UNIVERSITÄT BIELEFELD
Fakultät für Wirtschaftswissenschaften

Zentrum für Statistik

15th Young Researchers Workshop of the Centre for Statistics (ZeSt)

3 June 2025

Programme

10:00 - 10:05	Dietmar Bauer: Welcoming
10:05 - 10:35	Michael Balzer: Estimation of Simultaneous Equations Models with Interrelated Endogenous Networks, and an Application to Cancer Research in Science and Industry
10:35 - 11:05	Kevin Hoppe: Asymmetric Confidence Intervals for Average Treatment Effects on Binary Outcomes Using Difference- and Ratio-Effect Specifications
11:05 - 11:15	Short Break
11:15 - 11:45	Sophie Schmiegel: Development Process of a Clinical Decision Support System for Empiric Antibiotic Therapies in Sepsis Patients
11:45 - 12:15	Kurtulus Kidik: Model Selection and Forecasting in Matrix Autoregressive Models
12:15 - 13:30	Lunch Break
13:30 - 14:00	Jan-Ole Koslik: A Lego System for Latent Markov Models
14:30 - 15:00	Jonas Bauer: Modelling the unobservable: How different operationalizations can emerge from the same concept
15:00 - 15:15	Coffee Break
15:15 - 15:45	Katharina Amman: Financial Risk Modelling Using Markov Switching Generalized Additive Models for Location, Scale and Shape
15:45 - 16:15	Nele Bertling: Quantile-based Integrated Learning and Optimization for Contextual Stochastic Vehicle Routing
16:15 - 16:25	Discussion and Closing Remarks

Estimation of Simultaneous Equations Models with Interrelated Endogenous Networks, and an Application to Cancer Research in Science and Industry

Michael Balzer
Center for Mathematical Economics, Bielefeld University, Germany
michael.balzer@uni-bielefeld.de

Cancer research is crucial for advancing both scientific understanding and industrial innovation, leading to improvements in diagnostics, treatments, and prevention strategies. In the scientific domain, research explores the molecular and genetic basis of cancer, enabling breakthroughs in targeted therapies, immunotherapy, and personalized medicine. Cutting-edge technologies such as CRISPR gene editing, artificial intelligence, and next-generation sequencing have accelerated the discovery of biomarkers and therapeutic targets. In the industrial sector, collaboration among pharmaceutical companies, biotechnology firms, and research institutions has resulted in groundbreaking treatments, including monoclonal antibodies, CAR-T cell therapies, and novel chemotherapeutic agents. Despite these advances, challenges such as drug resistance, high treatment costs, and disparities in healthcare access persist. Overcoming these barriers requires sustained interdisciplinary collaboration between scientists, clinicians, and industry leaders. Therefore, understanding the relationship between academic and industrial cancer research is essential for fostering continued innovation. A key question is whether these domains act as strategic complements, reinforcing each other, or as substitutes, where progress in one reduces incentives in the other. In this work, we introduce a novel perspective, arguing that academic cancer research influences industrial cancer patent production not only as a direct input but also through its strategic effects, shaping innovation dynamics beyond knowledge transfer. We develop a mathematical model of strategic interactions that accounts for multiple activities across interrelated networks. We characterize the unique Nash equilibrium of this model and propose an empirical model related to simultaneous equations models for empirical model validation based on real-world data. To estimate the structural parameters, we propose a novel two-step estimation strategy which accounts for the potential endogenous nature of science and industry networks. We implement the estimator in the programming language R and investigate proper functionality by conducting extensive simulation studies. Lastly, we marry the model with data collected on cancer research in science and industry.

Asymmetric Confidence Intervals for Average Treatment Effects on Binary Outcomes Using Difference- and Ratio-Effect Specifications

Kevin Hoppe
Faculty of Psychology and Sports Science, Bielefeld University, Germany
kevin.hoppe@uni-bielefeld.de

Randomized controlled trials (RCTs) are the gold standard for estimating causal treatment effects. In the context of binary outcomes, difference- and ratio-based specifications of an average treatment effect (ATE) can be estimated from logistic regression models incorporating covariates. Using the Delta method, symmetric confidence intervals for a given specification can be constructed. However, since treatment effects build on individual outcome probabilities, the range of ATE estimates is naturally restricted. Symmetric intervals can therefore exceed the admissible range of values when effect sizes or estimation uncertainty are large. Moreover, especially with highly nonlinear ratio effects, symmetric confidence intervals may represent uncertainty on either side of the point estimate very differently than asymmetric intervals would. In this work, we examine two methods of obtaining asymmetric confidence intervals when covariates are stochastic. These methods not only adapt to valid effect boundaries but may also better accommodate effect nonlinearities. The first option is a transformation approach that maps effect estimates to a space where symmetric intervals are admissible and then transforms interval boundaries back. The second option is a bootstrap approach that estimates asymmetric quantiles of the ATE distribution by resampling. We present a simulation study that compares inference based on symmetric and asymmetric intervals under difference- and ratio-based ATE specifications and different levels of effect size and heterogeneity. Results indicate that, while symmetric intervals perform well overall, both bootstrap and transformation-based intervals adhere to the admissible range of estimates and can also help improve inference with large and highly heterogenous ratio-based effects.

Development Process of a Clinical Decision Support System for Empiric Antibiotic Therapies in Sepsis Patients

Sophie Schmiegel
Faculty of Business Administration and Economics, Bielefeld University, Germany
sophie.schmiegel@uni-bielefeld.de

Antibiotic therapies are the most important treatment against bacterial infections. However, their inadequate use poses a major threat to global health due to an increased incidence of antibiotic resistance. This particularly threatens sepsis patients, whose chances of survival greatly decrease if no effective empiric antibiotic therapy is administered. The selection of such empiric antibiotic therapy is challenging in clinical practice — a clinical decision support system could support physicians in deciding for a fast and targeted antibiotic therapy. To aim at a reliable system, which can be applied in clinical practice, preliminary work needs to be done. This includes gathering of medical expertise, data collection and preprocessing as well as the construction of first models. In this work, we focus exactly on this process and highlight the importance of a close exchange between medical experts and model constructors. We further identify challenges which impede the development of reliable models and present our approaches to model a physician's choice for an empiric antibiotic therapy. We consider these model results taking into account the identified challenges and draw conclusions for a successful model implementation

Model Selection and Forecasting in Matrix Autoregressive Models

Kurtulus Kidik
Faculty of Business Administration and Economics, Bielefeld University, Germany
kurtulus.kidik@uni-bielefeld.de

Matrix-valued time series are increasingly prevalent in fields such as economics, finance, neuroscience, and environmental science, where observations exhibit multidimensional structure, temporal dynamics and structured spatial relationships. Traditional vector-based models often fail to adequately capture these spatial and temporal interactions. Therefore, matrix autoregressive (MAR) models have emerged as a natural and effective framework for modeling such dual dependencies. This talk presents a comprehensive examination of model selection and forecasting within the MAR paradigm, emphasizing the importance of preserving matrix structure for accurate and interpretable analysis. We investigate various MAR models alongside vector autoregressive (VAR) models. Model selection is addressed through extensions of traditional criteria—such as the Akaike Information Criterion (AIC), Schwarz-Bayesian Information Criterion (BIC) — as well as cross-validation tailored to matrix settings. Forecasting performance is assessed using both simulated and real-world datasets. Extensive experiments demonstrate that approaches which preserve and exploit matrix structure consistently outperform vectorized or unstructured baselines in terms of forecasting accuracy and interpretability. Our results underscore the value of leveraging matrix-specific properties for enhanced predictive performance and provide practical guidelines for selecting appropriate MAR models in high-dimensional and complex time series scenarios.

A Lego System for Latent Markov Models

Jan-Ole Koslik
Faculty of Business Administration and Economics, Bielefeld University, Germany
jan-ole.koslik@uni-bielefeld.de

In this talk, I offer a unifying perspective on a wide range of models that share a common structure: unobserved (latent) states evolving over time according to a Markov process. Despite their diverse appearances and areas of application, models ranging from hidden Markov models, through state-space models, to Markov-modulated Poisson processes can all be understood within a single coherent framework, which we refer to as latent Markov models. This perspective not only clarifies conceptual connections across model classes but also motivates a unified approach to inference.

Building on this unification of inferential procedures, the second part of the talk introduces the LaMa R package. LaMa supports a broad range of discrete and continuous formulations for time and state spaces, and provides a modular, “Lego-style” system for constructing custom likelihood functions. Users can flexibly assemble models from pre-built components or define their own to suit specific needs. We will see LaMa in action through case studies that demonstrate its flexibility and practical utility.

Modelling the unobservable: How different operationalizations can emerge from the same concept

Jonas Bauer
Faculty of Business Administration and Economics, Bielefeld University, Germany
j.bauer@uni-bielefeld.de

SEM is a widely used statistical technique in fields like psychology and economics. One reason for its popularity is the possibility to include unobservable quantities of interest, i.e., theoretical concepts, in a statistical model. Effectively, every such theoretical concept must be operationalized as one of three possible construct types. In this talk, we focus on latent variables that reflect sets of observations and composites that are artificially forged out of these observations. A common recommendation states that an ideal construct should represent the theory about and nature of the theoretical concept. In practice, however, the same concept is often studied, argued for and published using different operationalizations. In this talk, we take a closer look at the recommendation for choosing a construct type and examine one possible explanation for publications with different construct types, namely a 2-component structural equation mixture model.

Financial Risk Modelling Using Markov Switching Generalized Additive Models for Location, Scale and Shape

Katharina Ammann
Faculty of Business Administration and Economics, Bielefeld University, Germany
katharina.ammann@uni-bielefeld.de

Financial Risk Modelling With Quantifying the risks associated with financial assets using metrics such as value-at-risk (VaR) and expected shortfall (ES) is essential for effective risk management. However, accurately modelling tail risk remains a challenge due to the complexity of financial return distributions and their time-varying characteristics. In this paper, we propose Markov-switching generalised additive models for location, scale, and shape (GAMLSS) as a flexible class of latent-state time series regression models to model these metrics. In contrast to conventional models like generalised autoregressive conditional heteroscedasticity (GARCH), which are commonly applied to VaR and ES forecasting, Markov-switching GAMLSS extends the modelling framework by allowing to model different state-dependent parameters of the response distribution — not only the mean and variance, but also the skewness and kurtosis — as potentially smooth functions of a given set of covariates. To evaluate model performance and ensure compliance with Basel regulatory standards, the feasibility of the proposed methodology is assessed through simulation experiments and a real-data application, where VaR and ES forecasts are compared with historical values using backtesting.

Quantile-based Integrated Learning and Optimization for Contextual Stochastic Vehicle Routing

Nele Bertling
Faculty of Business Administration and Economics, Bielefeld University, Germany
nele.bertling@uni-bielefeld.de

Traditional sequential learning and optimization (SLO) methods rely on expected value predictions, often leading to suboptimal decisions in contextual stochastic optimization problems such as the vehicle routing problems with uncertain travel times. Integrated learning and optimization (ILO) methods address this by directly minimizing the decision loss. In this work, we compare classical SLO and ILO methods with our quantile-based ILO approach, which selects the best-performing quantile during training to minimize the decision loss. Our evaluation considers normalized decision loss, the total training time, and the time spent solving the optimization problems. While SPO+ improves decision quality, its training time is significantly impacted by the cost of solving the optimization problems. In contrast, SLO as well as quantile-based ILO approaches, though sometimes less accurate, shift computational effort toward model training rather than execution and may offer greater scalability for larger optimization problems. Finally, we show that our quantile-based ILO approach using TabPFN as a prediction model consistently outperforms SPO+ while requiring a similar amount of time, albeit with the need for a GPU. These findings offer new insights into the trade-offs between accuracy, computational efficiency, and scalability in contextual stochastic optimization.