



# Digital Economy & Decision Analytics

Q-Kolleg mini-conference in 2016

Address: Icube Seminar Room, Level 1, Risk Management Institute, 21 Heng Mui Keng Terrace I<sup>3</sup> Building #04-03, Singapore 119613

Monday, 22 August 2016 (Morning)	
<b>9:00-9:10</b>	<b>Opening Address</b> Wolfgang Karl Härdle (Ladislaus von Bortkiewicz Chair of Statistics, Humboldt University of Berlin) Steven Kou (Risk Management Institute, National University of Singapore )
<b>9:10-9:30</b>	<b>Q3-D3-LSA</b> Wolfgang Karl Härdle (Ladislaus von Bortkiewicz Chair of Statistics, Humboldt University of Berlin)
<b>9:30-9:50</b>	<b>Forecasting Electricity Prices with Warping Functional AutoRegressive Dynamics</b> Jiejie Zhang (Department of Statistics and Applied Probability, National University of Singapore)
<b>9:50-10:10</b>	<b>CRIX or Evaluating Blockchain Based Currencies</b> Simon Trimborn (Ladislaus von Bortkiewicz Chair of Statistics, Humboldt University of Berlin)
<b>10:10-10:30</b>	<b>Forecasting Limit Order Book Liquidity Supply-Demand Curves with Functional AutoRegressive Dynamics</b> Wee Song Chua (Department of Statistics and Applied Probability, National University of Singapore)
<b>10:30-10:50</b>	<b>An Efficient Particle-Based Bayesian Method and Its Application to Estimating Derivative Pricing Models</b> Junye Li (ESSEC Business School, Singapore)
<b>10:50-11:10</b>	<b>Two New Tests for Equality of Several Covariance Functions</b> Jia Guo (Department of Statistics and Applied Probability, National University of Singapore)
<b>11:10-11:30</b>	<b>SCRIX</b> Lam Tuyen Nguyen (Ladislaus von Bortkiewicz Chair of Statistics, Humboldt University of Berlin)
<b>11:30-11:50</b>	<b>Discrimination and Classification of Face Data</b> Janice Paschek (Ladislaus von Bortkiewicz Chair of Statistics, Humboldt University of Berlin)
<b>11:50-12:10</b>	<b>Predicting Market Instability: New Dynamics Between Volume and Volatility</b> Zhi Qiao (Department of Physics, National University of Singapore)
<b>12:10-12:30</b>	<b>Quantile Treatment Effects in Difference in Differences Models under Dependence Restrictions and with only Two Time Periods</b> Tatsushi Oka (Department of Economics, National University of Singapore)
<b>Closing</b>	<b>Lunch</b>

Supported by:  **RMI**  
Advancing Risk Management  
for Singapore and Beyond

# Abstract

## **Q3-D3-LSA**

### **Wolfgang Karl Härdle**

QuantNet is an integrated web-based environment consisting of different types of statistics-related documents and program codes. Its goal is creating reproducibility and offering easy access in the scientific community by means of a powerful and specialized searching interface. To increase the information retrieval efficiency there is a need for incorporating semantic information. Our research is concentrated on semi-structured small or medium size corpora of documents. Our application platform is QuantNet. Three models will be examined: VSM, GVSM and LSA (Latent Semantic Analysis). Especially the LSA model has been successfully used for Information Retrieval (IR) purposes as a technique for capturing semantic relations between terms and inserting them into the similarity measure between two documents.

Our results show that different model configurations allow adapted similarity-based document clustering and knowledge discovery. In particular, different LSA configurations together with the hierarchical clustering reveal optimal results under our so-called  $M^3$  evaluation. QuantNet and the corresponding D3 based visualization can be found and applied under <http://quantlet.de/> and <http://quantlet.de/d3/ia/>, respectively.

## **Modeling Seasonality and Serial Dependence of Electricity Price Curves with Warping Functional Autoregressive Dynamics**

### **Jiejie Zhang (Jointly with Ying Chen, J.S. Marron)**

Electricity prices are high dimensional, serially dependent and have seasonal variations. We propose a Warping Functional AutoRegressive (WFAR) modeling that simultaneously accounts for the cross time-dependence and seasonal variations of the large dimensional data. In particular, electricity price curves are obtained by smoothing over the 24 discrete hourly prices on each day. In the functional domain, seasonal phase variations are separated from level amplitude changes in a warping process with the Fisher-Rao distance metric and the aligned (season-adjusted) electricity price curves are modeled in the functional autoregression framework. In real application in the California energy market from 5 July 1999 to 31 January 2001, the WFAR model provides superior out-of-sample forecasts with forecast error reduced 35% to 104% compared to several alternative models.

## **CRIX or Evaluating Blockchain based currencies**

### **Simon Trimborn**

The S&P500 or DAX30 are important benchmarks for the financial industry. These and other indices describe different compositions of certain segments of the financial markets. It is surprising, though, to see that emerging e-coins have not been mapped into an index yet because with cryptos like Bitcoin, a new kind of asset of great public interest has arisen. Usually, the index provider decides on a fixed number of index constituents which will represent the market segment. It is a huge challenge to set this fixed number and develop the rules to find the constituents, especially since markets change and this has to be taken into account. A method relying on the AIC is proposed to quickly react to market changes and therefore enable us to create an index, referred to as CRIX, for the cryptocurrency market. The codes used to obtain the results in this paper are available via [www.quantlet.de](http://www.quantlet.de).

## **Forecasting Limit Order Book Liquidity Supply-Demand Curves with Functional AutoRegressive Dynamics**

### **Wee Song Chua (Jointly with Ying Chen, Wolfgang Karl Härdle)**

Limit order book contains comprehensive information of liquidity on bid and ask sides. We propose a Vector Functional AutoRegressive (VFAR) model to describe the dynamics of liquidity supply and demand curves and use it to forecast the joint evolution of the multiple functional time series. We derive a closed-form maximum likelihood estimator under sieves and develop its asymptotic consistency. In real data analysis, it shows the VAR model produces a superior accuracy in liquidity prediction based on the limit order book for 12 stocks in NASDAQ traded from 2 Jan 2015 to 6 Mar 2015.

## **An Efficient Particle-Based Bayesian Method and Its Application to Estimating Derivative Pricing Models**

**Junye Li (Jointly with Andras Fulop)**

We propose a particle-based Bayesian method based on the quasi Monte Carlo methods and a sequential Monte Carlo sampler to estimate a class of derivative pricing models that take into account co-jumps of underlying prices and volatility, non-affineness, and self-excitation. This method can effectively control the noise of the likelihood estimate and makes it efficient to estimate such complex models. Empirical study using both S&P 500 index and variance swap rates shows that both non-affineness and self-excitation need to be introduced in modeling volatility dynamics.

## **Two New Tests for Equality of Several Covariance Functions**

**Jia Guo (Jointly with Jin-Ting Zhang)**

We propose two new tests for testing the equality of the covariance functions of several functional populations, namely a quasi GPF test and a quasi Fmax test. The asymptotic random expressions of the two tests under the null hypothesis are derived. We show that the asymptotic null distribution of the quasi GPF test is a chi-squared-type mixture whose distribution can be well approximated by a simple scaled chi-squared distribution. We also adopt a random permutation method for approximating the null distributions of the quasi GPF and Fmax tests. The random permutation method is applicable for both large and finite sample sizes. The asymptotical distributions of the two tests under a local alternative are investigated and they are shown to be root-n consistent. Simulation studies are presented to demonstrate the finite-sample performance of the new tests against three existing tests. They show that our new tests are more powerful than the three existing tests when the covariance functions at different time points have different scales. An illustrative example is also presented.

## **SCRIX**

**Lam Tuyen Nguyen (Jointly with Wee Song Chua)**

Since the launch of Bitcoin and its numerous followers cryptocurrencies become increasingly popular and accepted by users and investors. CRIX, a CRyptocurrency IndeX that serves as the benchmark for the cryptocurrency market is weighted by market capitalization. As a modification, we create a family of Smart CRyptocurrency Indices (SCRIX) by using alternative weighting methods, i.e. equal-weighted, minimum-volatility weighted, risk-weighted, Value atRisk, Expected Shortfall and Sharpe Ratio to give a risk-based view on the cryptocurrency market and to compare the performance of SCRIX against CRIX. Our findings demonstrate that weighting by Sharpe Ratio outshines among the various weighting methods, having an index higher than CRIX by 45 points on average, and is able to mimic CRIX movements up to 75% in terms of Mean Directional Accuracy.

## **Discrimination and Classification of Face Data**

**Janice Paschek (Jointly with Xiaofei Xu)**

Face recognition became one of the most active research fields, especially as the amount of available pictures and videos increased a lot in the past years. Therefore many algorithms for face recognition have been developed, which

differ in their performance for different database characteristics. This project aims to find the most efficient algorithms regarding time and accuracy for varying face recognition situations. On the basis of three databases, which include grayscale and colourful pictures and images of siblings, three different classification algorithms are tested, compared and evaluated. For that purposes k-Nearest-Neighbour (kNN), Support Vector Machines (SVM) and Fisher Linear Discriminant Analysis (FLDA) are considered as classification algorithms.

## **Predicting Market Instability: New Dynamics Between Volume and Volatility**

**Zhi Qiao**

Understanding, quantifying and predicting market fluctuation has become increasingly important in recent decades. It is widely accepted that there is a correlation between volume and volatility in a time series although debates still exist. Using empirical data and their distributions, we further investigate this correlation and discover new ways that volatility and volume interact, particularly when the levels of both are high. We find that the distribution of the volume-conditional volatility is well fit by a power law function with an exponential cutoff. We find that the volume-conditional volatility distribution scales with volume, and collapses these distributions to a single curve. We exploit the characteristics of the volume-volatility scatterplot to find a strong correlation between logarithmic volume and a quantity we define as local maximum volatility (LMV), which indicates the largest volatility observed in a given range of trading volumes. This finding supports our empirical analysis showing that volume is an excellent predictor of the maximum value of volatility for both same-day and near future time periods. We also use a joint conditional probability that includes both volatility and volume to demonstrate that invoking both allows us to better predict the largest next-day volatility than invoking either one alone.

## **Quantile Treatment Effects in Difference in Differences Models under Dependence Restrictions and with only Two Time Periods**

**Tatsushi Oka (Jointly with Tong Li and Brant Callaway)**

This paper shows that the Quantile Treatment Effect on the Treated (QTT) can be identified using a combination of (i) a Distributional Difference in Differences Assumption and (ii) an assumption on the dependence between the change in untreated potential outcomes and the initial level of untreated potential outcomes for the treated group. The second assumption recovers the unknown dependence from the observed dependence for the untreated group. This result extends previous research that required at least three periods of data for identifying the QTT under a similar setup. We also provide identification results when the assumptions hold only after conditioning on observed covariates. Under an additional assumption, we also show that the QTT is identified when only repeated cross sections are available. Finally, we consider estimation and inference -- we develop uniform confidence intervals and show the validity of an exchangeable bootstrap procedure.