UNIVERSITÄT BIELEFELD
Fakultät für Wirtschaftswissenschaften

Zentrum für Statistik

17th Young Researchers Workshop of Centre for Statistics (ZeSt)

27th July 2026

Programme

10:00 - 10:05	Welcoming
10:05 - 10:35	Kurtulus Kidik – Misspecified but parsimonious: the value of Kronecker structure for high-dimensional vector autoregressions
10:35 - 11:05	Thanh Thao Bui – A method for identifying dynamic error correlation in ordered probit models
11:05 - 11:35	Maya Vienken – Jump-augmented stochastic volatility: a discrete-time state-space approach
11:35 - 11:50	Coffee break
11:50 - 12:20	Sven Hellbusch – Forecasting electricity prices via a smooth transition regression model
12:20 - 12:50	Kevin Hoppe – Bayesian inference for average effects on binary outcomes in small samples
12:50 - 13:00	Discussion and closing remarks

Misspecified but parsimonious: the value of Kronecker structure for high-dimensional vector autoregressions

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

Many modern time series are naturally observed as matrices rather than vectors — for instance, several macroeconomic indicators measured across several countries each quarter. Stacking such data into a single long vector and fitting an unrestricted vector autoregression (VAR) is often infeasible: the number of parameters grows quadratically with the dimension and quickly exceeds the sample size. Matrix autoregressions address this problem by expressing the VAR coefficient matrix as a sum of a few Kronecker products of two small matrices — one acting on the rows and the other on the columns — reducing thousands of parameters to a few dozen.

In practice, however, real data carry no guarantee that this structural restriction holds. This talk examines what the structured estimator delivers under model misspecification and when its bias is a worthwhile trade-off. We address three main questions. First, we show that the estimator consistently recovers the best separable approximation of the true dynamics, and its approximation bias has an exact, computable characterization as the neglected tail of a particular singular-value spectrum. Second, a bias–variance calculation yields a sharp criterion for when the structured estimator outperforms the unrestricted VAR in forecasting; this advantage increases as the cross-sectional dimension grows and becoming decisive when the OLS-VAR is no longer estimable. Third, we demonstrate that choosing how many Kronecker terms to retain reduces to thresholding estimated singular values against a noise floor, in close analogy with selecting the number of factors in a factor model.

Monte Carlo evidence confirms each theoretical prediction. An empirical application to macroeconomic data further demonstrates that the proposed approach strictly dominates the feasible unrestricted OLS-VAR and, as a class, matches or outperforms ridge-VAR and factor-VAR benchmarks while using five to ten times fewer parameters, placing matrix autoregressions squarely on the accuracy–parsimony frontier.

A method for identifying dynamic error correlation in ordered probit models

Thanh Thao Bui
Faculty of Business Administration and Economics, Bielefeld University, Germany
thanh.bui@uni-bielefeld.de

Ordered probit models are widely used to analyze decision-makers' choices among ordered alternatives. In survey datasets where respondents answer a series of questions over time, the temporal correlation of the error term may represent as the dynamics and should be explicitly accounted for in the model. Ignoring this structure can lead to incorrect assumptions about the distribution of the error terms, resulting in biased parameter estimates and misleading statistical inference regarding the significance of the estimated parameters.

This research proposes a technique for identifying the dynamic structure of the error terms in the form of an autoregressive process of order (p) (AR((p))). The proposed approach is based on a flexible framework that considers pairs of observations from different time points, thereby avoiding repeated estimation of the ordered probit model and the specification of the high-dimensional variance–covariance matrix that arises from a large number of choice occasions. Generalized least squares (GLS) is used to combine information from the variance–covariance matrices, while the Yule–Walker estimators are employed to estimate the parameters of the AR((p)) process. The autocorrelation function (ACF) and partial autocorrelation function (PACF) correlograms, together with information criteria, are used to identify the appropriate dynamic error structure.

The results indicate that the proposed technique successfully identifies the dynamic structure of the error terms. Its performance is demonstrated through an extensive simulation study, which also establishes the consistency of the proposed estimator. The findings further suggest that the reliability of the technique improves with increasing sample size. The proposed method provides a useful tool for exploratory data analysis (EDA), enabling researchers to identify and characterize the temporal correlation structure of the error terms before estimating the full dynamic ordered probit model.

Jump-augmented stochastic volatility: a discrete-time state-space approach

Maya Vienken
Faculty of Business Administration and Economics, Bielefeld University, Germany
maya.vienken@uni-bielefeld.de

Major events such as the COVID-19 pandemic, the Russia–Ukraine war, and trade conflicts can cause sudden spikes in financial market uncertainty that standard stochastic volatility models fail to capture. We propose a discrete-time state-space model in which the latent log-volatility process is augmented by Bernoulli-triggered, exponentially distributed jumps. This choice keeps the state transition density available in closed form, as a mixture of a Gaussian and an exponentially modified Gaussian, from which posterior jump probabilities follow directly. We estimate the model by numerical maximum likelihood, finely discretising the state space and evaluating the marginal likelihood via the forward algorithm, which makes fitting several model variants across a large number of assets computationally feasible. Applied to crude oil returns, jump-augmented specifications outperform their non-jump counterparts across standard information criteria, with identified jumps aligning closely with known market-disrupting events. Across 100 assets spanning equity indices, commodities, FX, crypto and single stocks, AIC selects a specification with jumps in the volatility process in 80% of cases, including specifications with jumps in both returns and volatility. These results are preliminary in that they rest on one particular jump distribution — which we plan to extend to arbitrary jump distributions by performing the required convolution via numerical integration when closed-forms are not available. With this machinery in place, we will broaden the empirical study towards identifying the most appropriate modelling framework (regarding jump and error distribution) for different asset classes.

Forecasting electricity prices via a smooth transition regression model

Sven Hellbusch
Institute of Mathematical Economics, Bielefeld University, Germany
sven.hellbusch@uni-bielefeld.de

We analyze regime-dependent price formation in the German day-ahead electricity market and propose a forecasting model that reconciles predictive performance with structural interpretability. Increasing renewable penetration induces non-linearities driven by interactions between residual load and fuel costs that are not adequately captured by standard linear models, while more flexible machine learning approaches often lack economic transparency.

To forecast hourly day-ahead prices, while addressing the accuracy-interpretability tradeoff, we develop a Logistic Smooth Transition Regression (LSTR) model with a market-fundamental specification, where coefficients vary smoothly as functions of observable state variables. Using the sequential linearity tests of Teräsvirta (1994), we identify residual load as the primary transition variable and select a logistic transition function. A second transition based on natural gas prices captures additional non-linear structure. To ensure economic consistency of the regime structure, we specify the model such that the gas price transition is multiplicatively linked to the residual load transition. This interaction effectively removes the economically implausible regime of low residual load combined with high gas prices and yields a three-regime representation of the price process.

The resulting model distinguishes between a base regime, a quantity stress regime associated with tight residual load, and a joint stress regime in which both residual load and gas prices are elevated. These regimes arise endogenously and reflect distinct structural relationships between prices and market fundamentals.

We estimate the model on German day-ahead data and evaluate its out-of-sample forecasting performance in a rolling window framework against established regression benchmarks. Formal linearity tests provide strong evidence of non-linear dynamics consistent with the proposed transition structure. The LSTR model matches the predictive accuracy of leading linear benchmarks while offering a substantially richer characterization of the underlying price formation process.

A key advantage of the approach lies in its interpretability: regime-dependent coefficient estimates quantify how the marginal effects of load, renewable generation, and fuel prices vary across market states.

Overall, the results demonstrate that incorporating economically motivated transition structures within a smooth transition framework provides a tractable way to model non-linearities in electricity prices without sacrificing accuracy while increasing interpretability, making the approach particularly suitable for applications that require both accurate forecasts and structural insights.

Bayesian inference for average effects on binary outcomes in small samples

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

In an RCT, average effects on binary outcomes can be defined as contrasts of marginal outcome probabilities in terms of differences, simple ratios, or odds ratios. They are commonly estimated from parametric response models like logistic regression that include covariate effects to model effect heterogeneity. G-computation can be used to marginalize over estimated unit-level outcome probabilities under observed and counterfactual treatment assignments. Under a maximum likelihood framework, interval estimates of average effects can be obtained using a delta-method-based transformation approach or bootstrapping. However, these rest on large-sample asymptotic theory. Therefore, this work explores the use of Bayesian logistic regression models in small samples. A simulation study is presented that compares maximum-likelihood-based confidence intervals with intervals based on either the full posterior distribution of average effects or intervals obtained from a prior-modified variant of IRLS model estimation. Bayesian bootstrapping is explored as a method of accounting for the unknown population distribution of covariates when drawing inference beyond the sample. Preliminary results indicate that coverages of equal-tailed and HPD credible intervals similarly profit from Bayesian bootstrapping but fall slightly short of the nominal level when true effects are large. Transformation-based intervals perform better than expected but can suffer from numeric instability when using maximum likelihood estimation.