MSc Seminar

Big Data in Economics and Finance

Prof. Dr. Winfried Pohlmeier

WS 2019/20

Goal and Content:

Modern empirical research in economics and finance is confronted with a vast menu of modeling strategies resulting from different data sets, huge numbers of potential covariates and an increasing number of alternative estimation strategies. The overflow of covariates, potential specifications and estimators makes it difficult to select the best modeling strategy. Standard testing procedures fail or are simply infeasible due to the high dimensionality of the problem. The problem is prevalent e.g. in macroeconomic forecasting, in financial econometrics (e.g. portfolio and risk management), and in the evaluation of causal treatment effects of public policies. Alternative strategies based on algorithmic learning mechanisms (regularization, machine learning, data mining) are not yet well understood in what regards their statistical properties.

The goal of this seminar is to acquaint master students with the necessary toolbox of econometric methods related to high dimensional estimation problems. Students have to write an empirical research paper in which they apply a novel and/or advanced econometric method to shed more light on a real world problem.
Time and Date:
The seminar will take place as a block seminar on the following two days:

**Monday April 06th, 2020, 09-17:00, room tba**
**Tuesday April 07th, 2020, 09-17:00, room tba**

The submission deadline of seminar paper is **March 30th, 2020** to Mrs. Kretz (F329) by email and also in paper format (verena.kretz@uni-konstanz.de)

Prerequisite:
Econometrics I and Advanced Econometrics (or equivalent). Knowledge of Financial Econometrics and/or Time series Analysis is desirable. We expect that students have a decent programming background in either MatLab, R or Python or at least willing to invest sufficient effort to learn one of these languages.

Organizational Issues:
- For enrollment contact: econ.masterseminar@uni-konstanz.de
- For more information contact Winfried Pohlmeier, F319, Tel. 2660, Winfried.Pohlmeier@uni-konstanz.de
- There will be a brief information session on the content of the seminar topics and the seminar requirements on **Oct. 21st 2019, room and time tba.**
- Seminar participants have to decide on the topics until **Oct. 28th, 2019** via Email to verena.kretz@uni-konstanz.de
- ECTS: 6 credits
  - There will be topics for which joint work of two participants may be possible.
Seminar Topics:

The topics above are only first suggestions and are supposed to give a flavor of the seminar’s content. We will assign the final topics according to the student’s background and interests.

**Topic 1: Bagging the Value-at Risk and Expected Shortfall**

The Value-at-risk and Expected are the most important concepts to evaluate extreme downside risks. However, there are many different ways of computing these risk measures. Based on bagging, one can construct robust model average of different VaR and ES approaches. Based on the fundamental bagging idea proposed by Kazak and Pohlmeier (2019a,b) the empirical study use the stock market data to improve on the results of Halbleib and Pohlmeier (2012).

**Topic 2: Machine Learning and Causal Returns to Schooling**

Machine learning techniques such as regression trees and support vector machines are well suited for estimating to obtain precise propensity scores based on data sets with a large number of covariates. These can used to estimate the causal returns to schooling. The goal of the empirical application based on the British Cohort Study is to investigate, if machine learning techniques improve the estimates of traditional causal effects estimators.

**Topic 3: Machine Learning and Undetected Heterogeneity**

Machine learning techniques are well suited to detect heterogeneity to automatically in populations. This allows the econometrician to estimate heterogeneous treatment effect and to prose public programs particularly designed for these heterogeneous groups. In an empirical application, the student is supposed to estimate heterogeneous returns to schooling based on the British Cohort Study is to investigate.

**Topic 4: How does the stress of the classmates effect one’s own stress? An econometric network analysis of emotional crossover during class.**

Social interactions shape individuals’ behavior. Recently, a growing literature has attempted to investigate social interactions theoretically and econometrically using the social network approach. At the econometric level, a special interest lies in the estimation of peer effects, which capture the dependence of an individual's outcome on group behavior. The goal in this research is to identify the different sources that drive the observed correlation between the outcomes of interacting individuals. In this project, the dataset has a panel structure, where information on the stress experience of the students is observed across time. The goal is to estimate peer effects, while finding the most appropriate model specification.

**Topic 5: Predicting financial volatilities by MIDAS using high frequency information.**

In empirical work the variable of interest often is observed at a different (lower) frequency than (some of) the explanatory variables. A common way to solve this problem is to aggregate the variable observed at a higher frequency to match the frequency of the variable of interest. Mixed Data Sampling (MIDAS) regression is a parametric way to directly use lags of an explanatory variable measured at higher frequency to explain the relation to a variable of interest measured at lower frequency. The objective of this project is to implement the MIDAS regression approach to use high frequency intraday information on financial assets in order to forecast volatility.
**Topic 6: Portfolio Choice by clustering financial time series.**
This paper studies various procedures for financial time series clustering, aimed at creating groups of time series characterized by similar behavior (for example with regard to extreme events). The ultimate goal of many clustering exercises in the context of financial data is to exploit the clustering solutions in portfolio optimization. The goal of the paper is to use time series clustering techniques for constructing optimal portfolios.

**Topic 7: Clustering of Economic Time Series via Cross-Predictability – An Application to Realized Volatilities**
The realized volatility time series of different stocks often show similar patterns, which indicates that the Forecasting Performance could be improved by clustering similar time series. A recent paper suggests to cluster time series via their ability to predict the respectively other time series, hence cross-predictability. This Approach shall be applied to realized volatilities.

**Topic 8: Transfer Learning in the Neural Network Setup – An Application to Realized Volatilities**
Recent results show that the dynamics in the realized volatility time series of different stocks are often similar. This result may be used to improve the forecasting performance for individual stocks by using the Information and/or learned dynamics from other stocks.

**Topic 9: Realized Covariance Forecasting via Neural Networks**
The forecasting of whole covariance matrices is difficult for more than a few time series, as the dimension of problem increases quadratically in the number of time series. Various neural network architectures have proven to perform well even if the dimensionality of the problem is large. In particular, the performance of multilayer perceptrons, convolutional neural networks and linear models shall be compared.

**Topic 10: Simulating Swarm Behavior**
Individuals in swarms strongly react to the individuals around them. Simple behavioral rules can lead to complex patterns in the overall swarm dynamics. In this simulation study, such behavioral rules shall be developed and adjusted, to mimic the behavior of animals in swarms.

**Topic 11: Earthquake Anticipation – A Simulation Study**
There are many historical reports on earthquakes being anticipated by unusual behavior of animals. So far, there is no consensus on which mechanism this anticipatory behavior follows. However, the mechanism can be written down as a flexible theoretical framework. Making reasonable assumptions, such mechanisms shall be simulated to obtain a better understanding on how the various unknown elements of the model impact the observed animal activity.