of effects in categorical variables: A structural perspective

Alexander von Eye
Michigan State University

Wolfgang Wiedermann
University of Vienna

Latent classes explain the covariation of variables in classes of cases. Given the latent classes, these covariations disappear. In the present article, we apply this principle to manifest variables and fuse it with direction dependence methodology. In addition, we extend this principle to parameters other than covariation, and propose that not only univariate probability distributions of variables but also variable interactions can be placed on both the predictor and the outcome sides. Log-linear models are developed with multiple variables on the predictor side, the outcome side, and both sides. Sample models are presented in which univariate probability distributions are predicted as well as covariations among multiple dependent variables. In data examples, hypotheses about the self-perpetuation of aggressive behavior are tested. In the discussion, we relate the proposed methodology to latent class analysis.

Direction of Effects on Mediation Analysis

Wolfgang Wiedermann
University of Vienna

Alexander von Eye
Michigan State University

Linear regression models are usually employed to test mediation hypotheses. These models consider moments no higher than second order and do not allow one to make decisions about competing mediation models, that is, models in which the reverse flow of causality is considered. The recently more intensively discussed methods of direction dependence do consider higher moments. A direction of dependence methodology is introduced which allows one to evaluate hypotheses of directions of effects, and extend its application to mediation analysis. Significance tests for statistical inference on direction of effects are proposed and discussed. Results of a Monte-Carlo simulation of the performance of the tests under various data scenarios are presented and an empirical example from research on intimate partner violence is given.

Alternative State Trait Models for Longitudinal Data: Implications for the Assessment of Granger Causality

Phillip K. Wood
University of Missouri

The time-sequential relationships between constructs across waves of assessment has gained increasing attention by researchers as researchers use such associations to evaluate hypotheses of Granger causality between constructs. Although many studies have used studies composed of two waves of assessment, such designs cannot answer the question of whether estimates of these effects can be explained by reference to stable trait differences and/or general comorbidity across constructs. When, however, more than two measurement occasions are assessed, autoregressive covariation can be decomposed into covariation due to stable inter-individual trait differences and unique covariation from one measurement occasion to another. In similar fashion, such designs also allow the researcher to identify cross-construct covariation between adjacent measurement occasions in the presence of general trait-level covariation. Although the term "state-trait" models have been used for such longitudinal models, researchers have frequently ignored the question of whether the construct of interest has been successfully modeled via a single factor. By contrast, at the trait level, it is also possible that a given construct may be better approximated by a random intercept or two-factor model. In this paper, such competing longitudinal state-trait models are considered in the context of alcohol-tobacco comorbidity in a longitudinal study of college students at the Midwestern University. For both tobacco and alcohol constructs a random intercept model fit the data better than a single factor model and the estimated autoregressive and cross-lagged effects in such models is markedly different than that found when traits of tobacco and alcohol use (and their general comorbidity at the trait level) are not included in the model.

Conference Information

Organizers

Wolfgang Wiedermann
University of Vienna
Department of Psychology
Unit of Quantitative Methods
wolfgang.wiedermann@univie.ac.at

Alexander von Eye
Michigan State University
Department of Psychology
voneye@msu.edu

Conference Venue

Oskar Morgenstern Platz 1
12th floor
A-1090 Vienna
AUSTRIA
https://rossau.univie.ac.at/

Accommodation

Hotel Rathauspark
Rathausstrasse 17
A-1010 Wien
AUSTIRA
Phone: +(43) 1 40 412
http://www.austria-trend.at/Hotel-Rathauspark/en/

Supported by the Austrian Research Foundation and the University of Vienna