PROGRAMME AND ABSTRACTS

1st International Conference on
Econometrics and Statistics (EcoSta 2017)

http://cmstatistics.org/EcoSta2017

Hong Kong University of Science and Technology
15 – 17 June 2017

HKUST
BUSINESS SCHOOL
香港科大商學院
WORLD CLASS IN ASIA

CMStatistics
Computational and Methodological Statistics
http://www.CMStatistics.org
http://www.CFEnetwork.org
Computational and Financial Econometrics
CFEnetwork

ISBN: 978-9963-2227-2-8

Technical Editors: Gil Gonzalez-Rodriguez and Marc Hofmann.

Co-chairs:

Ana Colubi, Erricos J. Kontoghiorghes, Tsung-I Lin, Yasuhiro Omori, Byeong Park and Mike K.P. So.

Scientific Programme Committee:

Alessandra Amendola, Eric Beutner, Monica Billio, Cathy W.S. Chen, Ray-bin Chen, Ming-Yen Cheng, Jeng-Min Chiou, Sung Nok Chiu, Taeryon Choi, Bertrand Clarke, Aurore Delaigle, Jean-Marie Dufour, Yang Feng, Alain Hecq, Inchi Hu, Salvatore Ingrassia, Yongdai Kim, Robert Kohn, Carlos Lamarche, Degui Li, WK Li, Yi Li, Zudi Lu, Geoff McLachlan, Samuel Mueller, Marc Paolella, Tommaso Proietti, Artem Prokhorov, Igor Pruenster, Jeroen Rombouts, Huiyan Sang, Tak Kuen Siu, Xinyuan Song, Mark Steel, Jianguo Sun, Nobuhiko Terui, Alan Wan, Lan Wang, Toshiaki Watanabe, Yoon-Jae Whang, Heung Wong, Valentin Zelenyuk, Helen Zhang, Ping-Shou Zhong and Hongtu Zhu.

Local Organizing Committee:

Yingying Li, Xinghua Zheng, Lilun Du, Catherine Chunling Liu, Alan Wan, Geoffrey Tso and Min Xie.

Dear Colleagues,

It is a great pleasure to welcome you to the 1st International Conference on Econometrics and Statistics (EcoSta 2017). The Conference is co-organized by the Working Group on Computational and Methodological Statistics (CMStatistics), the Network of Computational and Financial Econometrics (CFENetwork), and the Hong Kong University of Science and Technology (HKUST) Business School.

The aim is for the conference to become a leading meeting in econometrics, statistics and their applications.

The EcoSta 2017 consists of almost 150 sessions, four keynote talks, four invited sessions, and about 600 presentations. There are over 650 participants. This is quite impressive for a first edition of a conference. It is indeed promising that the EcoSta conference will become a successful medium for the dissemination of high quality research in Econometrics and Statistics, and facilitate networking.

The Co-chairs observed that the collective effort of the scientific program committee, session organizers, and local organizing committee has produced a programme that spans all the areas of econometrics and statistics. The HKUST provides excellent facilities and a fantastic environment with an astonishing scenery on the outskirts of Hong Kong. The local host, volunteers, and sponsoring universities have substantially contributed through their effort to the successful organization of the conference. We thank them all for their support. Particularly we express our sincere appreciation to the host and main sponsor, the HKUST Business School.

It is hoped that the quality of both the scientific programme and the HKUST will provide the participants with a productive, stimulating conference, and an enjoyable stay in Honk Kong.

The Elsevier journals of Econometrics and Statistics (EcoSta) and Computational Statistics & Data Analysis (CSDA) are associated with CFEnetwork, CMStatistics, and the EcoSta 2017 conference. The participants are encouraged to submit their papers to special or regular peer-reviewed issues of EcoSta and CSDA, and to join the networks.

Finally, we are happy to announce that the 2nd International Conference on Econometrics and Statistics (EcoSta 2018) will take place at the City University of Hong Kong from Tuesday 19 to Thursday 21 of June 2018. You are invited to participate actively in these events.

Ana Colubi, Erricos J. Kontoghiorghes and Mike K.P. So
on behalf of the Co-Chairs

CMStatistics: ERCIM Working Group on COMPUTATIONAL AND METHODOLOGICAL STATISTICS

http://www.cmstatistics.org

The working group (WG) CMStatistics comprises a number of specialized teams in various research areas of computational and methodological statistics. The teams act autonomously within the framework of the WG in order to promote their own research agenda. Their activities are endorsed by the WG. They submit research proposals, organize sessions, tracks and tutorials during the annual WG meetings and edit journal special issues. The Econometrics and Statistics (EcoSta) and Computational Statistics & Data Analysis (CSDA) are the official journals of the CMStatistics.

Specialized teams

Currently the ERCIM WG has over 1650 members and the following specialized teams

BM: Bayesian Methodology
CODA: Complex data structures and Object Data Analysis
CPEP: Component-based methods for Predictive and Exploratory Path modeling
DMC: Dependence Models and Copulas
DOE: Design Of Experiments
EF: Econometrics and Finance
GCS: General Computational Statistics WG CMStatistics
GMS: General Methodological Statistics WG CMStatistics
GOF: Goodness-of-Fit and Change-Point Problems
HDS: High-Dimensional Statistics
ISDA: Imprecision in Statistical Data Analysis
LVSEM: Latent Variable and Structural Equation Models
MCS: Matrix Computations and Statistics
MM: Mixture Models
MSW: Multi-Set and multi-Way models
NPS: Non-Parametric Statistics
OHEM: Optimization Heuristics in Estimation and Modelling
RACDS: Robust Analysis of Complex Data Sets
SAE: Small Area Estimation
SAET: Statistical Analysis of Event Times
SAS: Statistical Algorithms and Software
SEA: Statistics of Extremes and Applications
SFD: Statistics for Functional Data
SL: Statistical Learning
SSEF: Statistical Signal Extraction and Filtering
TSMC: Times Series Modelling and Computation

You are encouraged to become a member of the WG. For further information please contact the Chairs of the specialized groups (see the WG's website), or by email at info@cmstatistics.org.

CFEnetwork COMPUTATIONAL AND FINANCIAL ECONOMETRICS

http://www.CFEnetwork.org

The Computational and Financial Econometrics (CFEnetwork) comprises a number of specialized teams in various research areas of theoretical and applied econometrics, financial econometrics and computation, and empirical finance. The teams contribute to the activities of the network by organizing sessions, tracks and tutorials during the annual CFEnetwork meetings, and by submitting research proposals. Furthermore the teams edit special issues currently published under the Annals of CFE. The Econometrics and Statistics (EcoSta) is the official journal of the CFEnetwork.

Specialized teams

Currently the CFEnetwork has over 1000 members and the following specialized teams

AE: Applied Econometrics
BE: Bayesian Econometrics
BM: Bootstrap Methods
CE: Computational Econometrics
ET: Econometric Theory
FA: Financial Applications
FE: Financial Econometrics
TSE: Time Series Econometrics

You are encouraged to become a member of the CFEnetwork. For further information please see the website or contact by email at info@cfenetwork.org.

SCHEDULE

2017-06-14	2017-06-15	2017-06-16	2017-06-17
	Opening, 08:45 - 09:00	**F** EcoSta2017 08:30 - 09:50	**K** EcoSta2017 08:30 - 09:50
	A - Keynote EcoSta2017 09:00 - 09:50		
	Coffee Break 09:50 - 10:25	**Coffee Break** 09:50 - 10:25	**Coffee Break**, 09:50 - 10:15
	B EcoSta2017 10:25 - 12:30	**G** EcoSta2017 10:20 - 12:25	**L** EcoSta2017 10:15 - 11:30
			M - Keynote EcoSta2017 11:40 - 12:30
	Lunch Break 12:30 - 14:00	**Lunch Break** 12:30 - 14:00	**Lunch Break** 12:30 - 14:00
		H - Keynote EcoSta2017 14:00 - 14:50	
	C EcoSta2017 14:00 - 15:40	**I** EcoSta2017 15:00 - 16:40	**N** EcoSta2017 14:00 - 15:40
	Coffee Break 15:40 - 16:10	**Coffee Break** 16:40 - 17:10	**Coffee Break** 15:40 - 16:10
	D EcoSta2017 16:10 - 17:25	**J** EcoSta2017 17:10 - 18:50	**O** EcoSta2017 16:10 - 17:50
Registration & Ice Breaker 17:30 - 19:00	**E - Keynote** EcoSta2017 17:40 - 18:30		**Closing**, 17:55 - 18:10
			Closing Drink, 18:15 - 18:40
	Welcome Reception 18:45 - 20:15	**Conference Dinner** 19:45 - 22:15	

MEETINGS AND SOCIAL EVENTS

SPECIAL MEETINGS by invitation to group members

- The *Econometrics and Statistics (EcoSta) Editorial Board* meeting will take place on Saturday 17th of June 2017, 09:50-10:50, Room LSK1032. The meeting is by invitation only.
- The *Econometrics and Statistics (EcoSta) Editorial Board* dinner will take place on Saturday 17th of June 2017, 20:00-22:30. A minibus will depart from the conference venue at 19:20. The dinner is by invitation only.

SOCIAL EVENTS

- *Registration & Ice breaker, Wednesday 14th of June 2017, from 17:30 - 19:00*. The Ice breaker is open to all registrants and accompanying persons. It will take place at the Hall of the ground floor **(see map at page IX)**.
- *The coffee breaks* will take place at the Conference Lodge **(see map at page VIII)**. You must have your conference badge in order to attend the coffee breaks.
- *Welcome Reception, Thursday 15th of June 2017, from 18:45 - 20:15*. The Welcome Reception is open to all registrants and accompanying persons who have purchased a reception ticket. It will take place at the Lo Ka Chung University Center **(see maps at page VIII)**. Conference registrants must bring their conference badge and ticket and any accompanying persons shall bring their reception tickets in order to attend the reception. Preregistration is required due to health and safety reasons, and limited capacity of the venue. Entrance to the reception venue will be strictly allowed only to those who have a ticket. The Welcome Reception is fully sponsored by the HKUST Business School.
- *Conference Dinner, Friday 16th of June, from 19:45 to 22:15*. The conference dinner is optional and registration is required. It will take place at the Conference Lodge **(see map at page VIII)**. Conference registrants and accompanying persons shall bring their conference dinner tickets in order to attend the conference dinner.
- *Conference Buffet Lunches*. The conference lunches are optional and registration is required. The Lunch will take place at the Conference Lodge the 15th, 16th and 17th of June 2017. Conference registrants and accompanying persons shall bring their conference lunch tickets in order to attend the conference lunches.
- *Conference Lunch box*. The conference lunch box is optional and registration is required. The Lunch box will be arranged at the 7th floor, Lounge and Function room **(see map at pages VIII and X)** of the venue (Lee Shau Kee Business Building) the 15th, 16th and 17th of June 2017. Conference registrants and accompanying persons shall have the corresponding Lunch box ticket in order to attend the lunch each day.
- *Closing Drink, Saturday 17th of June 2017, from 18:15 - 18:40*. The Closing Drink is open to all registrants. It will take place at the Hall of the ground floor **(see map at pages VIII and IX)**.

GENERAL INFORMATION

Addresses of venues

- Lee Shau Kee Business Building, The Hong Kong University of Science and Technology (HKUST) Business School, Clear Water Bay, Kowloon Hong Kong.

Registration

The registration will be open from Wednesday afternoon 14th until Saturday 17th June 2017 at the ground floor of the conference venue.

Lecture rooms

The paper presentations will take place at the ground and first floor of the Lee Shau Kee Business Building **(see map in page IX)**. The different rooms are shown in the following floor plans of the venue. We advise that you visit the venue in advance. The opening and keynote talks will take place at the Auditorium LTA – Citi Lecture Theater (see map in page VIII) and the closing talk will take place at room LSKG001 of the venue (see map in page IX).

Presentation instructions

The lecture rooms will be equipped with a PC and a computer projector. The session chairs should obtain copies of the talks on a USB stick before the session starts (use the lecture room as the meeting place), or obtain the talks by email prior to the start of the conference. Presenters must provide the session chair with the files for the presentation in PDF (Acrobat) or PPT (Powerpoint) format on a USB memory stick. This must be done at least ten minutes before each session. Chairs are requested to keep the sessions on schedule. Papers should be presented in the order they are listed in the programme for the convenience of attendees who may wish to go to other rooms mid-session to hear particular papers. In the case of a presenter not attending, please use the extra time for a break or a discussion so that the remaining papers stay on schedule. The PC in the lecture rooms should be used for presentations. An IT technician will be available during the conference and should be contacted in case of problems.

Posters

The poster sessions will take place at the Hall of the ground floor. The posters should be displayed only during their assigned session. The authors will be responsible for placing the posters in the poster panel displays and removing them after the session. The maximum size of the poster is A0.

Internet connection

The information for the wireless Internet connection will be displayed by the registration desk.

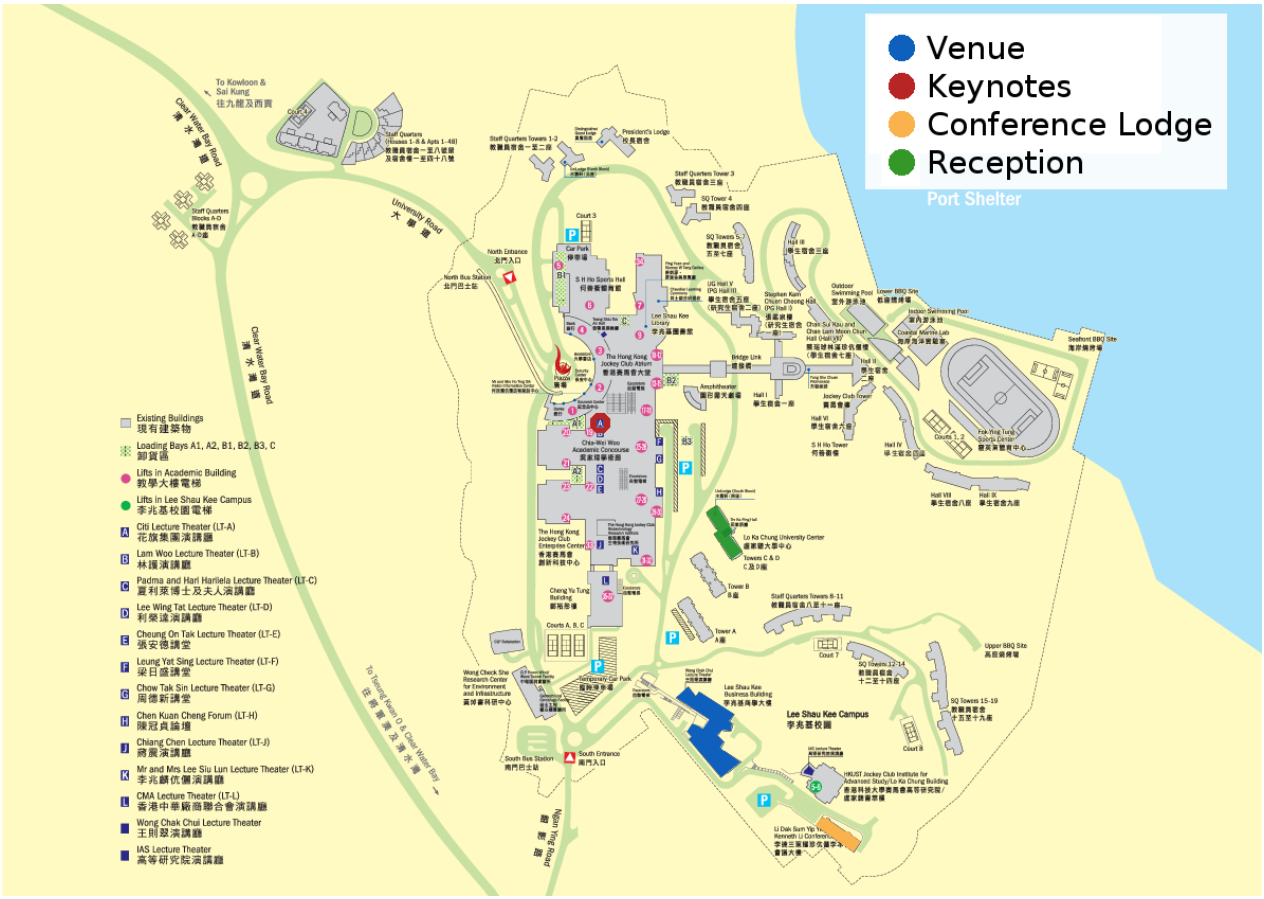
Information and messages

You may leave messages for each other on the bulletin board by the registration desks. General information about restaurants, useful numbers, etc. can be obtained from the registration desk.

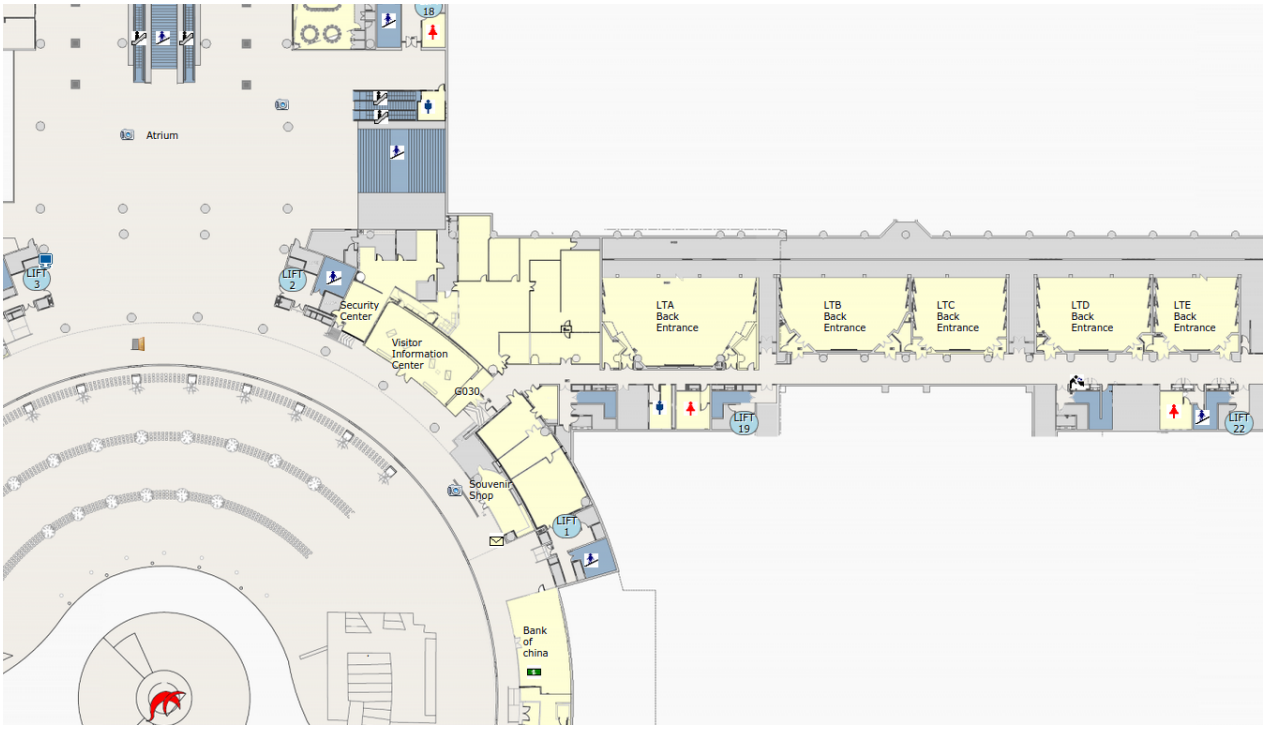
Exhibitors

Elsevier.

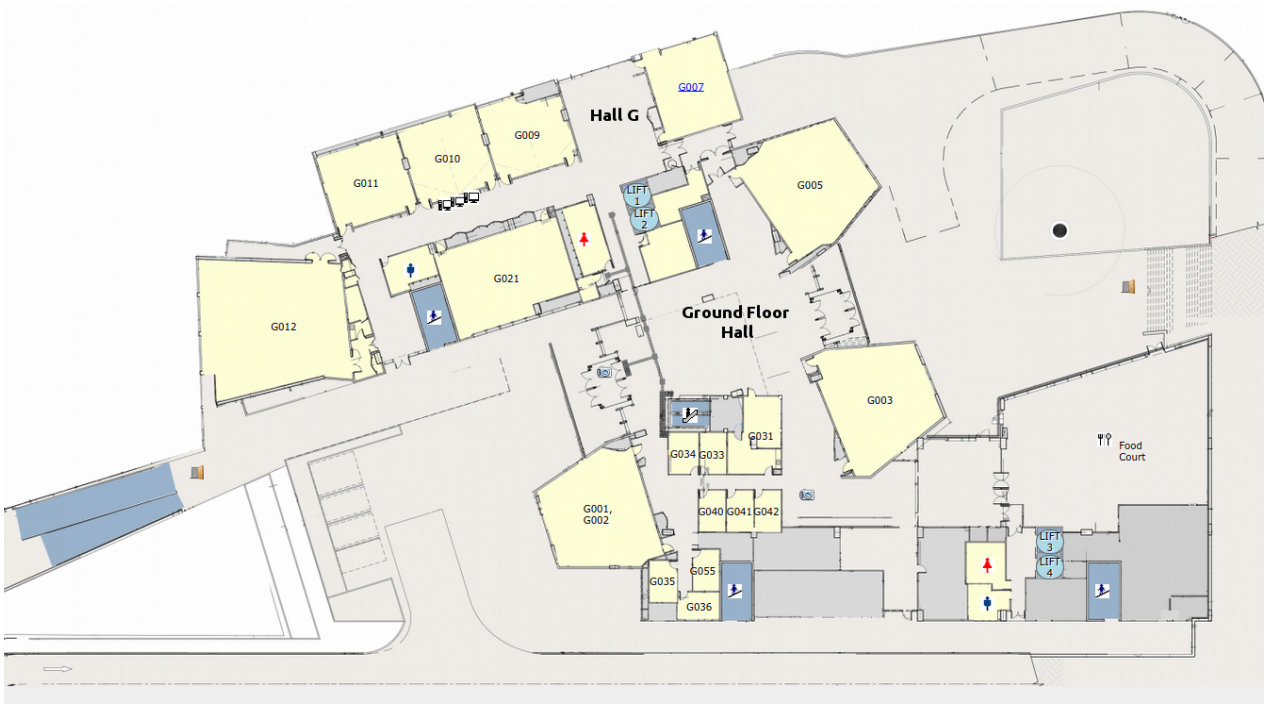
Map of the venue and nearby area



LTA - Keynotes

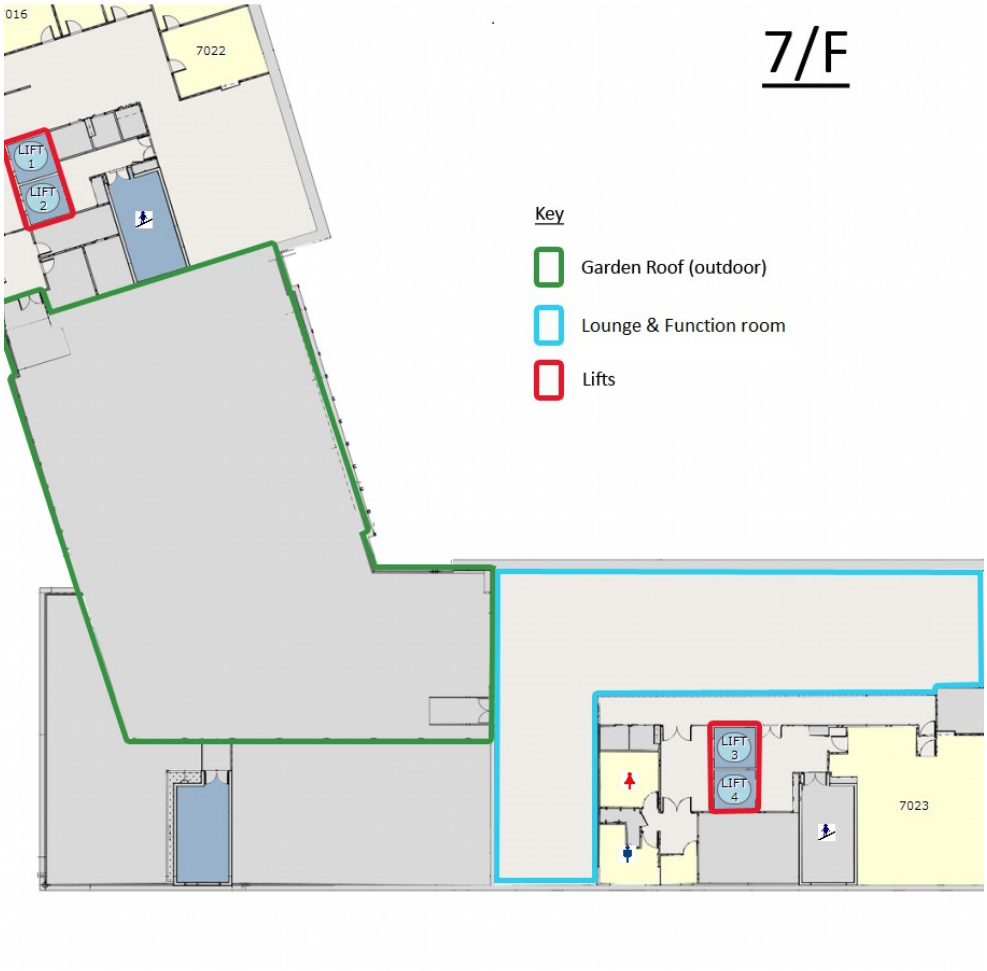


Venue: LSK - Ground Floor



Venue: LSK - First Floor

Venue: LSK - Conference Lunch box



PUBLICATION OUTLETS

Econometrics and Statistics (EcoSta)

http://www.elsevier.com/locate/ecosta

Econometrics and Statistics (EcoSta), published by Elsevier, is the official journal of the networks Computational and Financial Econometrics (CFENetwork) and Computational and Methodological Statistics (CMStatistics). It publishes research papers in all aspects of econometrics and statistics and comprises two sections:

Part A: Econometrics. Emphasis is given to methodological and theoretical papers containing substantial econometrics derivations or showing a potential of a significant impact in the broad area of econometrics. Topics of interest include the estimation of econometric models and associated inference, model selection, panel data, measurement error, Bayesian methods, and time series analyses. Simulations are considered when they involve an original methodology. Innovative papers in financial econometrics and its applications are considered. The covered topics include portfolio allocation, option pricing, quantitative risk management, systemic risk and market microstructure. Well-founded applied econometric studies that demonstrate the practicality of new procedures and models are of interest as well. Such studies should involve the rigorous application of statistical techniques, including estimation, inference and forecasting. Topics include volatility and risk, credit risk, pricing models, portfolio management, and emerging markets. Innovative contributions in empirical finance and financial data analysis that use advanced statistical methods are encouraged. The results of the submissions should be replicable. Applications consisting only of routine calculations are not of interest to the journal.

Part B: Statistics. Papers providing important original contributions to methodological statistics inspired in applications are considered for this section. Papers dealing, directly or indirectly, with computational and technical elements are particularly encouraged. These cover developments concerning issues of high-dimensionality, re-sampling, dependence, robustness, filtering. In general, the interaction of mathematical methods, numerical implementations and the extra burden of analysing large and/or complex datasets with such methods in different areas such as medicine, epidemiology, biology, psychology, climatology and communication. Innovative algorithmic developments are also of interest, as are the computer programs and the computational environments that implement them as a complement.

The journal consists, preponderantly, of original research. Occasionally, reviews and short papers from experts are published, which may be accompanied by discussions. Special issues and sections within important areas of research are occasionally published. The journal publishes as a supplement the Annals of Computational and Financial Econometrics.

Call For Papers Econometrics and Statistics (EcoSta)

http://www.elsevier.com/locate/ecosta

Papers containing novel components in econometrics and statistics are encouraged to be submitted for publication in special peer-reviewed, or regular issues of the new Elsevier journal Econometrics and Statistics (EcoSta) and its supplement Annals of Computational and Financial Econometrics. The Econometrics and Statistics (EcoSta) is inviting submissions for the special issues:

- (Part A: Econometrics) Annals of Computational and Financial Econometrics
- (Part A: Econometrics) Special Issue on Forecast combinations.
- (Part A: Econometrics) Special Issue on Risk management.
- (Part B: Statistics) Special Issue on Quantile regression and semiparametric methods.
- (Part B: Statistics) Special Issue on Statistics of extremes and applications.

The deadline for paper submissions is the 30th June 2017. Papers should be submitted using the Elsevier Electronic Submission tool EES: http://ees.elsevier.com/ecosta (in the EES please select the appropriate special issue). For further information please consult http://www.cfenetwork.org or http://www.cmstatistics.org.

Call For Papers Computational Statistics & Data Analysis (CSDA)

http://www.elsevier.com/locate/csda

Papers containing strong computational statistics, or substantive data-analytic elements can also be submitted to special peer-reviewed, or regular issues of the journal Computational Statistics & Data Analysis (CSDA). The CSDA is planning for 2017 the following special issues:

- 4th Special Issue on advances in mixture models
- Special Issue on Biostatistics
- High-dimensional and functional data analysis

The deadline for paper submissions is the 30th November 2017. Papers should be submitted using the Elsevier Electronic Submission tool EES: http://ees.elsevier.com/csda (in the EES please select the appropriate special issue). Any questions may be directed via email to: csda@dcs.bbk.ac.uk.

Contents

Thursday 15.06.2017 09:00 - 09:50 Room: LTA - Citi Lecture Theater Chair: Ana Colubi Keynote talk 1

A statistical tale of subgroup analysis for managerial decision making
Speaker: **Xuming He, University of Michigan, United States**

Clinical trials are often used for the assessment of therapeutic benefits. Subgroup analysis is routinely used in the clinical studies to understand heterogeneity of treatment effects in subgroups of patients. When used appropriately, subgroup analysis helps personalized treatments of patients and better designs of new studies. When used indiscriminately, subgroup analysis leads to costly false discoveries. We will start with a partial review of subgroup identification and confirmation methods, and examine the benefits and pitfalls of subgroup analysis associated with decision making in the pharmaceutical industry. We will discuss some important questions of how the risk for subgroup pursuit needs to be quantified to help managerial decision making. In this process, we identify several interesting statistical questions.

Thursday 15.06.2017 17:40 - 18:30 Room: LTA - Citi Lecture Theater Chair: Alan Wan Keynote talk 2

Quantile spectral analysis for locally stationary time series
Speaker: **Marc Hallin, Universite Libre de Bruxelles, Belgium** Stefan Birr, Holger Dette, Stanislav Volgushev

Classical spectral methods are subject to two fundamental limitations: they only can account for covariance-related serial dependencies, and they require second-order stationarity. Much attention has been devoted lately to quantile-based spectral methods that go beyond covariance-based serial dependence features. At the same time, covariance-based methods relaxing stationarity into much weaker *local stationarity* conditions have been developed for a variety of time-series models. We are combining those two approaches by proposing quantile-based spectral methods for locally stationary processes. We therefore introduce a time-varying version of the copula spectra that have been recently proposed in the literature, along with a suitable local lag-window estimator. We propose a new definition of local *strict* stationarity that allows us to handle completely general non-linear processes without any moment assumptions, thus accommodating our quantile-based concepts and methods. We establish a central limit theorem for the new estimators, and illustrate the power of the proposed methodology by means of a simulation study. Moreover, in two empirical studies, we demonstrate that the new approach detects important variations in serial dependence structures both across time and across quantiles. Such variations remain completely undetected, and are actually undetectable, via classical covariance-based spectral methods.

Friday 16.06.2017 14:00 - 14:50 Room: LTA - Citi Lecture Theater Chair: Erricos Kontoghiorghes Keynote talk 3

Robust normal mixtures for financial portfolio allocation
Speaker: **Marc Paolella, University of Zurich, Switzerland**

A new approach for multivariate modeling and prediction of asset returns is proposed. It is based on a two-component normal mixture, estimated using a fast new variation of the minimum covariance determinant (MCD) method made suitable for time series. It outperforms the (shrinkage-augmented) MLE in terms of out-of-sample density forecasts and portfolio performance. In addition to the usual stylized facts of skewness and leptokurtosis, the model also accommodates leverage and contagion effects, but is i.i.d., and thus does not embody, for example, a GARCH-type structure. Owing to analytic tractability of the moments and the expected shortfall, portfolio optimization is straightforward, and, for daily equity returns data, is shown to substantially outperform the equally weighted and classical long-only Markowitz framework, as well as DCC-GARCH (despite not using any kind of GARCH-type filter).

Saturday 17.06.2017 11:40 - 12:30 Room: LTA - Citi Lecture Theater Chair: Mike So Keynote talk 4

Recent developments in Bayesian inference for time series
Speaker: **Michael Pitt, Kings College London, United Kingdom**

The pseudo-marginal algorithm is a variant of the Metropolis Hastings scheme which samples asymptotically from a target probability density when we are only able to estimate unbiasedly an unnormalized version of it. It has found numerous applications in statistics and econometrics as there are many scenarios where the likelihood function is intractable but can be estimated unbiasedly using Monte Carlo samples. Several recent contributions will be discussed which optimise the trade off between computational complexity and statistical efficiency. A modification of the pseudo-marginal algorithm, termed the correlated pseudo-marginal algorithm, is introduced. Guidelines are provided for the optimal settings of this algorithm. The computational gains of the new algorithm are demonstrated by examining large time series, including the estimation of continuous time volatility models.

Thursday 15.06.2017 10:25 - 12:30 Parallel Session B – EcoSta2017

EO116 Room LSK1005 NEW DEVELOPMENT IN ANALYZING LARGE COMPLEX DATA Chair: Xuming He

EO0152: **Lack-of-fit tests for quantile regression models**
Presenter: **Xingdong Feng**, Shanghai University of Finance and Economics, China
The aim is to novelly transforms lack-of-fit tests for parametric quantile regression models into checking the equality of two conditional distributions of covariates. We then can borrow these successful test statistics from the rich literature of two-sample problems, and this gives us much flexibility in constructing a reliable test according our experiences on covariates. As an illustration, three two-sample test statistics are considered. The first one will lead to an already known test in this context when a Cramer-von Mises test statistic is employed. The second one is a practical two-sample test statistic especially for multivariate distributions, and the resulting test is still applicable in real applications when the number of covariates is moderate or even large. In the last case, we provide a lack-of-fit test based on two-sample tests on moments for the high dimensional data. Its usefulness is demonstrated by simulation experiments and a real example.

EO0591: **Test for high dimensional regression coefficients using refitted cross-validation variance estimation**
Presenter: **Wei Zhong**, Xiamen University, China
Testing a hypothesis for high dimensional regression coefficients is of fundamental importance in the statistical theory and applications. The aim is to develop a new test for coefficients in high dimensional linear regression models based on an estimated U-statistics of order two. With the aid of martingale central limit theorem, we prove that the asymptotic null distributions of the proposed test are normal under two different distribution assumptions. The idea of refitted cross-validation (RCV) approach is utilized to reduce the bias of the sample variance in the estimation of the test statistic. We assess the finite-sample performance of the proposed test by examining its size and power via Monte Carlo simulations which show that the new test based on the RCV estimator of the variance achieves higher powers, especially for the sparse cases. We also illustrate the application of the proposed test by an empirical analysis of a microarray data set on Yorkshire gilts.

EO0651: **Model-free variable selection**
Presenter: **Junhui Wang**, City University of Hong Kong, Hong Kong
Co-authors: Xin He, Shaogao Lyu
Variable selection is central to sparse modeling, and many methods have been proposed under various model assumptions. We will present a model-free variable selection method that allows for general variable effects. As opposed to most existing methods based on an explicit functional relationship, the proposed method attempts to identify non-informative variables that are conditional independent with the response by simultaneously examining the sparsity in multiple conditional quantile functions. It does not require specification of the underlying model for the response, which is appealing in sparse modeling with a relatively large number of variables. The proposed method is implemented via an efficient computing algorithm, which couples the majorize-minimization algorithm and the proximal gradient descent algorithm. The effectiveness of the proposed method is also supported by a variety of simulated and real-life examples. Its asymptotic estimation and variable selection consistencies are established without explicit model assumption.

EO0829: **High dimensional variable selection for longitudinal data with covariate measurement error and dropout**
Presenter: **Yang Bai**, Shanghai University of Finance and Economics, China
A novel approach is developed for high-dimensional variable selection in longitudinal data with covariate measurement error and dropout. Specifically, new penalized estimating equations are proposed for variable selection, while measurement error and dropout are dealt with simultaneously. A modified coordinate descent algorithm for solving this specific variable selection problem is developed. The asymptotic properties of the proposed estimators are established in high dimensional settings. We demonstrate that the proposed approach is consistent in variable selection and achieves the oracle property under regularity conditions. Finite sample performance of the proposed method is evaluated via both extensive Monte Carlo simulation and a real data analysis in systematic lupus erythematosus.

EO0249: **Statistical inference for the single index hazards model**
Presenter: **Sheng Xu**, The Hong Kong Polytechnic University, Hong Kong
Co-authors: Jicai Liu, Catherine Liu, Xihong Lin
A class of consistent estimators is developed for the parameter vector for the single-index hazards model, where the model allows nonparametric modeling of covariate effects in a parsimonious way through a single index and circumvent the curse of dimensionality. A class of consistent profile likelihood estimators is proposed by estimating the nonparametric part via local linear smoothing tool, and hence the estimator of the parametric component is shown to be asymptotically normal and even achieves the semiparametric efficiency bound. Motivated by the semiparametric efficient score, we propose two classes of estimation equations such that their roots enjoys doubly robust properties. We present some algorithm procedures for validity of model checking. Extensive numerical analysis are implemented so as to assess the finite-sample properties of our proposed estimators compared with other existed methods. We also demonstrate our method by several real-world data application.

EO018 Room LSK1007 STATISTICAL MODELLING FOR NETWORK DATA Chair: Yang Feng

EO0163: **Graph-based change-point analysis for object data**
Presenter: **Hao Chen**, University of California at Davis, United States
After observing snapshots of a network, the aim is to check whether there has been a change in dynamics. We develop a general nonparametric framework for change-point detection that relies on a distance metric on the sample space of observations. This new approach, which relies on graph-based tests, can be applied to high dimensional data, as well as data from non-Euclidean sample spaces, such as network data. An analytic approximation for the false positive error probability is derived and shown to be reasonably accurate by simulation. We illustrate the method through the analysis of a phone-call network from the MIT Reality Mining.

EO0203: **A semidefinite program for the stochastic blockmodel**
Presenter: **David Choi**, Carnegie Mellon University, United States
Semidefinite programs have recently been developed for the problem of community detection, which may be viewed as a special case of the stochastic blockmodel. We develop a semidefinite program that can be tailored to other instances of the blockmodel, such as non-assortative networks and overlapping communities. We establish label recovery in sparse settings, with conditions that are analogous to recent results for community detection. In settings where the data is not generated by a blockmodel, we give an oracle inequality that bounds excess risk relative to the best blockmodel approximation. Simulations are presented for community detection, for overlapping communities, and for latent space models.

EO0519: **Estimating network memberships by simplex vertices hunting**
Presenter: **Tracy Ke**, University of Chicago, United States

Consider an undirected mixed-membership network with n nodes and K communities. For each node i, we model the membership by a Probability Mass Function (PMF) $\pi_i = (\pi_i(1), \pi_i(2), \ldots, \pi_i(K))'$, where $\pi_i(k)$ is the probability that node i belongs to community k. We call node i 'pure' if π_i is degenerate and 'mixed' otherwise. The primary interest is to estimate π_i, $1 \leq i \leq n$. We propose Mixed-SCORE as a new spectral method for mixed membership estimation. At the heart of Mixed-SCORE is a (tall by very skinny) matrix of entry-wise ratios, formed by dividing by the first few eigenvectors of the network adjacency matrix over the leading eigenvector of the same matrix in an entry-wise fashion. The main surprise is that the rows of the entry-wise ratio matrix form a cloud of points in a low-dimensional space with the silhouette of a simplex, which simplex carries all information we need for estimating the memberships. We apply Mixed SCORE to four network data sets (a coauthorship and a citee network for statisticians, a political book network, and a football network) and obtain meaningful results. We propose a Degree-Corrected Mixed Membership (DCMM) model, and use it to solidify our discoveries with delicate spectral analysis and Random Matrix Theory.

EO0628: **Random walk models of network formation**
Presenter: **Peter Orbanz**, Columbia University, United States

A class of network models are described that insert edges by connecting the starting and terminal vertices of a random walk on the network graph. Within the taxonomy of statistical network models, this class is distinguished by permitting the location of a new edge to explicitly depend on the structure of the graph, but being nonetheless statistically and computationally tractable. In the limit of infinite walk length, the model converges to an extension of the preferential attachment model. Theoretical properties will be discussed, such as power laws, and show that inference of model parameters is possible from a single graph generated by the model.

EO0192: **Multivariate spatial autoregression for large scale social network**
Presenter: **Xuening Zhu**, Peking University, China

The rapid growth of social network platforms generates a large amount of social network data. As a result, multivariate responses and the corresponding predictors can be collected for social network users. To statistically model such type of data, the multivariate spatial autoregression (MSAR) model is proposed and studied. To estimate the model, the maximum likelihood estimator (MLE) is obtained under certain technical conditions. However, it is found that the computational cost of MLE is expensive. In order to fix this problem, a least squares type estimator is developed. The corresponding asymptotic properties are investigated. To gauge the finite sample performance of the proposed estimators, a number of numerical studies are conducted. Lastly, a Sina Weibo dataset is analyzed for illustration purpose.

EO136 Room LSK1011 STATISTICAL METHODS FOR FUNCTIONAL DATA Chair: Yuhang Xu

EO0189: **Nested hierarchical functional data modeling and inference for evaluating lunar effects on root gravitropism**
Presenter: **Yuhang Xu**, University of Nebraska - Lincoln, United States
Co-authors: Dan Nettleton

In a plant science study, the process of roots bending in response to gravity is recorded. The data are collected from seeds representing a large variety of genotypes and have a three-level nested hierarchical structure. We allow the mean function of the bending rate to depend on the lunar day and model the variation between genotypes, groups of seeds imaged together, and individual seeds by hierarchical functional random effects. We estimate the covariance functions by a fast penalized tensor product spline approach, perform multi-level functional PCA using the BLUP of the principal component scores, and improve the efficiency of mean estimation by iterative decorrelation. We choose the number of principal components using a conditional AIC and test the lunar day effect using generalized likelihood ratio test statistics based on the marginal and conditional likelihoods. We propose a permutation procedure to evaluate the null distribution of the test statistics. Simulation studies show that our model selection criterion selects the correct number of principal components with high frequency, and the likelihood-based tests based on functional PCA have higher power than a test based on working independence. Our data analysis suggests that the root bending behavior may be associated with moon phases.

EO0234: **Function-on-function regression for highly densely observed spiky functional data**
Presenter: **Ruiyan Luo**, Georgia State University, United States

Modern techniques allow data to be recorded with high sample rate, which leads to highly densely observed spiky curves. For example, the mass spectrometry data contains a number of narrow and high peaks which are interests of scientists, and the EEG curves exhibit high local variations over the whole time interval. The existing methods for function-on-function regression assume that the coefficient functions are smooth (usually they are assumed to belong to the Sobolev space), and impose smoothness regularities in various ways. However, the smoothness assumption makes it difficult to model the associations between high local variations in response curve and predictor curves. We model the coefficient functions in a more general family of function spaces, where various levels of local variations are possible. We propose a new regularization method to replace the traditional smoothing method in functional data analysis, and apply it to our recently developed signal compression method in function-on-function regression. In addition to capturing the association between high local variations, such as rapid peaks, this method has good prediction performance and can handle multiple functional predictors with thousands of observation time points.

EO0235: **Nonparametric operator-regularized covariance function estimation**
Presenter: **Raymond Wong**, Iowa State University, United States
Co-authors: Xiaoke Zhang

A class of nonparametric covariance function estimators is developed by utilizing spectral regularization of an operator, which is associated with a typically infinite dimensional reproducing kernel Hilbert space. By construction, these estimators are positive semi-definite and hence valid covariance functions. A related representer theorem is established to provide a finite dimensional representation of such estimators. In order to achieve low-rank estimations, trace-norm regularization is studied in detail. A specific computational algorithm is developed and this estimator is shown to enjoy excellent rates of convergence under either fixed or random designs. The empirical performance of the proposed trace-norm-regularized estimator is demonstrated in a simulation study, while its practical utility is illustrated in an analysis of a traffic data set.

EO0348: **Hypothesis testing in functional linear models**
Presenter: **Yu-Ru Su**, Fred Hutchinson Cancer Research Center, United States
Co-authors: Chong-Zhi Di, Li Hsu

Functional data arise frequently in biomedical studies, where it is often of interest to investigate the association between functional predictors and a scalar response variable. While functional linear models (FLM) are widely used to address these questions, hypothesis testing for the functional association in the FLM framework remains challenging. A popular approach to testing the functional effects is through dimension reduction by functional principal component (PC) analysis. However, its power performance depends on the choice of the number of PCs, and is not systematically studied. We first present the power performance of the Wald-type test with varying thresholds in selecting the number of PCs for the functional covariates, and show that the power is sensitive to the choice of thresholds. To circumvent the issue, we propose a new method of ordering and selecting principal components to construct test statistics. The proposed method takes into account both the association

with the response and the variation along each eigenfunction. We establish its theoretical properties and assess the finite sample properties through simulations. Our simulation results show that the proposed test is more robust against the choice of threshold while being as powerful as, and often more powerful than, the existing method. We then apply the proposed method to the cerebral white matter tracts data obtained from a diffusion tensor imaging tractography study.

EO0838: **Component selection and estimation for functional additive models**
Presenter: **Hao Zhang**, University of Arizona, United States
Functional additive model provides a flexible yet simple framework for regressions involving functional predictors. The utilization of data-driven basis in an additive rather than linear structure naturally extends the classical functional linear model. However, the critical issue of selecting nonlinear additive components has been less studied. We propose a new regularization framework for joint component selection and estimation in the context of the Reproducing Kernel Hilbert Space. The proposed approach takes advantage of the functional principal components which greatly facilitates the implementation and the theoretical analysis. The selection and estimation are achieved by penalized least squares using a penalty which encourages the sparse structure of the additive components. Theoretical properties, such as the existence and the rate of convergence are investigated. The empirical performance is demonstrated through simulation studies and a real data application.

EO108 Room LSK1001 MODEL AVERAGING, SELECTION AND SHRINKAGE Chair: Alan Wan

EO0205: **PMSE performance of two different types of preliminary test estimators under a multivariate t error**
Presenter: **Haifeng Xu**, Xiamen University, China
Co-authors: Kazuhiro Ohtani
Assuming that the error terms follow a multivariate t distribution, the exact formula is derived for the predictive mean squared error (PMSE) of two different types of preliminary test estimators: (1) a homogeneous pre-test (HO-PT) estimator whose components are the adjusted minimum mean squared error (AMMSE) estimator and the minimum mean squared error (MMSE) estimator; (2) a heterogeneous pre-test (HE-PT) estimator whose components are the AMMSE estimator and the Stein-rule (SR) estimator. It is shown analytically that the HE-PT estimator dominates the SR estimator if a critical value of the pre-test is chosen appropriately. Also, we compare the PMSE of the HO-PT, HE-PT, MMSE, AMMSE, SR and PSR estimators by numerical evaluations. Our results show that 1. the HO-PT and HE-PT estimators dominate the OLS estimator for all combinations when the degrees of freedom is not more than 5; 2. if the number of independent variables is 3, and the critical value of the pre-test is chosen appropriately, then the HE-PT estimator dominates the PSR estimator even when error terms follow a multivariate t distribution.

EO0391: **Dominance of the positive-part shrinkage estimator when each individual regression coefficient is estimated**
Presenter: **Akio Namba**, Kobe University, Japan
Co-authors: Haifeng Xu
Assuming that there exist omitted variables in the specified model, the aim is to analytically derive the exact formula for the mean squared error (MSE) of a general family of shrinkage estimators for each individual regression coefficient. It is shown analytically that when our concern is to estimate each individual regression coefficient, the positive-part shrinkage estimators dominate the shrinkage estimators under some conditions even when the relevant regressors are omitted. Also, by numerical evaluations, we showed the effects of our theorem for several specific cases. It is shown that the positive-part shrinkage estimators dominate shrinkage estimators for large region of parameter space even when there exist omitted variables in the specified model.

EO0221: **Conditionally optimal weights: Optimal combination with predictable errors**
Presenter: **Andrey Vasnev**, University of Sydney, Australia
A classical unconditional framework is extended and conditionally optimal weights are constructed that it can be used to combine individual forecasts. Often there is an information set which is available when the combination is constructed. The previous forecast errors can be included in this information set, but it could also contain other variables. If conditionally on this information the forecast errors are predictable, then the conditional mean squared error (MSE) of the combination needs to be minimized rather than unconditional variance used in the classical framework. We prove that the new conditionally optimal weights produce combinations with smaller expected MSE than the classical unconditionally optimal weights. Our empirical study of the European Central Bank Survey of Professional Forecasters confirms theoretical findings and shows that the new weights outperform a wide range of other methods.

EO0219: **Model averaging in a multiplicative heteroscedastic model**
Presenter: **Alan Wan**, City University of Hong Kong, Hong Kong
In recent years, the literature of frequentist model averaging in econometrics has grown significantly. Models with different mean structures have been considered, but variance considerations have been left out. We consider a regression model with multiplicative heteroscedasticity and develop a model averaging method that combines maximum likelihood estimators of unknown parameters in both the mean and variance functions of the model. Our weight choice criterion is based on a minimisation of a plug-in estimator of the model average estimator's squared prediction risk. We prove that the new estimator possesses an asymptotic optimality property. Our investigation of finite-sample performance by simulations demonstrates that the new estimator frequently exhibits very favourable properties compared to some existing heteroscedasticity-robust model average estimators. The model averaging method hedges against the selection of very bad models and serves as a remedy to variance function mis-specification, which often discourages practitioners from modeling heteroscedasticity altogether. The proposed model average estimator is applied to the analysis of two data sets on housing and economic growth.

EO0225: **Focused information criterion and model averaging for large panels with a multifactor error structure**
Presenter: **Chu-An Liu**, Academia Sinica, Taiwan
Co-authors: Chang-Ching Lin, Shou-Yung Yin
Model selection and model averaging is considered for panel data models with a multifactor error structure. We investigate the limiting distribution of the common correlated effects estimator in a local asymptotic framework and show that the trade-off between bias and variance remains in the asymptotic theory. We then propose a focused information criterion and a plug-in averaging estimator for large heterogeneous panels and examine their theoretical properties. The novel feature of the proposed method is that it aims to minimize the sample analogue of the asymptotic mean squared error and can be applied to cases irrespective of whether the rank condition holds or not. Monte Carlo simulations show that both proposed selection and averaging methods generally achieve lower expected squared error than other methods. The proposed methods are applied to analyze the consumer response to gasoline taxes.

EO112 Room LSK1034 NEW DEVELOPMENTS IN FINANCIAL ECONOMETRICS Chair: Daniel Preve

EO0242: **Asymptotic properties of maximum likelihood estimators of a multiplicative time-varying correlation GARCH model**
Presenter: **Timo Terasvirta**, Aarhus University, Denmark
A new multivariate volatility model that belongs to the family of conditional correlation GARCH models is introduced. The GARCH equations of this model contain a multiplicative deterministic component to describe long-run movements in volatility and, in addition, the correlations are deterministically time-varying. Parameters of the model are estimated jointly using maximum likelihood. Consistency and asymptotic normality of maximum likelihood estimators is proved. Numerical aspects of the estimation algorithm are discussed. A bivariate empirical example is provided.

EO0449: **High dimensional minimum variance portfolio estimation with high-frequency data**
Presenter: **Yingying Li**, Hong Kong University of Science and Technology, Hong Kong
Co-authors: Tony Cai, Jianchang Hu, Xinghua Zheng
The estimation of high dimensional minimum variance portfolio (MVP) with high frequency financial data is considered. High frequency returns can exhibit heteroskedasticity and possiblybe contaminated by microstructure noise. Under some sparsity assumptions on the precision matrix, we propose an estimator of MVP, which asymptotically achieves the minimum variance. Simulation and empirical studies demonstrate that our proposed portfolio performs favorably.

EO0525: **A mixture autoregressive model based on Student's t-distribution**
Presenter: **Mika Meitz**, University of Helsinki, Finland
Co-authors: Daniel Preve, Pentti Saikkonen
A new mixture autoregressive model is proposed based on Student's t-distribution. A key feature of our model is that the conditional t-distributions of the component models are based on autoregressions that have t-distributions as their stationary distributions. That autoregressions with such stationary distributions exist is not immediate. Our formulation implies that the conditional variance of each mixture component is not constant but of (nonlinear) ARCH type. Compared to previous mixture autoregressive models our model may therefore be useful in applications where the data exhibits rather strong conditional heteroskedasticity. Our formulation also has the theoretical advantage that conditions for stationarity and ergodicity are always met and these properties are much more straightforward to establish than is common in nonlinear autoregressive models. An empirical example employing a realized kernel series based on S&P 500 data shows that the proposed model performs well in volatility forecasting.

EO0212: **The risk return relationship: Evidence from index return and realised variance series**
Presenter: **Minxian Yang**, The University of New South Wales, Australia
The risk return relationship is analysed in bivariate models for return and realised variance (RV) series. Based on daily time series from 21 international market indices for more than 13 years (January 2000 to February 2013), the empirical findings support the arguments of risk return tradeoff, volatility feedback and statistical balance. It is argued that the empirical risk return relationship is primarily shaped by two important data features: the negative contemporaneous correlation between the return and RV, and the difference in the autocorrelation structures of the return and RV. The findings do not support the risk premium effect of the shock to volatility as documented by recent studies that do not take into account of the contemporaneous correlation.

EO0508: **New distribution theory for the estimation of structural break point in mean**
Presenter: **Xiaohu Wang**, The Chinese University of Hong Kong, Hong Kong
Co-authors: Liang Jiang, Jun Yu
Based on the Girsanov theorem, the exact distribution of the maximum likelihood estimator of structural break point in a continuous time model is obtained. The exact distribution is asymmetric and tri-modal, indicating that the estimator is biased. These two properties are also found in the finite sample distribution of the least squares (LS) estimator of structural break point in the discrete time model, suggesting the classical long-span asymptotic theory is inadequate. A continuous time approximation to the discrete time model is built, and an in-fill asymptotic theory for the LS estimator is developed. The in-fill asymptotic distribution is asymmetric and tri-modal and delivers good approximations to the finite sample distribution. To reduce the bias in the estimation of both the continuous time and the discrete time models, a simulation-based method based on the indirect estimation (IE) approach is proposed. Monte Carlo studies show that IE achieves substantial bias reductions.

EO010 Room LSK1027 MODELLING FINANCIAL AND INSURANCE RISKS Chair: Tak Kuen Siu

EO0277: **On modeling credit defaults: A probabilistic Boolean network approach**
Presenter: **Wai-Ki Ching**, The University of Hong Kong, Hong Kong
One of the central issues in credit risk measurement and management is modeling and predicting correlated defaults. A novel model is introduced to investigate the relationship between correlated defaults of different industrial sectors and business cycles as well as the impacts of business cycles on modeling and predicting correlated defaults using the Probabilistic Boolean Network (PBN). The key idea of the PBN is to decompose a transition probability matrix describing correlated defaults of different sectors into several BN matrices which contain information about business cycles. An efficient estimation method based on entropy approach is used to estimate the model parameters. Using real default data, we build a PBN for explaining the default structure and make reasonably good prediction of joint defaults in different sectors.

EO0502: **Longevity-product valuation under correlated financial and mortality risks**
Presenter: **Rogemar Mamon**, University of Western Ontario, Canada
The pricing of annuity and guaranteed annuity option primarily depends on mortality and interest risk factors. Independence between these two risk factors greatly facilitates the valuation but such an assumption is not always realistic. We propose a pricing framework where the dependence between interest and mortality rates is modelled explicitly. We employ the change of measure technique in conjunction with the comonotonicity theoretic approach to approximate annuity and guaranteed annuity option prices. Our practical method provides accurate pricing values and is efficient as it circumvents the simulation-within-simulation problem.

EO0453: **Rating of financial products by implied risk aversion and optimized expected utility risk measures**
Presenter: **Joern Sass**, University of Kaiserslautern, Germany
Co-authors: Holger Fink, Sebastian Geissel, Frank Thomas Seifried
An optimal expected utility risk measure (OEU) is introduced which is generated by a utility function via an associated optimal investment problem: A financial product is evaluated by finding the capital to be borrowed and added to the position in order to maximize the discounted certainty equivalent of the future payoff. Properties of OEU are derived and put in relation to alternative risk measures. For constant relative risk aversion and for proper discounting, OEU is non-trivial and coherent. OEU reacts in a more sensitive way to slight changes of the probability of a financial loss than (average) value at risk. This motivates to use implied risk aversion based on OEU as a coherent rating methodology for structured financial products. This takes into account both upside potential and downside risks and is easily interpreted in terms of an individual investor's

risk aversion. The optimizer provides the additional information how much of the financial product can be added to the portfolio optimally. The approach is illustrated by a case study comprising different stock return models, about 15 000 put and call options, and a comparison with alternative rating systems.

EO0470: **Optimal portfolio and insurance problems under a value-at-risk constraint**
Presenter: **Cedric Yiu**, The Hong Kong Polytechnic University, Hong Kong
The portfolio selection problem will be considered. For ordinary investors or insurers who invest in the financial market, they need to manage portfolios continuously. At the same time, they need to fulfil the regulatory requirement governed by the value-at-risk constraints. This problem is addressed. For the investment, the goal is to maximize the expected utility of terminal wealth. By using the principle of dynamic programming, the Hamilton-Jacobi-Bellman (HJB) equation can be derived. We will examine a few scenarios with different stochastic processes such as regime switching, and discuss how to solve the resulting HJB equation. Furthermore, we will investigate the impacts of the risk constraint on the optimal strategies.

EO0652: **A self-exciting threshold jump-diffusion model for option valuation**
Presenter: **Tak Kuen Siu**, Macquarie University, Australia
A self-exciting threshold jump-diffusion model for option valuation will be discussed. This model incorporates regime switches without introducing an exogenous stochastic factor process. A generalized version of the Esscher transform is adopted to select a pricing kernel. The valuation of both the European and American contingent claims is considered. A piecewise linear partial differential-integral equation governing a price of a standard European contingent claim is obtained. For an American contingent claim, a formula decomposing a price of the American claim into the sum of its European counterpart and the early exercise premium is provided. An approximate solution to the early exercise premium based on the quadratic approximation technique is obtained for a particular case where the jump component is absent. Some numerical results for European and American options are presented for the case without jumps.

EO094 Room LSK1033 ADVANCES IN TIME SERIES ANALYSIS Chair: Kaiji Motegi

EO0311: **On a measure of lack of fit in nonlinear cointegrating regression**
Presenter: **Ke Zhu**, University of Hong Kong, Hong Kong
A Portmanteau test is proposed for the adequacy of nonlinear cointegrating regression models. This Portmanteau test is shown to be asymptotically pivotal, and is applicable to a wide class of integrable and non-integrable regression functions with endogenous regressors derived by either short or long memory innovations. Moreover, the applicability scope of this portmanteau test is generalized to include an additive nonlinear cointegrating regression model, whose consistency results are investigated as an independent interest. Finally, the importance of this portmanteau test is demonstrated by simulated and real data.

EO0538: **Estimation of a nonlinear cointegrating regression**
Presenter: **Kosaku Takanashi**, Keio University, Japan
Nonlinear cointegrating regression is studied. Both nonparametric and parametric estimation are considered in multivariate setting beyond bivariate case. The crucial point of the problem is that the Browinan motion with more than three dimension is non-recurrent. Therefore, the limit distribution obtained by an ergodic theorem for d dimensional Brownian motion degenerate into zero. To meet this problem, we introduce an analytical Markov process theory and Kipnis-Varadhan type central limit theorem. We establish asymptotic theory for a nonlinear least squares estimator and the Nadaraya-Watson estimator.

EO0246: **Robust GEL method for linear hypothesis of infinite variance processes**
Presenter: **Fumiya Akashi**, Waseda University, Japan
Testing general linear hypotheses on the coefficient of ARMA models with possibly infinite variance error terms is considered. If the model may have infinite variance, it is well known that the rate of convergence of fundamental statistics (e.g., sample mean) depends on the unknown tail-index of innovation processes, and the limit distribution is not represented in a closed form in general. As a result, it is often difficult in practice to decide cut-off points of confidence intervals or critical values of tests when we apply classical methods such as the maximum likelihood ratio test directly. To overcome the difficulties, the self-weighted generalized empirical likelihood (GEL) method is constructed, and the proposed GEL statistic converges to the chi-square distribution regardless whether the model has infinite variance or not. That is, the proposed test statistic is shown to be robust against the heavy-tails of the model. Therefore, various important tests involving model diagnostics can be carried out with no prior estimation for the unknown quantity of the models such as the tail-index of the innovations. Some simulation experiments also illustrate that the proposed test performs better than the classical self-weighted Wald-type test.

EO0600: **Dynamic mixed frequency synthesis for GDP nowcasting**
Presenter: **Kenichiro McAlinn**, Duke University, United States
A new method is developed for mixed frequency modeling. We call it mixed frequency synthesis (MFS), and it utilizes the newly developed framework of Bayesian predictive synthesis (BPS). MFS synthesizes the information from multiple frequencies in a theoretically coherent Bayesian framework. We demonstrate the efficacy of MFS by a topical macroeconomic exercise of nowcasting GDP using higher frequency data. This study highlights insights into dynamic relationships among multiple frequencies, as well as the potential for improved forecast accuracy at multiple horizons.

EO0355: **Testing for weak form efficiency of stock markets**
Presenter: **Kaiji Motegi**, Kobe University, Japan
Co-authors: Jonathan Hill
White noise tests are performed on daily stock returns over rolling windows. Tests of weak form efficiency in stock returns are performed predominantly under the null hypothesis of independence or a martingale difference property. These properties rule out higher forms of dependence that may exist in stock returns that are otherwise serially uncorrelated. It is therefore of interest to test whether returns are white noise, allowing for a wider range of conditionally heteroskedastic processes, but also for non-martingale difference white noise. Assisted by the dependent wild bootstrap, we use sup-LM, Cramer-von Mises, and max-correlation statistics in order to test the white noise hypothesis. Evidently the dependent wild bootstrap has only been used in a full data sample, hence a key shortcoming has gone unnoticed: in rolling windows, the block structure unintentionally inscribes an artificial periodicity in confidence bands. We eliminate periodicity by randomizing the block size across bootstrap samples and windows. We find that the degree of market efficiency varies across countries and sample periods. Chinese and Japanese markets have a high degree of efficiency so that we cannot reject the white noise hypothesis. The same goes for the U.K. and the U.S., provided trading occurs during non-crisis periods. When U.K. and U.S. markets face greater uncertainty, we tend to observe negative autocorrelations that are large enough to reject the white noise hypothesis.

EO166 Room LSK1032 APPLIED STATISTICAL MODELING Chair: Jingheng Cai

EO0353: **Multilevel heterogeneous factor analysis models with Bayesian covariance lasso prior**
Presenter: **Junhao Pan**, Sun Yat-sen University, China
Multilevel confirmatory factor analysis (CFA) models are widely used in the social sciences to analyze data with heterogeneity. We extend previous work on multilevel CFA model to account for the dependency among observed indicators. A Lasso covariance prior was proposed to model the entire inverse residual covariance matrix of the observed indicators as a sparse positive definite matrix that contains only a few off-diagonal elements bounded away from zero, and Markov Chain Monte Carlo (MCMC) procedures was also developed to perform Bayesian inference. The proposed multilevel heterogeneous CFA model achieves model parsimony and generally fits the data better, while the factor structure (that is, the number of factors and how the observed indicators are loaded on the factors) is kept intact. Both simulated and real data sets were analyzed to evaluate the validity and practical usefulness of the proposed procedure.

EO0373: **Flexible Bayesian quantile regression for semiparametric geoadditive mixed models**
Presenter: **Jiang Xuejun**, Souther University of Science and Technology, China
A geoadditive mixed regression model is built under the flexible Bayesian quantile regression framework. Instead of considering the asymmetric Laplace distribution as residual distribution, the error term is assumed to be distributed as an infinite mixture of Gaussian densities. A stochastic constraint is imposed to ensure inference on the quantile of interest. Based on that, a Bayesian version of semiparametric geoadditive mixed regression model is proposed, where both one-dimensional curve and two-dimensional surface fitting for modeling interactions are developed by using P-splines. In addition, unobserved heterogeneity is incorporated as spatial random effects to account for variation among different regions. The main advantage of our method is that it not only allows for the full span of quantile-restricted error distribution for Bayesian quantile regression with more flexibility and feasibility, but also can describe the large-scale geographic trend along with the local spatial correlation. Finally, a data set of earthquake in Mainland China is used to illustrate the usefulness and effectiveness of the proposed methodology.

EO0366: **Modelling and forecasting online auction price**
Presenter: **Weiwei Liu**, Lanzhou university, China
Online auctions have become increasingly popular in recent years. There is a growing body of research on this topic, whereas modeling online auction price curves constitutes one of the most interesting problems. Most research treats price curves as deterministic functions, which ignores the random effects of external and internal factors. To account for the randomness, a more realistic model using stochastic differential equation is proposed. The online auction price is modeled by a stochastic differential equation in which the deterministic part is equivalent to the second-order differential equation model proposed previously. The model also includes a component representing the measurement errors. Explicit expressions for the likelihood function are also obtained, from which statistical inference can be conducted. Finally, the forecast accuracy of the proposed model is compared with the ODE (ordinary differential equation) approach. Simulation results show that the proposed model performs better at the end of the online auction.

EC0253: **Response surface methodology in plant breeding**
Presenter: **Reka Howard**, University of Nebraska - Lincoln, United States
The Response Surface Methodology (RSM) is introduced as a strategy to find the combination of attribute levels that results in accurate predictions for a given genomic prediction (GP) method, and compare GP methods. We illustrate RSM with a simulated example where the response we optimize is the difference between prediction accuracy using the parametric best linear unbiased prediction (BLUP) and the nonparametric support vector machine (SVM). The greatest impact on the response is due to the genetic architecture of the population and the heritability. When epistasis and heritability are highest, the advantage of using the SVM versus the BLUP is greatest.

EO0544: **On the impact of the portfolio theory with leverage aversion on the performance of a fund**
Presenter: **Yan Liu**, Ocean University of China, China
The innovation of financial derivatives greatly changes the behaviors of the investors from the long selling to short selling. The traditional way to build a portfolio cannot reflect the degree of risk aversion against leverage, which increases the risk faced by investors invisibly. Therefore, an optimization procedure proposed previously which incorporates the mean, variance and leverage is considered. Considering the volatility of the stock market and the demand of handling high dimensional stock yields, we apply single index model to reduce the estimation errors of the sample variance-covariance matrix. A real data study of Chinese stock market shows that a good application of the new method can significantly improve the performance of a fund.

EO016 Room LSKG003 ADVANCES IN NONPARAMETRIC METHODS AND APPLICATIONS Chair: Taeryon Choi

EO0394: **Harmonic Bayesian prediction under alpha-divergence**
Presenter: **Yuzo Maruyama**, The University of Tokyo, Japan
Bayesian shrinkage methods for constructing predictive distributions are investigated. We consider the multivariate normal model with a known covariance matrix and show that the Bayesian predictive density with respect to Stein's harmonic prior dominates the best invariant Bayesian predictive density, when the dimension is greater than three. Alpha-divergence from the true distribution to a predictive distribution is adopted as a loss function.

EO0419: **Adaptive regularized Hotelling's T2 test for high dimensional data**
Presenter: **Jong Soo Lee**, University of Massachusetts Lowell, United States
A test statistic is presented for high dimensional data modified from Hotelling's T^2 test. Although there have been much interests in Hotelling's T^2 test and many modifications have been suggested for high dimensional data, the proposed test combines and extends some of the existing methods to demonstrate its improvements. The theoretical properties of the proposed test will be shown using the random matrix theory and other tools, and numerical properties will be examined and compared via simulation studies.

EO0541: **A spatiotemporal nonparametric Bayesian variable selection model to multi-subject fMRI data**
Presenter: **Kuo-Jung Lee**, National Cheng-Kung University, Taiwan
A joint analytical framework is considered for the analysis of task-related brain activity in multi-subject fMRI experiments. A novel nonparametric Bayesian variable selection approach is proposed to detect the subject- and group-level neural activation induced by a stimulus and to investigate the association or clustering of spatially remote voxels exhibiting similar characteristics in brain activities. In subject-level analysis, we consider a voxel-dependent hemodynamic response function with a shape parameter in the model and use a spatial generalized linear mixed-effects model to capture the spatial dependence among nearby voxels. In group-level, we achieve clustering of the voxels by using a hierarchical Dirichlet process prior on brain activation probability. We explore the performance of the proposed model on simulated data and on a real fMRI data.

EO0433: **Bayesian estimation of distributions of causes of death with verbal autopsy surveys**
Presenter: **Tsuyoshi Kunihama**, Kwansei Gakuin University, Japan

A population distribution of causes of death provides crucial information for dealing with public health issues. However, in many developing countries with limited medical certification and registration of deaths, there is big uncertainty in vital statistics on causes of death in the populations. As a practical way of obtaining population representative data of deaths by cause, verbal autopsy surveys collect information by interviewing family members about the signs, symptoms and medical history of the deceased. A new Bayesian method is developed for estimation of population distributions of causes of death, taking into account characteristics in verbal autopsy data. The proposed approach is based on a multivariate probit model where associations among items in questionnaires are flexibly induced by latent factors. We measure strength of conditional dependence of symptoms with causes in the proposed framework. Using the Population Health Metrics Research Consortium gold standard data, we assess performance of the proposed method and estimate important questionnaire items highly associated with causes of death.

EO0854: **Nonparametric Bayesian multivariate meta-regression with spatial structure: An application in environmental epidemiology**
Presenter: **Yeonseung Chung**, Korea Advanced Institute of Science and Technology, Korea, South
Co-authors: Gyuseok Sim, Antonella Zanobetti, Ho Kim, Joel Schwartz

In biomedical research, meta-analysis has been a useful tool combining evidence from multiple studies to investigate an exposure-response association. Because of the hierarchical nature in meta-analysis, a two-stage analytical approach has been used in many studies for its computational convenience and flexibility. One of the applications is an epidemiological study for the health effect of environmental exposure. The study analyzes the time series data over multiple locations and the multi-site time series design is essentially a form of meta-regression. The currently used multivariate meta-regression is a form of multivariate normal linear regression which assumes linearity in meta-predictors, residual independence, normality and homoscedasticity. However, these assumptions may be limited when study-specific estimates are heterogeneous, spatially-correlated and explained by meta-predictors in a nonlinear manner with non-normal residuals. In this paper, we propose a more flexible two-stage multivariate meta-regression by replacing the second stage by a nonparametric Bayesian multivariate spatial nonlinear regression. The proposed method was compared with the currently used method through a simulation study and applied to the data for the 135 US cities to investigate the temperature-mortality association.

EO298 Room LSK1010 CHANGE POINT ANALYSIS IN A HIGH-DIMENSIONAL SETTING Chair: Minya Xu

EO0515: **Test for temporal homogeneity of high-dimensional means with application to fMRI studies**
Presenter: **Jun Li**, Kent State University, United States
Co-authors: Ping-Shou Zhong

Motivated by the region of interest analysis in the fMRI studies, we consider the problem of testing temporal homogeneity of p-dimensional population mean vectors from the repeated measurements of n subjects over T times. To cope with the challenges brought by fMRI data, we propose a test statistic that takes into account not only the large p, large T and small n situation, but also the complex temporospatial dependence of the fMRI data. The asymptotic distribution of the proposed test statistic is established under mild conditions. When the null hypothesis of temporal homogeneity is rejected, we further propose a binary segmentation method shown to be consistent for multiple change-points identification. An application to fMRI data is provided to demonstrate the performance of the proposed methods.

EO0560: **Multiple changepoint detection via deep learning**
Presenter: **Qiqi Lu**, Virginia Commonwealth University, United States

Change points are extremely important features to consider when homogenizing time series and analyzing its trends and variations. A new change point detection method is introduced using supervised deep learning algorithms. Deep learning is a branch of machine learning and uses a cascade of multiple layers of nonlinear processing units for feature extraction and transformation. The advantage of using deep neural networks is that the method does not need to assume any specific probability model or correlation structure of the time series to be tested. We focus on the appropriate design of deep neural networks for the task of change point detection by specifying the multiple layers and developing the training procedure. The method is applied in the analysis of three time series: an annual precipitation series from New Bedford, MA, a century of monthly temperatures from Tuscaloosa, AL, and the North Atlantic basin tropical cyclone record.

EO0565: **Multiple change-points detection in high dimension**
Presenter: **Yunlong Wang**, Nankai University, China

Change-point detection is an integral component of statistical modeling and estimation. For high-dimensional data, classical methods based on the Mahalanobis distance are typically inapplicable. We propose a novel testing statistic by combining a modified Euclidean distance and an extreme statistic, and its null distribution is asymptotically normal. The new method naturally strikes a balance between the detection abilities for both dense and sparse changes, which gives itself an edge to potentially outperform existing methods. Furthermore, the number of change-points is determined by a new Schwarz's information criterion together with a pre-screening procedure, and the locations of the change-points can be estimated via the dynamic programming algorithm in conjunction with the intrinsic order structure of the objective function. Under some mild conditions, we show that the new method provides consistent estimation with an almost optimal rate. Simulation studies show that the proposed method has satisfactory performance of identifying multiple change-points in terms of power and estimation accuracy, and two real data examples are used for illustration.

EO0474: **Fast change-points detection in high-dimension: A combination of global and local segmentations**
Presenter: **Changliang Zou**, Nankai University, China

A novel procedure is proposed by combining local and global segmentation for change-point detection in high dimension. The new method naturally strikes a balance between the detection accuracy and computation speed, which gives itself an edge to potentially be useful for massive datasets. We show that the new method provides consistent estimation with an almost optimal rate under mild conditions. Simulation studies show that the proposed method has satisfactory performance of identifying multiple change-points in terms of estimation accuracy.

EO0247: **Detecting variance change-points for blocked time series and dependent panel data**
Presenter: **Minya Xu**, Guanghua School of Management, Peking University, China
Co-authors: Ping-Shou Zhong

A class of weighted differences of averages (WDA) statistics is proposed to test and estimate possible change-points in variance for time series with weakly dependent blocks and dependent panel data without specific distributional assumptions. We derive the asymptotic distributions of the test statistics for testing the existence of a single variance change-point under the null and local alternatives. We also study the consistency of the change-point estimator. Within the proposed class of the WDA test statistics, a standardized WDA test is shown to have the best consistency rate and is recommended for practical use. An iterative binary searching procedure is suggested for estimating the locations of possible multiple change-points in variance, whose consistency is also established. Simulation studies are conducted to compare detection power and number of wrong rejections of the proposed procedure to that of a cumulative sum (CUSUM) based test and a likelihood ratio-based test. Finally, we apply the proposed method to a stock index dataset and an unemployment rate dataset. Supplementary materials for this article are available online.

EO084 Room LSKG001 MODELLING WITH NON-GAUSSIAN DISTRIBUTIONS Chair: Geoffrey McLachlan

EO0616: **Looking skew from Antonella Capitanio's perspective**
Presenter: **Cinzia Viroli**, University of Bologna, Italy
Co-authors: Angela Montanari
In recent years, research on skew-normal, skew-t and related flexible families of distributions, that can be constructed by a perturbation of the corresponding symmetric ones, has been really intense and prolific. We retrace the most important contributions to the field by Prof. Antonella Capitanio (1964-2016). Her pioneering works are an inevitable reference for anybody willing to enter this research domain.

EO0582: **Modelling and clustering via finite mixtures of skew factor analyzers**
Presenter: **Sharon Lee**, University of Queensland, Australia
Co-authors: Geoffrey McLachlan
Mixtures of factor analyzers (MFA) provide a popular and powerful tool for simultaneous clustering and dimension reduction. The traditional MFA assumes a joint normal distribution for the latent factors and errors. We consider a robust generalization of the MFA model where the latent factors and errors jointly follow the so-called canonical fundamental skew t-distribution (CFUST). This approach provides an effective way for modelling and clustering high-dimensional heterogeneous data that exhibit non-normal features. The adoption of a CFUST distribution allows the mixtures of CFUST factor analyzers model to accommodate flexible skewness, encompassing a number of commonly used models as special and/or limiting cases. Parameter estimation can be carried out by maximum likelihood via an EM-type algorithm. The usefulness of the proposed methodology is demonstrated using simulated and real datasets.

EO0490: **Finding group structures with a two-way clustering approach**
Presenter: **Shu-Kay Ng**, Griffith University, Australia
Suppose the data are represented by an $n \times p$ matrix of categorical indicators corresponding to n individuals and p observed features for each individual, possibly $n \gg p$. And it is of interest to explore the group structures about the p features as well as the n individuals. A two-way semi-parametric clustering approach to tackle this problem is presented. The first clustering adopts a non-parametric method to analyse (high-dimensional) categorical vectors so as to identify groups of features that are significantly correlated with one another. The clustering results could inform potential association among features in an associative network of features. By controlling for the false discovery rate, the method offers protection against false positive findings of feature groups. The second clustering then adopts a (parametric) mixture model to identify groups of individuals with different patterns in the associative network of features obtained in the first clustering. Key applications of this two-way clustering approach are in the fields of health sciences, where outcomes are measured in a categorical form. The method is illustrated using a national morbidity data set; the group structures of the p morbidities and the n survey participants with respect to their morbidity patterns are explored.

EO0654: **Multithreaded implementation of the EM algorithm in its application to skew normal mixture models**
Presenter: **Kaleb Leemaqz**, University of Queensland, Australia
Co-authors: Sharon Lee, Geoffrey McLachlan
Mixture models provide a powerful tool for the modelling and analysis of heterogeneous data. In recent years, extensions of normal and t mixture models that adopt skew component densities have gained increasing popularity in the statistical literature. These models can be fitted by maximum likelihood via the EM algorithm. However, the implementation of the EM algorithm can become very time-consuming due to the complicated expressions involved. With modern machines typically possessing more than one CPU core, multithreaded variants of the EM algorithm can be easily implemented to harness the power of these processors. A block-based implementation is discussed and illustrated using mixtures of multivariate skew normal distributions. Examples on real datasets are given to demonstrate the performance gain in time using this multithreaded approach.

EO0693: **The multimix class of mixture models for clustering mixed categorical and continuous data**
Presenter: **Lynette Hunt**, University of Waikato, New Zealand
Co-authors: Murray Jorgensen
One possible approach to clustering data is to assume that the data to be clustered come from a finite mixture of populations. The mixture likelihood approach has been well developed and is much used, especially for mixtures where the component distributions are multivariate normal. However, when clustering large multivariate data sets, it is rare to find data with all attributes being either continuous or categorical. The Multimix class of mixture models was proposed in the nineties. This approach enables the clustering of mixed categorical and continuous data, and can also cope with mixed data where data are missing at random in the sense. We demonstrate the Multimix approach to clustering mixed categorical and continuous data, and illustrate the methodology by clustering a large data set.

EO080 Room LSK1014 VARIABLE SELECTION, DIMENSION REDUCTION, AND OUTLIER DETECTION Chair: Sung Nok Chiu

EO0710: **Nonparametric variable selection for additive models**
Presenter: **Zhenghui Feng**, School of Economics & Wang Yanan Institute for studies in Economics, Xiamen University, China
Co-authors: Lu Lin, Ruoqing Zhu, Lixing Zhu
For multivariate nonparametric regression models, existing variable selection methods with penalization require high-dimensional nonparametric approximations in objective functions. When the dimension is high, none of methods with penalization in the literature are readily available. Also, ranking and screening approaches cannot have selection consistency when iterative algorithms cannot be used due to inefficient nonparametric approximation. A novel and easily implemented approach is proposed to make existing methods feasible for selection with no need of nonparametric approximation. Selection consistency can be achieved. As an application to additive regression models, we then suggest a two-stage procedure that separates selection and estimation steps. An adaptive estimation to the smoothness of underlying components can be constructed such that the consistency can be even at parametric rate if the underlying model is really parametric. Simulations are carried out to examine the performance of our method, and a real data example is analyzed for illustration

EO0716: **Robust estimation for high-dimensional Gaussian graphical models**
Presenter: **Kei Hirose**, Kyushu University, Japan
Co-authors: Hironori Fujisawa
Gaussian graphical modeling has been widely used to explore various network structures, such as gene regulatory networks and social networks. We often use a penalized maximum likelihood approach with the L1 penalty for learning a high-dimensional graphical model. However, the penalized maximum likelihood procedure is sensitive to outliers. To overcome this problem, we introduce a robust estimation procedure based on the gamma-divergence. The proposed method has a redescending property, which is known as a desirable property in robust statistics. The parameter estimation procedure is constructed using the Majorize-Minimization algorithm, which guarantees that the objective function monotonically decreases at each iteration.

EO0717: **Multivariate functional data visualization and outlier detection**
Presenter: **Wenlin Dai**, KAUST, Saudi Arabia
Co-authors: Marc Genton
A new graphical tool, the magnitude-shape (MS) plot, is proposed for visualizing both the magnitude and shape outlyingness of multivariate functional data. The proposed tool builds on the recent notion of functional directional outlyingness, which measures the centrality of functional data by simultaneously considering the level and the direction of their deviation from the central region. The MS-plot intuitively presents not only levels but also directions of magnitude outlyingness on the horizontal axis or plane, and demonstrates shape outlyingness on the vertical axis. A dividing curve or surface is provided to separate non-outlying data from the outliers. Both the simulated data and the practical examples confirm that the MS-plot is superior to existing tools for visualizing centrality and detecting outliers for functional data.

EO0714: **Variable selection for mixed measurement level data in dimension reduction methods and its computation**
Presenter: **Yuichi Mori**, Okayama University of Science, Japan
Co-authors: Masahiro Kuroda, Masaya Iizuka
There are several variable selection methods in dimension reduction methods. Among them Tanaka and Mori's modified principal component analysis (M.PCA) is one of useful variable selection methods in the context of PCA, which derives principal components that are computed using only a selected subset but represent all of the variables. Since M.PCA is a method for numerical variables, if categorical variables in the data could be quantified in an appropriate manner, M.PCA can be applied to mixed measurement level data. To do this, we implement optimal scaling algorithm with alternating least squares (ALS) in M.PCA, because the algorithm quantifies all categorical variables to homogeneous numerical ones. Then, we can select a reasonable subset of variables from the data including any measurement level variables. We call this nonlinear M.PCA. As a feature of variable selection in nonlinear M.PCA, there are three types of selection according to where the quantification is implemented in the selection process. Furthermore, for the problem of computational cost in finding a reasonable subset among possible combinations, we apply an acceleration algorithm, which is proposed to accelerate the convergence of the ALS algorithm, to nonlinear M.PCA. The performances are evaluated in some numerical experiments.

EO0719: **Dimension reduction via seeded canonical correlation analysis**
Presenter: **Jae Keun Yoo**, Ewha Womans University, Korea, South
Canonical correlation analysis (CCA) is a classical and popular statistical methodology in multivariate analysis, especially in studying a relation of two sets of variables. The CCA selects pairs of linear combinations of the two sets by maximizing Pearson-correlation. One limit of CCA should the requirement to larger sample sizes than the numbers of variables for each set. In this talk, a seeded canonical correlation approach to overcome the limit is proposed by adopting seeded dimension reduction. Simulation studies support the approach, and real data analysis are presented.

EO054 Room LSK1009 STATISTICAL METHODS FOR BIG DATA INTEGRATION Chair: Hongtu Zhu

EO0795: **Strong independence screening for ultra-high-dimensional survival data**
Presenter: **Xueqin Wang**, Sun Yat-sen University, China
Ranking by marginal utility provides an efficient way to reduce the data from ultra-high dimension to portable size. In order to handle the complex big data in great variability, the statistic that can measure the nonlinear relationship between response and marginal predictor were extensively discussed recently. Comparing to the regression analysis, it is more challenging when the response is the survival time with possible censoring. We propose a novel method to measure the marginal dependency between survival time and predictors. A screening criteria is presented to determine an active set to include important predictors and exclude unimportant predictors. It is shown that the proposed procedure enjoys good statistical properties. Its performance in finite sample size is evaluated via simulations and illustrated by a real data analysis.

EO0789: **Bayesian scalar on image regression with non-ignorable non-response**
Presenter: **Xinyuan Song**, Chinese University of Hong Kong, Hong Kong
Co-authors: Xiangnan Feng, Tengfei Li, Hongtu Zhu
Medical imaging data have been widely used in modern health care, particularly in the prognosis, screening, diagnosis, and treatment of various diseases. We consider a scalar-on-image regression model that uses ultrahigh dimensional imaging data as explanatory covariates. The model is used to investigate important risk factors for the scalar response of interest, which is subject to non-ignorable missingness. We propose the use of an efficient functional principle component analysis method to reduce the dimensions of the imaging observations. Given that non-ignorable non-response distorts the accuracy of statistical inference and generates misleading results, we propose an imaging exponential tilting model for the examination of the potential influence of imaging observations along with scalar variables on the probability of missingness. An instrumental variable, such as a covariate associated with the response but conditionally independent of the probability of missingness, is introduced to facilitate model identifiability. Statistical inference is conducted in a Bayesian framework with Markov chain Monte Carlo algorithms. Simulation studies show that the proposed method exhibits satisfactory finite sample performance. The methodology is applied to a study on the Alzheimer's Disease Neuroimaging Initiative dataset.

EO0835: **An extended Mallows model for ranked data aggregation**
Presenter: **Han Li**, Shenzhen University, China
The rank aggregation problem is studied, which aims to find a consensus ranking by aggregating multiple ranking lists. To tackle the problem probabilistically, we formulate an elaborate ranking model for full as well as partial rankings by generalizing the Mallows model. Our model assumes that the ranked data is generated through a multi-stage ranking process, which is explicitly governed by parameters that measure the overall quality and stability of the process. The new model is quite flexible and has a closed form expression. Under mild conditions, we can derive a few useful theoretical properties of the model. Furthermore, we propose an efficient statistic called rank coefficient to detect highly correlated rankings and a hierarchical ranking model to fit the data. When the ranked data is big, we propose a simplified version of our algorithm. Through extensive simulation studies and real applications, we evaluate the merits of our models and demonstrate that they outperform the state-of-the-art methods in diverse scenarios.

EO0784: **SFM: Surface functional models for functional data in product spaces**
Presenter: **Hongtu Zhu**, University of Texas MD Anderson Cancer Center, United States
In medical imaging analysis and computer vision, there is a growing interest in analyzing various functional data in product spaces. The aim is to introduce a surface functional modeling framework for the analysis of functional data in the product space of two metric spaces. The two metric spaces may have different grid point patterns, including the number and distribution of grid points. We characterize the nine different subtypes of the grid point patterns. We propose to estimate the mean function by using local linear smoothers with a general weighting scheme. We systematically study the asymptotic properties of the local linear smoothers of the mean function. We derive the optimal bandwidth rates for all the nine subtypes. Simulations and real data analysis confirm our theoretical results.

EO0839: **Transformation biometrical genetic models for the analysis of non-normal twin data**
Presenter: **Xiaobo Guo**, Sun Yat-Sen University, China

In classical twin design studies, the ACE or ADE models for estimating heritability primarily uses two methods: structural equation modeling and linear mixed models. However, both methods heavily rely on the assumption that the phenotypic distribution is normal. Clearly, this assumption is unlikely to be satisfied due to the complexity of phenotypes. We propose a transformation biometrical genetic model which is satisfied to any phenotypic distributions. A Bayesian framework is provided to estimate the parameters. Simulation and real data analysis are conducted to conform the utility of the proposed model.

EO206 Room LSK1003 HIGH DIMENSIONAL BAYESIAN TIME SERIES MODELING AND FORECASTING Chair: Yasuhiro Omori

EO0400: **Multivariate stochastic volatility model with realized volatilities and pairwise realized correlations**
Presenter: **Yasuhiro Omori**, University of Tokyo, Japan
Co-authors: Yuta Yamauchi

The multivariate stochastic volatility modes with dynamic pairwise correlations is proposed: (1) we obtain the stable parameter estimates for dynamic correlation models using the realized measures, (2) we make full use of intraday information by using pairwise realized correlations, (3) the covariance matrices are guaranteed to be positive definite, (4) we avoid the arbitrariness of the ordering of asset returns, (5) we propose the flexible correlation structure model (e.g. such as setting some correlations to be identically zeros if necessary), and (6) the parsimonious specification for the leverage effect is proposed. Our proposed models are applied to daily returns of nine U.S. stocks with their realized volatilities and pairwise realized correlations, and are shown to outperform the existing models with regard to portfolio performances.

EO0485: **Nowcasting using news topics: Big data versus big bank**
Presenter: **Leif Anders Thorsrud**, Norges Bank, Norway

The agents in the economy use a plethora of high frequency information, including news media, to guide their actions and thereby shape aggregate economic fluctuations. Traditional nowcasting approaches have to a relatively little degree made use of such information. We show how unstructured textual information in a business newspaper can be decomposed into daily news topics and used to nowcast quarterly GDP growth. Compared with a big bank of experts, here represented by official central bank nowcasts and a state-of-the-art forecast combination system, the proposed methodology performs at times up to 15 percent better, and is especially competitive around important business cycle turning points. Moreover, if the statistical agency producing the GDP statistics itself had used the news-based methodology, it would have resulted in a less noisy revision process. Thus, news reduces noise.

EO0458: **Dynamic mixed frequency pooled copula**
Presenter: **Audrone Virbickaite**, University of Konstanz, Germany
Co-authors: Hedibert Lopes

Modeling dependence between random variables through copula is proposed by exploiting the information available from daily and intradaily data. The two alternative copula models, obtained from low and high frequency data, are combined via density pooling approach and show superior performance in terms of predictive log Bayes Factors. The intraday copula is estimated via three alternative specifications: independent, hierarchical and mixture model. Proposed pooled model is applied to 3-variate and 10-variate financial data, traded at NYSE.

EO0326: **Model selection in the time-variant cointegration model**
Presenter: **Roberto Leon-Gonzalez**, GRIPS, Japan

Calculating predictive likelihoods or marginal likelihoods in Time-Variant Cointegrating models can be very time consuming and challenging in large dimensions. We adapt the Stochastic Search Variable Selection (SSVS) algorithm in order to use the Gibbs sampling algorithm for model selection. However, we find that the SSVS algorithm can suffer from a high correlation between the model indicator and the other parameters. To solve this problem we design an algorithm that proposes a joint update of the model indicator and other model parameters such as the cointegrating space and the adjustment coefficients. We find that this joint update solves the problem and allows the algorithm to switch among models irrespective of initial values. We illustrate the methodology with an analysis of 10 electricity prices from Mexico.

Thursday 15.06.2017 14:00 - 15:40 Parallel Session C – EcoSta2017

EI301 Room LSKG001 NON- AND SEMI-PARAMETRIC INFERENCE Chair: Ping-Shou Zhong

EI0279: **Flexible parametric approach to classical measurement error variance estimation without auxiliary data**
Presenter: **Ingrid Van Keilegom**, KU Leuven, Belgium
Co-authors: Aurelie Bertrand, Catherine Legrand
Measurement error in the continuous covariates of a model generally yields bias in the estimators. It is a frequent problem in practice, and many correction procedures have been developed for different classes of models. However, in most cases, some information about the measurement error distribution is required. When neither validation nor auxiliary data (e.g., replicated measurements) is available, this specification turns out to be tricky. We develop a likelihood-based procedure to estimate the variance of classical additive error of Gaussian distribution, without additional information. The performance of this estimator is investigated both in an asymptotic way and through finite-sample simulations. The usefulness of the obtained estimator when using the Simulation-Extrapolation algorithm, a widely used correction method, is then analyzed in the Cox proportional hazards model through other simulations. Finally, the whole procedure is illustrated on real data.

EI0407: **Fitting a two phase threshold multiplicative error model**
Presenter: **Hira Koul**, Michigan State University, United States
The class of multiplicative error models are particularly suited to model nonnegative time series such as financial durations, realized volatility, and squared returns. Threshold models are also known to play an important role in time series analysis. We shall present a lack-of-fit test for fitting a two-phase threshold model to the conditional mean function in a multiplicative error model. The proposed testing procedure can also be applied to a class of autoregressive conditional heteroscedastic threshold models. A simulation study shows some superiority of the proposed test over some commonly used existing tests. We shall illustrate the testing procedure by some data examples.

EI0609: **Moving block bootstrap procedures for distribution estimation for sample quantiles of strong mixing sequences**
Presenter: **Stephen Lee**, The University of Hong Kong, Hong Kong
Compared to distribution estimation of smooth functionals of strong mixing data, consistency and optimality properties of moving block bootstrap schemes are much less known for distribution estimation of sample quantiles. Existing results focus primarily on either the standard block bootstrap, which requires the bootstrap series be of approximately the same length n as the observed data series, or subsampling, where a bootstrap series is formed by a single block. We show that optimality of the moving block bootstrap is in fact achieved by neither the standard method nor subsampling, but by setting the length of the bootstrap series to be of order $n^{2/3}$ and the block length be of order $n^{1/3}$. Empirical results are presented which compare the performance of the moving block bootstrap under different lengths of bootstrap series.

EO244 Room LSK1033 MACRO AND FINANCIAL ECONOMETRICS Chair: Kyu Ho Kang

EO0153: **A Bayesian foreign currency portfolio optimization of conditional value-at-risk**
Presenter: **Kyu Ho Kang**, Korea university, Korea, South
A Bayesian method is presented for implementing a conditional value-at-risk (C-VaR) portfolio optimization method for foreign currency investments. Our portfolio strategy seeks to minimize the C-VaR with an expected return constraint. This consists of two stages. The first stage is to simulate the posterior predictive densities of the currency returns. To this end, we propose a new multivariate stochastic volatility (MSV) model with time-varying conditional correlations, and provide an efficient MCMC algorithm for the posterior inference. In the second stage, given the currency return density forecasts, we conduct optimal portfolio choice minimizing the C-VaR through a numerical optimization. We subsequently evaluate our portfolio strategy in terms of out-of-sample C-VaR prediction. For an empirical application, we use data on weekly returns for USD/EUR, USD/JPY, andUSD/KRW. According to our out-of-sample portfolio choice experiments, the MSV with fat tail produces the most accurate C-VaR forecasts. In addition, we find that the optimal portfolio weights change over time drastically even when the transaction cost is considered.

EO0239: **Measuring the output gap with professional forecasts: A bivariate approach**
Presenter: **Yunjong Eo**, University of Sydney, Australia
The aim is to measure the output gap based on bivariate unobserved components models with U.S. quarterly real GDP growth and professional forecasters' prediction. The unobserved components (UC) model decomposes real GDP growth into a trend (a random walk with drift) and a stationary component. The model parameters are restricted in that the UC-based one-period-ahead prediction is consistent with that from professional forecasters. This additional information from the forecasts helps identify the model parameters and produce intuitive estimates of the output gap. The bivariate UC decomposition implies a larger cycle-trend variance ratio and more persistent cycles than the univariate decomposition. Trend and cycle innovations are estimated to be less negatively correlated in comparison to the univariate model. As a result, the bivariate estimates of cyclical components are positively associated with NBER reference cycles while the univariate estimates are not. The bivariate UC model allows for structural interpretation of permanent and transitory shocks.Furthermore, the bivariate UC model is found to have a structural break in the mean growth rate of real GDP around 1974, which is not found in the univariate model.

EO0240: **Likelihood inference for dynamic linear models with Markov switching parameters: On the efficiency of the Kim Filter**
Presenter: **Young Min Kim**, Korea University, Korea, South
Co-authors: Kyu Ho Kang
The Kim filter (KF) approximation is commonly used for the likelihood calculation of dynamic linear models with Markov regime switching parameters. Despite its popularity, its approximation error has not been examined in a rigorous way yet. The reliability of the KF approximation is studied. In order to measure the approximation error, we compare the outcomes of the KF method with those of the auxiliary particle filter (APF). The APF is a numerical method requiring a longer computing time, but its error can be sufficiently minimized by increasing simulation size. We conduct extensive simulation and empirical studies for comparison purposes. According to our experiments, the likelihood values obtained from the KF approximation are practically identical to those of the APF. Consequently, the KF approximation is a fast and efficient approach for maximum likelihood estimation of regime-switching dynamic linear models. The main contribution is to justify the use of the Kim filter.

EO0462: **A long-run approach to money, unemployment, and asset prices**
Presenter: **Kuk Mo Jung**, Division of International Studies, Hanyang University, Korea, South
A long-run relationship between money (inflation or interest rates), unemployment, and asset prices are studied. Using panel data, we first document novel evidence that a statistically significant joint relationship between the three variables exists. Specifically, three long-run relationships are found: 1. A positive relationship between inflation (or interest rates) and unemployment; 2. A negative relationship between unemployment and asset prices; 3. A negative relationship between inflation (or interest rates) and asset prices. These findings are robust to a number of different estimation strategies. We provide a unified framework that incorporates money, unemployment, and asset prices in a micro-founded way. The model predicts the empirically found joint relationship in the long run. The calibration exercise also shows that the model can account for a sizable portion of the

long-run joint relationship among those three variables for major OECD countries over the post WW-II period. Lastly, the model suggests some novel and interesting policy implications, especially for open market operations in equities, one of the newly debated topics in the aftermath of global financial crisis.

EO256 Room LSK1003 BUSINESS ANALYTICS Chair: Ray Chung

EO0173: **On the factor structure of high-dimensional asset markets**
Presenter: **CY Sin**, National Tsing Hua University, Taiwan
The study of factor structure on asset markets can be traced back to the seventies. It was shown that given a factor structure, the mean returns of financial assets are approximately linear functions of the factor loadings. Assuming a market with many assets which results in risk-free investment opportunities, it was imposed the no arbitrage constraint and it was concluded that the corresponding d eigenvectors converge and play the role of factor loadings. In view of the recent flourishing literature on high-dimensional models,the so-called "krigings over space and time" factor model is extended in three-fold: (i) The number of assets goes to infinity; (ii) The number of factors d is allowed to go to infinity; and (iii) Macroeconomic and fundamental factors are incorporated. Our model is applied to the stocks listed in the Taiwan Stock Exchange.

EO0451: **On deep machine learning and time series models: A case study with the use of Keras**
Presenter: **Carlin Chu**, The Open University of Hong Kong, Hong Kong
The purpose is to illustrate the ways to model financial problems with the use of a machine learning technique, artificial neural network. Keras, a high-level neural network library written in Python, will be used as the tool to tackle the problems. The emphasis will be on comparing the performance of Long-Short Term Memory (LSTM) recurrent neural network with Bollerslev's time series model for forecasting realized volatility. Other possible usages of Keras in financial area will also be addressed.

EO0697: **Realized asymmetric long memory stochastic volatility models**
Presenter: **Manabu Asai**, Soka University, Japan
A realized long memory asymmetric stochastic volatility model is presented, which captures more flexible asymmetric patterns as compared with several existing models. We suggest a quasi-maximum likelihood estimation, based on its spectral density. We apply the model to the return and realized kernel of three stocks traded on the New York Stock Exchange. Overall, the results of the out-of-sample forecasts show the adequacy of the new asymmetric and long memory volatility model for the period including the global financial crisis.

EO0711: **Bayesian randomized response technique with multiple sensitive attributes & its application to behavioral modeling**
Presenter: **Ray Chung**, Hang Seng Management College, China
Co-authors: Amanda Chu
Randomized response technique (RRT) is a classical and effective method to mitigate the effect from the distortion of responses arising from respondents providing dishonest answers to sensitive questions. Traditional RRT usually focuses on the case of a single sensitive attribute, and discussion of the case of multiple sensitive attributes is limited. To implement RRT with multiple sensitive attributes, we propose a generic Bayesian approach for estimating covariance matrices with incomplete information (resulting from the randomization procedure in the RRT case). The proposed Bayesian approach provides benefits such as (i) accommodating the positive definite condition and other intrinsic parameter constraints in the posterior to improve statistical precision, (ii) incorporating Bayesian shrinkage estimation for covariance matrices despite incomplete information, and (iii) adopting a quasi-likelihood to achieve Bayesian semiparametric inference for enhancing flexibility. Based on a simulation experiment, the Bayesian RRT method performs much better than a method of moments at estimating the covariance structures of randomized responses. Moreover, the Bayesian RRT method is applied to a study on the development of a behavioral model explaining illegal waste dumping behavior in Hong Kong.

EO208 Room LSK1005 HIGH DIMENSIONAL MATRICES AND NETWORKS Chair: Qing Mai

EO0185: **Iterative model selection for biclustered matrix completion**
Presenter: **Eric Chi**, North Carolina State University, United States
Co-authors: Liuyi Hu, Arvind Saibaba, Arvind Rao
Performing matrix completion with side information on row-by-row and column-by-column similarities is considered. We build upon recent proposals for matrix estimation with smoothness constraints with respect to row and column graphs. We present a novel iterative procedure for directly minimizing an information criterion in order to select an appropriate amount row and column smoothing, namely perform model selection. We also discuss how to exploit the special structure of the problem to scale up the estimation and model selection procedure via the Hutchinson estimator.

EO0598: **Community detection in multi-layer networks**
Presenter: **Yuguo Chen**, University of Illinois at Urbana-Champaign, United States
In recent years there has been an increased interest in statistical analysis of data with multiple types of relations among a set of entities. Such multi-relational data can be represented as multi-layer graphs where the set of vertices represents the entities and multiple types of edges represent the different relations among them. For community detection in multi-layer graphs, we consider two random graph models, the multi-layer stochastic blockmodel and a model with a restricted parameter space. We derive consistency results for community assignments of the maximum likelihood estimators in both models. The simulation studies and real data applications confirm the superior performance of the multi-layer approaches in comparison to the baseline procedures.

EO0398: **Neyman-Pearson (NP) classification algorithms and NP receiver operating characteristic (NP-ROC)**
Presenter: **Xin Tong**, University of Southern California, United States
Co-authors: Yang Feng, Jingyi Li
In many binary classification applications, such as disease diagnosis and spam detection, practitioners commonly face the need to limit type I error so that it remains below a desired threshold. To address this need, the Neyman-Pearson (NP) classification paradigm is a natural choice; it minimizes type II error while enforcing an upper bound alpha on the type I error. Although the NP paradigm has a century-long history in hypothesis testing, it has not been well recognized and implemented in statistical classification schemes. Common practices that directly limit the empirical type I error to no more than do not satisfy the type I error control objective because the resulting classifiers are still likely to have type I errors much larger than alpha. As a result, the NP paradigm has not been properly implemented for many classification scenarios in practice. We develop the first umbrella algorithm that implements the NP paradigm for all scoring-type classification methods, including popular methods such as logistic regression, support vector machines and random forests. Powered by this umbrella algorithm, we propose a novel evaluation metric for classification methods under the NP paradigm: NP receiver operating characteristic (NP-ROC) bands, an extension of the popular receiver operating characteristic (ROC) curves. NP-ROC bands will serve as a new effective tool to evaluate, compare and select binary classifiers that aim to control type I error.

EO0689: **An EM algorithm for Bayesian graphical model selection**
Presenter: **Feng Liang**, University of Illinois at Urbana-Champaign, United States
A Bayesian framework for graphical model selection (or in particular, for estimating the precision matrix of a high-dimensional Gaussian vector), in which adaptive shrinkage and sparsity are induced by a mixture of Laplace priors. From the penalized likelihood perspective, the prior choice utilized here also gives rise to a smooth non-convex approximation to the L0 penalty. Optimal error rates for estimation consistency in terms of various matrix norms along with selection consistency for sparsity structure recovery are shown under mild conditions. For fast and efficient computation, an EM algorithm is proposed. The performance of our algorithm has been demonstrated on both synthetic and real data sets.

EO305 Room LSKG007 DATA ANALYTICS AND MACHINE LEARNING METHODS FOR RISK AND INSURANCE Chair: Gareth Peters

EO0245: **The Whittle likelihood estimation for Gegenbauer long memory time series with discrete distribution**
Presenter: **Hongxuan Yan**, University of Sydney, Australia
The aim is to study the application of the Whittle likelihood estimation method on the generalized linear Gegenbauer ARMA(GL-GARMA) type model with discrete distribution. We analyze the properties of spectral density function and periodogram. We compare the different Whittle likelihood estimation methods for long memory model with Gaussian distribution. In addition, we discuss the features of extending the Whittle likelihood estimation from the Gaussian distribution to non-Gaussian distribution. In order to represent the spectral density function, we derive the co-variance of some discrete distributions. Besides, we use Taylor expansion to approximate the log link function. We study the behavior and some properties of products of Gegenbauer polynomial in higher orders, such as the envelope, asymptotic behavior and tail behavior. We also study the trends of products of Gegenbauer polynomial effected by different value of long memory parameter, Gegnbauer parameter and the lag. We analyze as well the cumulative sum of these products of Gegenbauer polynomial terms. We find that all the higher order of product of Gegenbauer terms can be closely approximated by the linear combination of the first order. We apply cubic tensor spline to approximate these linear coefficients. Hence, the approximation of the spectral density function and Whittle likelihood function of model are derived.

EO0303: **Bayesian estimation of Gegenbauer long memory processes with stochastic volatility: Methods and applications**
Presenter: **Andrew Phillip**, The University of Sydney, Australia
Co-authors: Jennifer Chan, Shelton Peiris
A time series model is discussed which has generalized long memory in the mean process with stochastic volatility errors using a new Bayesian posterior simulator that couples advanced posterior maximisation techniques, as well as traditional latent stochastic volatility estimation procedures. Details are provided on the estimation process, data simulation, and out of sample performance measures. We conduct several rigorous simulation studies and verify our results for in and out of sample behaviour. We further compare the goodness of fit of the generalized process to the standard long memory model by considering two empirical studies on the US Consumer Price Index and the US Equity Risk Premium.

EO0648: **Functional coefficient semiparametric GARCH via Bayesian model averaging**
Presenter: **Wilson Ye Chen**, University of Technology Sydney, Australia
Co-authors: Richard Gerlach
As the dynamic structure of the financial markets is subject to dramatic changes, a model capable of providing consistently accurate volatility estimates must not make strong assumptions on how prices change over time. Most volatility models impose a particular parametric functional form that relates an observed price change to a volatility forecast (news impact function). We propose a new class of functional coefficient semiparametric volatility models where the news impact function is allowed to be any smooth function, and study its ability to estimate volatilities compared to the well known parametric proposals, in both a simulation study and an empirical study with real financial data. We estimate the news impact function using a Bayesian model averaging approach, implemented via a carefully developed Markov chain Monte Carlo (MCMC) sampling algorithm. Using simulations we show that our flexible semiparametric model is able to learn the shape of the news impact function from the observed data. When applied to real financial time series, our new model suggests that the news impact functions are significantly different in shapes for different asset types, but are similar for the assets of the same type.

EO0645: **Quantile range-based volatility measure for modelling and forecasting volatility using high frequency data**
Presenter: **Kok-haur Ng**, University of Malaya, Malaysia
Co-authors: Jennifer Chan, Shay-kee Tan
Volatility of asset prices in financial market is not directly observable. Return-based models have been proposed to estimate the volatility using daily close price. Recently, many new range-based volatility measures were proposed to estimate the financial volatility. A quantile Parkinson (QPK) measure is proposed to estimate daily volatility. We show how the Parkinson (PK) measure can robustify in the presence of intraday extreme returns. Results from extensive simulation studies show that the QPK measure is more efficient than intraday (open-to-close) squared returns and PK measures in the presence of intraday extreme returns. To demonstrate the applicability of QPK measure, we analyse the daily Standard and Poor 500 indices by fitting the QPK measure to the conditional autoregressive range (CARR) models. Results shows that choosing a suitable quantile level for the QPK measure will reduce its variance and hence improve its efficiency. In addition, the QPK measure using asymmetric CARR model gives the best in-sample model fit based on Akaike information criterion and provides the best out-of-sample forecast based on root mean squared forecast error and other measures. Mincer Zarnowitz test is carried out to confirm the unbiasedness of the forecasted volatility. Different levels of value-at-risk and conditional value-at-risk forecasts are also provided.

EO152 Room LSK1009 RECENT ADVANCES IN NONPARAMETRIC INFERENCE Chair: Young Kyung Lee

EO0281: **New structural models for in-sample density forecasting**
Presenter: **Young Kyung Lee**, Kangwon National University, Korea, South
Co-authors: Enno Mammen, Jens Perch Nielsen, Byeong Park
Recent proposals of density forecasting models are generalized, and theory is developed for this class of models. In density forecasting, the density of observations is estimated in regions where the density is not observed. Identification of the density in such regions is guaranteed by structural assumptions on the density that allows exact extrapolation. The structural assumption is made that the density is a product of one-dimensional functions. The theory is quite general in assuming the shape of the region where the density is observed. Such models naturally arise when the time point of an observation can be written as the sum of two terms (e.g. onset and incubation period of a disease). The developed theory also allows for a multiplicative factor of seasonal effects. Seasonal effects are present in many actuarial, biostatistical, econometric and statistical studies. Smoothing estimators are proposed that are based on backfitting. Full asymptotic theory is derived for them. A practical example from the insurance business is given producing a within year budget of reported insurance claims. A small sample study supports the theoretical results.

EO0291: **Hypothesis testing in high dimensions**
Presenter: **Ming-Yen Cheng**, National Taiwan University, Taiwan
High dimensionality is one of the key features in big data, and it poses many new challenges in the analysis and inference. Some hypothesis testing problems will be discussed occurring when we make inference based on different types of high-dimensional data. In particular, the focus will be on

analysis of variance with functional data, two-sample testing with multivariate data, and goodness-of-fit testing for some semiparametric regression models. These problems have their own challenging issues, and we will present some feasible approach(es) to each of them. Theoretical and numerical results, and applications to datasets coming from medical and climate studies will be given to demonstrate the efficacy and advantages of the proposed methods. Some future research directions will be discussed.

EO0347: **Statistical inference on party positions from texts: Statistical modeling, bootstrap and adjusting for time effects**
Presenter: **Eun Ryung Lee**, Sungkyunkwan University, Korea, South

A central task in comparative politics is to locate party positions in a certain political space. For this, several empirical methods have been proposed using text as data sources. In general, the analysis of texts to extract information is a difficult task. Its data structure is very complex and political texts usually contain a large number of words such that a simultaneous analysis of word counts becomes challenging. We consider Poisson models for each word count simultaneously and provide a statistical analysis suitable for political text data. In particular, we allow for multi-dimensional party positions and develop a data-driven way of determining the dimension of positions. Additionally, we consider a novel model which allows the political lexicon to change over time and develop an estimation procedure based on LASSO and fused LASSO penalization techniques to address high-dimensionality via significant dimension reduction. Furthermore, to address the potential dependence structure of the word counts over time, we included integer-valued time series processes into our modeling approach and we implemented a suitable bootstrap method to construct confidence intervals for the model parameters. We apply our approach to party manifesto data from German parties over all seven federal elections after German reunification. The data studies confirm that our procedure is robust, runs stable and leads to meaningful and interpretable results

EO0492: **Two-step estimation for varying coefficient regression models with censored data**
Presenter: **Seong Jun Yang**, Hankuk University of Foreign Studies, Korea, South
Co-authors: Ingrid Van Keilegom, Cedric Heuchenne

Estimators of the coefficient functions for the varying coefficient model are proposed where the response is subject to random right censoring. The model includes different coefficient functions depending on various covariates. Since multivariate smoothing is unavoidable under the model, smooth backfitting is applied to avoid “the curse of dimensionality". The estimation method is based on the Bucklely-James type transformation, where the estimators achieved by Koul-Susarla-Van Ryzin type transformation are used for primary estimators of the coefficient functions. Asymptotic normality of the proposed estimators are given, and numerical studies are shown to illustrate the reliability of the estimators.

EO154 Room LSK1034 RECENT ADVANCES IN TIME SERIES ANALYSIS Chair: Wai-Keung Li

EO0297: **Hybrid quantile regression estimation for time series models with conditional heteroscedasticity**
Presenter: **Yao Zheng**, University of Hong Kong, Hong Kong
Co-authors: Qianqian Zhu, Guodong Li, Zhijie Xiao

Estimating conditional quantiles of financial time series is essential for risk management and many other applications in finance. It is well-known that financial time series display conditional heteroscedasticity. Among the large number of conditional heteroscedastic models, the generalized autoregressive conditional heteroscedastic (GARCH) process is the most popular and influential one. So far, feasible quantile regression methods for this task have been confined to a variant of the GARCH model, the linear GARCH model, owing to its tractable conditional quantile structure. The widely used GARCH model is considered. An easy-to-implement hybrid conditional quantile estimation procedure is developed based on a simple albeit nontrivial transformation. Asymptotic properties of the proposed estimator and statistics are derived, which facilitate corresponding inferences. To approximate the asymptotic distribution of the quantile regression estimator, we introduce a mixed bootstrapping procedure, where a time-consuming optimization is replaced by a sample averaging. Moreover, diagnostic tools based on the residual quantile autocorrelation function are constructed to check the adequacy of the fitted conditional quantiles. Simulation experiments are carried out to assess the finite-sample performance of the proposed approach. The usefulness of the proposed estimator and inference tools is further illustrated by an empirical application.

EO0357: **Narrowest-over-threshold detection of multiple change-points and change-point-like features**
Presenter: **Yining Chen**, London School of Economics and Political Science, United Kingdom
Co-authors: Piotr Fryzlewicz, Rafal Baranowski

A new, generic and flexible methodology is proposed for nonparametric function estimation, in which we first estimate the number and locations of any features that may be present in the function, and then estimate the function parametrically between each pair of neighbouring detected features. Examples of features handled by our methodology include change-points in the piecewise-constant signal model, kinks in the piecewise-linear signal model, and other similar irregularities. Our methodology works with only minor modifications across a range of generalised change-point scenarios, and we achieve such a high degree of generality by proposing and using a new multiple generalised change-point detection device, termed Narrowest-Over-Threshold (NOT). The key ingredient of NOT is its focus on the smallest local sections of the data on which the existence of a feature is suspected. Crucially, this adaptive localisation technique prevents NOT from considering subsamples containing two or more features, a key factor that ensures the general applicability of NOT. For selected scenarios, we show the consistency and near-optimality of NOT in detecting the number and locations of generalised change-points. The NOT estimators are easy to implement and rapid to compute: the entire threshold-indexed solution path can be computed in close-to-linear time. Importantly, the NOT approach is easy to extend by users to tailor to their own needs. Our methodology is implemented in the R package not.

EO0371: **On a spiked model for large volatility matrix estimation from noisy high-frequency data**
Presenter: **Keren Shen**, The University of Hong Kong, Hong Kong
Co-authors: Jeff Yao, Wai-Keung Li

Recently, inference about high-dimensional integrated covariance matrices (ICVs) based on noisy high-frequency data has emerged as a challenging problem. In the literature, a pre-averaging estimator (PA-RCov) is proposed to deal with microstructure noise. Using the large-dimensional random matrix theory, it has been established that the eigenvalue distribution of the PA-RCov matrix is intimately linked to that of the ICV through the Marcenko-Pastur equation. Consequently, the spectrum of the ICV can be inferred from that of the PA-RCov. However, extensive data analyses demonstrate that the spectrum of the PA-RCov is spiked, that is, a few large eigenvalues (spikes) stay away from the others which form a rather continuous distribution with a density function (bulk). Therefore, any inference on the ICVs must take into account this spiked structure. As a methodological contribution, we propose a spiked model for the ICVs where spikes can be inferred from those of the available PA-RCov matrices. The consistency of the inference procedure is established and is checked by extensive simulation studies. In addition, we apply our method to some real data from the US and Hong Kong markets. It is found that our model clearly outperforms existing ones, both from the empirical and statistical points of view.

EO0448: **Network GARCH model**
Presenter: **Dong Li**, Tsinghua University, China

The multivariate GARCH (MGARCH) models are popularly used for analyzing financial time series data. However, statistical inference for MGARCH models is quite challenging due to high dimensionality. To circumvent this difficulty, we propose here a network GARCH model, which

can substantially reduce computational complexity. The proposed model makes use of information derived from an appropriately defined network structure. By doing so, the number of unknown parameters in the new model is much reduced. Strict and weak stationarity of the network GARCH model is rigorously established.In order to estimate the model, a quasi-maximum likelihood estimator (QMLE) is developed, and its asymptotics are investigated. Simulation studies are carried out to assess the performance of the QMLEin finite samples and empirical examples are analyzed to illustrate the usefulness of network GARCH models.

EO074 Room LSK1014 ADVANCES IN EXACT AND APPROXIMATE BAYESIAN COMPUTATION Chair: Robert Kohn

EO0327: **Computationally efficient Bayesian estimation of high dimensional copulas with discrete and mixed margins**
Presenter: **Minh-Ngoc Tran**, University of Sydney, Australia
Co-authors: David Gunawan, Kosuke Suzuki, Josef Dick, Robert Kohn
Estimating copulas with discrete marginal distributions is challenging, especially in high dimensions, because computing the likelihood contribution of each observation requires evaluating 2^J terms, with J the number of discrete variables. Currently, data augmentation methods are used to carry out inference for discrete copula and in practice, the computation becomes intractable when J is large. We propose two new fast Bayesian approaches for estimating high dimensional copulas with discrete margins, or a combination of discrete and continuous margins. Both methods are based on recent advances in Bayesian methodology that work with an unbiased estimate of the likelihood rather than the likelihood itself, and our key observation is that we can estimate the likelihood of a discrete copula unbiasedly with much less computation than evaluating the likelihood exactly or with current simulations methods that are based on augmenting the model with latent variables. The first approach builds on the pseudo marginal method. The second approach is based on a variational Bayes approximation to the posterior.

EO0443: **On general sampling schemes for particle Markov chain Monte Carlo methods**
Presenter: **Christopher K Carter**, UNSW, Australia
Co-authors: Eduardo Mendes, Robert Kohn
Particle Markov Chain Monte Carlo methods (PMCMC) are used to carry out inference in non-linear and non-Gaussian state space models, where the posterior density of the states is approximated using particles. Current approaches usually carry out Bayesian inference using a particle Marginal Metropolis-Hastings algorithm (PMMH), a particle Gibbs sampler (PG),or a particle Metropolis within Gibbs sampler (PMwG). We give a general approach for constructing sampling schemes that converge to target distributions given in the literature. Our approach shows how the three ways of generating variables mentioned above can be combined flexibly. The advantage of our general approach is that the sampling scheme can be tailored to obtain good results for different applications. We investigate the properties of the general sampling scheme, including conditions for uniform convergence to the posterior. We illustrate our methods with examples of state space models where one group of parameters can be generated in a straightforward manner in a PG step by conditioning on the states,and a second group of parameters are generated without conditioning on the states because of the high dependence between such parameters and the states.

EO0576: **MCMC algorithms for Bayesian nonparametric estimation of copulas**
Presenter: **Mohamad Khaled**, University of Queensland, Australia
Co-authors: Robert Kohn
The focus is on computational aspects of nonparametric Bayesian estimation of copulas models. New results on mixture approximation of arbitrary copulas are exploited to construct a Bayesian framework for nonparametric copula modeling and estimation. Specifically, exact and approximate MCMC algorithms are derived.

EO0395: **Asymptotic properties of approximate Bayesian computation**
Presenter: **Gael Martin**, Monash University, Australia
Co-authors: David Frazier, Christian Robert, Judith Rousseau
Approximate Bayesian computation (ABC) is becoming an accepted tool for statistical analysis in models with intractable likelihoods. With the initial focus being primarily on the practical import of ABC, exploration of its formal statistical properties has begun to attract more attention. We consider the asymptotic behaviour of the posterior obtained from ABC and the ensuing posterior mean. We give general results on: (i) the rate of concentration of the ABC posterior on sets containing the true parameter (vector); (ii) the limiting shape of the posterior; and (iii) the asymptotic distribution of the ABC posterior mean. These results hold under given rates for the tolerance used within ABC, mild regularity conditions on the summary statistics, and a condition linked to identification of the true parameters. Important implications of the theoretical results for practitioners of ABC are highlighted.

EO178 Room LSK1001 FACTOR MODELS AND FINANCIAL ECONOMETRICS Chair: Mengmeng Ao

EO0359: **Recovering mean-variance efficiency under factor models**
Presenter: **Mengmeng Ao**, Xiamen University, China
Co-authors: Yingying Li, Xinghua Zheng
Mean-variance efficiency is investigated under the approximate factor model framework. We propose a novel approach to estimate the mean-variance optimal portfolio on the basis of modeling asset returns by approximate factor structure when both factors and factor loadings can be unknown. We prove that as the number of assets gets large, our estimated portfolio asymptotically achieves maximum expected return and in the meanwhile with a risk controlled by the given constraint. Strong evidences from simulation and extensive empirical studies further support the superior properties of our portfolio.

EO0675: **Prediction under check loss in Gaussian models with unknown covariance**
Presenter: **Gourab Mukherjee**, University of Southern California, United States
A host of modern applications require prediction under asymmetric loss functions. We develop new Empirical Bayes methods that can produce optimal prediction under asymmetric check losses. The check loss function is piecewise linear and penalizes underestimation and overestimation in different ways. Because of the nature of this loss, our inferential target is a pre-chosen quantile of the predictive distribution rather than the mean of the predictive distribution. Prediction here differs in fundamental respects from estimation or prediction under symmetric quadratic loss which is considered in most high-dimensional statistics literature. Under unknown covariance structure, we develop a new method for constructing efficient asymptotic risk estimates for conditionally linear estimators. Minimizing these risk estimates we obtain asymptotically optimal Empirical Bayes prediction rule.

EO0346: **Testing equality of principle components in factor models**
Presenter: **Ningning Xia**, Shanghai University of Finance and Economics, China
Co-authors: Jianqing Fan, Yingying Li, Xinghua Zheng
A test is developed for structural breaks of factor loadings in factor models. We focus on the null hypothesis that factor loadings are constant over time. Because the number of factor loading parameters goes to infinity as the sample size grows, conventional tests cannot be used. Based

on the fact that the presence of a structural change in factor landings yields a direction change of principle components obtained from the sample covariance matrix, we test the equality of sample principle components and compare the pre- and post break subsample covariance matrix. Our test is consistent under the alternative hypothesis in which a fraction of or all factor loadings have structural changes. The Monte Carlo results show that our test has good finite-sample size and power.

EO0643: **Taming the factor zoo**
Presenter: **Dacheng Xiu**, University of Chicago, United States
Co-authors: Stefano Giglio, Guanhao Feng
The asset pricing literature has produced hundreds of potential risk factors. Organizing this "zoo of factors" and distinguishing between useful, useless, and redundant factors require econometric techniques that can deal with the curse of dimensionality. We propose a model-selection method that allows us to systematically evaluate the contribution to asset pricing of any new factor, above and beyond what is explained by a high-dimensional set of existing factors. Our procedure selects the best parsimonious model out of the large set of existing factors, and uses it as the control in evaluating the contribution of new factors. Our statistical inference allows for model selection mistakes, and is therefore robust to model misspecification. We show that despite the existence of hundreds of factors proposed in the last 30 years, several new factors – introduced in the last 5 years – have significant additional explanatory power over existing ones. We find that using a parsimonious model with a constant number of factors yields increasing explanatory power for the cross-sectional variation of expected returns, suggesting that the asset pricing literature has been adding more powerful factors over time, and not just spurious or redundant ones.

EO024 Room LSK1007 ADVANCES IN CHANGE POINTS, MISSING DATA AND NEURAL NETWORKS Chair: Frederick Kin Hing Phoa

EO0384: **An efficient analysis of change points via swarm intelligence**
Presenter: **Frederick Kin Hing Phoa**, Academia Sinica, Taiwan
Co-authors: Hsin-Hao Chen, Livia Lin-Hsuan Chang
Evolutionary algorithm is a new and promising method to statistical optimization. Recently, a nature-inspired metaheuristic method was proposed, namely the Swarm Intelligence Based (SIB) method, for efficient optimization in discrete and continuous domains, but it is of restricted use due to fixed particle size. We introduce a new operation called VARY to the standard SIB framework. It allows the adjustment on the number of change points to be included during the optimization process. We apply the enhanced algorithm for analyzing composite functions that consist of multiple change points. Numerical results show that our algorithm accurately detect the location of change points by using a small number of particles and iterations. Comparing to existing methods in the literature, our algorithm outperforms them in terms of accurate number and location of change points. Our method is demonstrated on the analysis of the data of global surface temperature, indicating that 1906, 1945 and 1963 are the three change-point years for the global surface temperature.

EO0668: **Semiparametric model for regression analysis with missing covariates**
Presenter: **Yang Zhao**, University of Regina, Canada
Maximum likelihood methods for regression model with missing values in covariates require modelling of the covariate distribution, which often needs certain variables to be discrete. We propose a semiparametric model for regression analysis with missing covariates where a piece-wise non-paprametric model will be proposed for the covariate distribution. The method can deal with both continuous and discrete variables. The resulting estimation method is easy to implement and the asymptotic properties of the maximum likelihood estimation follow under certain conditions. The proposed semiparametric approach is robust and can be extended to deal with arbitrary missing data patterns. Extensive simulation studies for different models indicate that the proposed method is acceptable for practical implementation. A real data example is used to illustrate the method.

EO0497: **Minimum contamination and beta-aberration criteria for screening quantitative factors**
Presenter: **Chang-Yun Lin**, National Chung Hsing University, Taiwan
For quantitative factors, the minimum beta-aberration criterion is commonly used for examining the geometric isomorphism and searching for optimal designs. We investigate the connection between the minimum beta-aberration criterion and the minimum contamination criterion. Results reveal that ranking designs by the two criteria can be extremely inconsistent and hence the optimal designs selected by them are likely to be different. We provide statistical justifications showing that the minimum contamination criterion well controls the expected total mean square error of the estimation and demonstrate that it is more powerful than the minimum-aberration criterion on identifying geometrically non-isomophic designs.

EO144 Room LSK1010 HIGH DIMENSIONAL INFERENCE FOR COMPLEX DATA Chair: Lilun Du

EO0396: **Smoothed full-scale approximation of Gaussian process models for computations of large spatial datasets**
Presenter: **Bohai Zhang**, University of Wollongong, Australia
Co-authors: Huiyan Sang, Jianhua Huang
Gaussian process (GP) models encounter computational difficulties with large spatial datasets since its computational complexity grows cubically with sample size n. Although the Full-Scale Approximation (FSA) using a block modulating function provides an effective way for approximating GP models, it has several shortcomings such as the less smooth prediction surface on block boundaries and sensitiveness to the knot set under small-scale data dependence. To address these issues, we propose a Smoothed Full-Scale Approximation (SFSA) method for the analysis of large spatial dataset. The SFSA leads to a class of scalable GP models, whose covariance functions consist of two parts: A reduced-rank covariance function capturing large-scale spatial dependence and a covariance adjusting local covariance approximation errors of the reduced-rank part both within blocks and between neighboring blocks. The proposed method provides a unified view of approximation methods for GP models, encompassing several existing computational methods for large spatial datasets into one common framework. We illustrate the effectiveness of the SFSA approach through simulation studies and a precipitation dataset.

EO0434: **Screening-based Bregman divergence estimation with NP-dimensionality**
Presenter: **Xiao Guo**, University of Science and Technology of China, China
Co-authors: Chunming Zhang, Yi Chai
Feature screening via the marginal screening has gained special attention for high dimensional regression problems. However, their results are confined to the generalized linear model (GLM) with the exponential family of distributions. This inspires us to explore the suitability of applying screening procedures to more general models, for example without assuming either the explicit form of distributions or parametric forms between response and covariates. We extend the marginal screening procedure, by means of Bregman divergence (BD) as the loss function, to include not only the GLM but also the quasi-likelihood model. A sure screening property for the resulting screening procedure is established under this very general framework, assuming only certain moment conditions and tail properties, where the dimensionality p_n is allowed to grow with the sample size n as fast as $\log(p_n) = O(n^a)$ for some $a \in (0,1)$. Simulation and real data studies illustrate that a two-step procedure, which combines the

feature screening in the first step and a penalized-BD estimation in the second step, is practically applicable to identify the set of relevant variables and achieve good estimation of model parameters, with the computational cost much less than those without using the screening step.

EO0477: **Parsimonious model averaging for high-dimensional data**
Presenter: **Xinyu Zhang**, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China
Model averaging generally provides better prediction than model selection, but the existing model averaging methods cannot lead to parsimonious models. Parsimony is an especially important property when modeling high-dimensional data. Studying model averaging for high-dimensional data, we suggest a criterion for choosing weights. The resulting model averaging estimators of coefficients have many zeros and thus leads to a parsimonious model. The asymptotic distribution of the estimators is also provided. Furthermore, the proposed procedure is asymptotically optimal in the sense that its squared loss and risk are asymptotically identical to those of the best but infeasible model averaging estimator. Numerical analysis in comparison with the existing model averaging and selection methods strongly favors our new model averaging procedure.

EO0664: **Weighted false discovery rate control in large-scale multiple testing**
Presenter: **Wenguang Sun**, University of Southern California, United States
The use of weights provides an effective strategy to incorporate prior domain knowledge in large-scale inference. Weighted multiple testing in a decision-theoretic framework will be discussed. We develop oracle and data-driven procedures that aim to maximize the expected number of true positives subject to a constraint on the weighted false discovery rate. The asymptotic validity and optimality of the proposed methods are established. The results demonstrate that incorporating informative domain knowledge enhances the interpretability of results and precision of inference. Numerical results are presented using both simulated and real data, which shows that the proposed method controls the error rate at the nominal level, and the gain in power over existing methods is substantial in many settings.

EO228 Room LSKG003 ASYMPTOTIC STATISTICS OF RANDOM PROCESSES Chair: Masayuki Uchida

EO0557: **Doubly robust estimator for net survival rate in analyses of cancer registry data**
Presenter: **Satoshi Hattori**, Kurume University, Japan
Cancer population studies based on cancer registry databases are widely conducted to address various research questions. In general, cancer registry databases do not collect information on cause of death. The net survival rate is defined as the survival rate if a subject would not die for any causes other than cancer. This counterfactual concept is widely used for the analyses of cancer registry data. Almost all the statistical methods for cancer registry data assume independent censoring. However, in practice of cancer registry analyses, covariate-dependent censoring frequently arises and existing methods allowing the covariate-dependent censoring rely on correct regression modelling of censoring time or that of the net survival. We propose a new estimator for survival time, which is shown to be doubly robust in the sense that it is consistent at least one of the regression models for survival time and for censoring time. We examine the theoretical and empirical properties of our proposed estimator by asymptotic theory and simulation studies. We also apply the proposed method to cancer registry data for gastric cancer patients in Osaka, Japan.

EO0534: **Ergodicity of some reversible proposal MCMC and its application to Bayesian inference for stochastic processes**
Presenter: **Kengo Kamatani**, Osaka University, Japan
Ergodicity of some Markov chain Monte Carlo (MCMC) methods with reversible proposal kernel is studied. This class of MCMC is useful for Bayesian inference problems since we can adjust MCMC algorithm directly to the invariant probability measure of the proposal kernel. Ergodicity theorem provides some useful information to choose appropriate MCMC for Bayesian inference problem. We illustrate the effect by the simulation study for discretely observed stochastic processes.

EO0535: **Stable quasi-likelihood regression**
Presenter: **Hiroki Masuda**, Kyushu University, Japan
Some parametric estimation results are presented for a regression model that can cover not only stochastic differential equations but also some semimartingale regression models. We introduce a class of non-Gaussian quasi-likelihood estimators, and prove its asymptotic mixed normality under the locally stable property of the driving noise process. In sharp contrast to the Gaussian quasi-likelihood based counterpart, the result holds without any moment condition and any stability even when attempting to estimate the trend parameter. Some numerical experiments are shown to illustrate effectiveness of the proposed methodology.

EO0558: **Quasi likelihood analysis and model selection for stochastic processes**
Presenter: **Nakahiro Yoshida**, University of Tokyo, Japan
The quasi likelihood analysis (QLA) gives a basis of statistical inference for stochastic processes. The polynomial type large deviation (PLD) inequality featuring in the QLA provides estimates of the tail of the quasi likelihood random field and hence L_p-boundedness of the quasi maximum likelihood estimator and the quasi Bayesian estimator. The QLA is one of the mathematical fundamentals in the theory of information criteria for model selection. The PLD inequality follows from only the local asymptotic quadratic structure of the quasi log likelihood function. The QLA can extend to the penalized QLA in an abstract manner, keeping PLD. Consequently, for example, L_p-boundedness of the penalized estimator and a precise estimate of the probability of selection consistency are obtained. The penalized QLA so obtained can be applied with high universality to various dependent structures. We will discuss applications of the QLA to information criteria and penalized methods for estimation of point processes and diffusion type processes.

EO176 Room LSK1032 LARGE SCALE FINANCIAL DATA Chair: Yingying Li

EO0849: **High dimensional minimum variance portfolio under factor models**
Presenter: **Yi Ding**, The Hong Kong University of Science and Technology, Hong Kong
Co-authors: Yingying Li, Xinghua Zheng
The high dimensional minimum variance portfolio (MVP) problem is studied under approximate factor models. We extend the theoretical results of POET covariance estimator and propose a MVP estimator that bears risk close to the theoretical minimum risk with the ratio converging to one. The convergence properties of the MVP estimator are established under two asymptotic risk scenarios: (i) when the minimum risk converges to zero with increasing number of assets; and (ii) when the minimum risk is bounded away from zero. Simulation and extensive empirical studies on S&P 100 stock returns demonstrate that our MVP estimator controls risk effectively.

EO0611: **Model-free knockoffs for high-dimensional variable selection**
Presenter: **Yingying Fan**, University of Southern California, United States
Many contemporary large-scale applications involve building interpretable models linking a large set of potential covariates to a response in a nonlinear fashion. Although this modeling problem has been extensively studied, it remains unclear how to effectively control the fraction of false discoveries in general high-dimensional nonlinear models. To address such a practical problem, we propose a new framework of model-free knockoffs, which reads from a different perspective the knockoff procedure originally designed for controlling the false discovery rate in linear models. The key innovation of our method is to construct knockoff variables probabilistically instead of geometrically. This enables model-free

knockoffs to deal with arbitrary (and unknown) conditional models and any dimensions, including when the dimensionality p exceeds the sample size n, while the original knockoffs procedure is constrained to homoscedastic linear models with $n > p$. Our approach requires the design matrix be random(independent and identically distributed rows) with a covariate distribution that is known, although we show our procedure to be robust to unknown/estimated distributions. To our knowledge, no other procedure solves the controlled variable selection problem in such generality, but in the restricted settings where competitors exist, we demonstrate the superior power of knockoffs through simulations and real data analysis.

EO0509: **Estimation of integrated covariance matrix for intra-day trading data**
Presenter: **Clifford Lam**, London School of Economics and Political Science, United Kingdom
While the use of intra-day price data increases the sample size substantially for asset allocation, the usual realized covariance matrix still suffers from bias contributed from the extreme eigenvalues when the number of assets is large. We introduce a novel nonlinear shrinkage estimator for the integrated volatility matrix which shrinks the extreme eigenvalues of a realized covariance matrix back to acceptable level, and enjoys a certain asymptotic efficiency at the same time, all at a high dimensional setting where the number of assets can have the same order as the number of data points. Compared to various other methods suitable for the high dimensional setting, our estimator demonstrates favorable performances in both simulations and a real data analysis in portfolio allocation. This include a novel maximum exposure bound and an actual risk bound when our estimator is used in constructing the minimum variance portfolio.

EO0495: **Tests for high-dimensional covariance matrices when heteroskedasticity is present**
Presenter: **Xinxin Yang**, HKUST, China
Co-authors: Xinghua Zheng, Jiaqi Chen, Hua Li
Tests for high-dimensional covariance matrices are studied when data exhibit heteroskedasticity. The observations are modeled as $\mathbf{Y}_i = \omega_i \mathbf{Z}_i$, where $\mathbf{Z}_i$'s are i.i.d. $p-$dimensional random vectors with mean $\mathbf{0}$ and covariance Σ, and ω_i's represent random scalars reflecting the heteroscedasticity. We aim to test $H_0 : \Sigma \propto \Sigma_0$, in the high-dimensional setting where both p and n grow to infinity proportionally. To remove the heteroskedasticity we self-normalize the observations and establish a CLT for the linear spectral statistics (LSS) of $\widetilde{\mathbf{S}}_n := \frac{p}{n}\sum_{i=1}^n \mathbf{Y}_i\mathbf{Y}_i^T/|\mathbf{Y}_i|^2 = \frac{p}{n}\sum_{i=1}^n \mathbf{Z}_i\mathbf{Z}_i^T/|\mathbf{Z}_i|^2$. Tests based on the CLT neither assume a specific parametric distribution for the data nor involve extra terms containing the forth moments of $\mathbf{Z}_i$'s. Numerical studies show that our tests work well even when $\mathbf{Z}_i$'s are heavy-tailed.

EO246 Room LSK1011 RECENT DEVELOPMENTS IN SUFFICIENT DIMENSION REDUCTION AND GRAPHICAL MODELS Chair: Bing Li

EO0321: **A statistical framework for data integration of heterogeneous and correlated data through graphical models**
Presenter: **Yuping Zhang**, University of Connecticut, United States
Various types of high-dimensional correlated data are emerging. We present a coherent statistical framework for integrating various types of data from distinct but related conditions through graphical models. Our statistical framework is designed for modeling multiple networks with shared interactions from heterogeneous high-dimensional datasets. The performance of our approach is illustrated through simulations and real applications.

EO0606: **Estimation and inference for brain connectivity analysis**
Presenter: **Lexin Li**, University of California Berkeley, United States
Brain connectivity analysis is now at the foreground of neuroscience research. A connectivity network is often characterized by a graph, where nodes represent neural elements such as neurons and brain regions, and links represent statistical dependency and interactions among those neural elements. Such a graph is commonly derived from neuroimaging data such as electroencephalography and functional magnetic resonance imaging. We discuss a number of projects addressing brain connectivity network analysis, including estimation of multiple networks across groups, hypothesis testing of inferring and comparing networks, and association modeling of networks and other biological phenotypes.

EO0691: **Structured sufficient dimension reduction and its applications**
Presenter: **Francesca Chiaromonte**, The Pennsylvania State University, United States
Co-authors: Yang Liu, Bing Li
Structured Ordinary Least Squares (sOLS) is described. It is a method for Sufficient Dimension Reduction in regression problems where both the predictors and the statistical units present a group structure. We also introduce some recent generalization of sOLS to settings in which the response is binary, or the data is collected spatially creating the need to correct for correlated observations. Finally, we illustrate the use of sOLS and its variants with applications to genomics data, where we investigate features affecting the prevalence of non-coding functional elements and de novo mutations in the human genome, as well as Medicare Provider Utilization and Payment data, where we investigate the determinants of per-capita health care costs for the elderly population in the US.

EO0830: **Model-free variable selection with matrix-valued predictors**
Presenter: **Yuexiao Dong**, Temple University, United States
Co-authors: Zeda Li
A novel framework is introduced for model-free variable selection with matrix-valued predictors. To test the importance of rows, columns, and submatrices of the predictor matrix in terms of predicting the response, three types of hypotheses are formulated under a unified framework. An asymptotic test as well as a simple permutation test procedure are used to approximate the null distribution of the test statistics for all three tests. A nonparametric maximum ratio criteria (MRC) is proposed for the purpose of model-free variable selection. Unlike the traditional stepwise regression procedures that require calculating p-values at each step, MRC is a non-iterative procedure that does not require p-value calculation and is guaranteed to achieve variable selection consistency under mild conditions. The effectiveness of the proposed methods are evaluated through extensive numerical studies and an application to the electroencephalography (EEG) dataset.

Thursday 15.06.2017 16:10 - 17:25 Parallel Session D – EcoSta2017

EO212 Room LSKG007 MODELLING AND ESTIMATION IN FINANCIAL TIME SERIES Chair: Clifford Lam

EO0199: **Spatial lag model with time-lagged effects and spatial weight matrix estimation**
Presenter: **Cheng Qian**, London School of Economics and Political Science, United Kingdom
Co-authors: Clifford Lam
A spatial lag model is considered with different spatial weight matrices for different time-lagged spatial effects, while allowing both the sample size T and the panel dimension N to grow to infinity together. To overcome potential misspecifications of these spatial weight matrices, we estimate each one by a linear combination of a set of M specified spatial weight matrices, with M being finite. Moveover, by penalizing on the coefficients of these linear combinations, oracle properties for these penalized coefficient estimators are proved, including their asymptotic normality and sign consistency. Other parameters of the model are estimated by profile-least square type of estimators after introducing covariates which serve similar functions as instrumental variables. Asymptotic normality for our estimators are developed under a previous framework of functional dependence used which is a measure of time series dependence. The proposed methods are illustrated using both simulated and real financial data.

EO0378: **A nonparametric eigenvalue-regularized integrated covariance matrix estimator for asset return data**
Presenter: **Phoenix Feng**, London School of Economics, United Kingdom
Co-authors: Clifford Lam
The estimation of an integrated covariance matrix is important in finance for portfolio allocation. To ameliorate the bias contributed from the extreme eigenvalues of the realized covariance matrix when the dimension p of the matrix is large relative to the average daily sample size n, and the contamination by microstructure noise, various researchers attempted regularization with specific assumptions on the true matrix itself, like sparsity or factor structure, which can be restrictive at times. With non-synchronous trading and contamination of microstructure noise, we propose a nonparametrically eigenvalue-regularized integrated covariance matrix estimator (NERIVE) which does not assume specific structures for the underlying integrated covariance matrix. We show that NERIVE is almost surely positive definite, with extreme eigenvalues shrunk nonlinearly under the high dimensional framework $p/n \to c > 0$. We also prove that almost surely, the minimum variance optimal weight vector constructed using NERIVE has maximum exposure and actual risk upper bounds of order $p^{-1/2}$. Incidentally, the same maximum exposure bound is also satisfied by the theoretical minimum variance portfolio weights. All these results hold true also under a jump diffusion model for the log-price processes with jumps removed using the wavelet method.

EO0590: **Functional linear model with dependent regressors in high dimensions**
Presenter: **Xinghao Qiao**, London School of Economics, United Kingdom
Co-authors: Shaojun Guo, Cheng Chen
The functional linear model is one of the most widely used tools of functional data analysis. Existing approaches assume either independent and identically distributed functional regressors or a fixed number of dependent functional regressors. We propose a new class of partially functional linear models to characterize the linear relationship between a scalar response and high dimensional regressors, involving both dependent scalar and functional variables. We develop a penalized least squares approach to simultaneously perform variable selection for dependent scalar and functional regressors. We investigate the theoretical properties of our proposed method under mild conditions and illustrate the sample performance through an extensive set of simulation studies and one real world example.

EO238 Room LSKG001 EXTREME VALUE MODELING AND RISK ANALYSIS Chair: Emily Kang

EO0216: **Optimal fingerprinting in detection and attribution of changes in climate extremes**
Presenter: **Zhuo Wang**, Shenzhen University, China
Co-authors: Jun Yan, Xuebin Zhang
No fully satisfactory analogue of optimal fingerprinting is available for detection and attribution analysis of changes in climate extremes. The latest development incorporates the signals under certain external forcing into the locations of generalized extreme value (GEV) distributions, where the regression coefficients are the primary interest of inferences. We propose two marginal methods and one joint modeling method. The combined score equations (CSE) method combines the scores of the marginal GEV models through a weight matrix to achieve high efficiency. Under working independence, it reduces to independence likelihood (IL), which turned out to be robust to uncertainty in signal estimation. The third method assumes a max-stable process for the dependence structure and estimates the parameters with the pairwise likelihood (PL). The properties of the estimators were compared in simulation studies. When the spatial dependence is strong and there is little uncertainty in signals, the CSE method provides a close analogue to optimal fingerprinting. The CSE and IL methods were applied to detection and attribution analyses off our annual daily temperature extremes in East Asia, and anthropogenic impact was found in changes in all observed temperature extremes.

EO0402: **Similarity-based clustering for stock market extremes**
Presenter: **Raphael Huser**, King Abdullah University of Science and Technology, Saudi Arabia
Co-authors: Rodrigo Rubio, Miguel de Carvalho
The analysis of the magnitude and dynamics of extreme losses in a stock market is essential from an investors viewpoint. We develop methods of similarity-based clustering for statistics of heteroscedastic extremes, which allow us to group stocks according to their extreme-value index and scedasis function, respectively controlling the magnitude and dynamics of extreme losses. Clustering is here performed in a product-space and a tuning parameter is used to control if more emphasis should be put on the latter or the former. This provides a practical tool to get more insight into stock market extreme synchronization, and can thus be practically useful to investors to make buy/sell decisions. Our analysis reveals an interesting mismatch between the magnitude/dynamics of extreme losses on the London stock exchange and the corresponding economic sectors of stocks.

EO0679: **The impact of competition on prices with numerous firms**
Presenter: **Deyuan Li**, Fudan University, China
Many existing homogeneous-good markets exhibit large mark-ups, including some markets with many competing firms. We analyze simple monopolistic-competition models and show that idiosyncratic demand shocks driven by standard noise distributions can produce large equilibrium markups that are insensitive to the degree of competition. For example, with Gaussian noise and n firms, markups are proportional asymptotically to $1/\sqrt{\ln n}$; consequently, a hundred-fold increase in n, from 10 to 1000 competing firms, only halves the equilibrium markup. The elasticity of the markup with respect to n asymptotically equals the tail exponent from extreme value theory. Only noise distributions with extremely thin tails have negative asymptotic markup elasticities.

EO224 Room LSK1009 MODERN STATISTICAL METHODS FOR COMPLEX DATA Chair: Junhui Wang

EO0223: **Sparse tensor response regression with applications in neuroimaging analysis**
Presenter: **Will Wei Sun**, University of Miami School of Business Administration, United States
Co-authors: Lexin Li
In neuroimaging analysis, it is of keen interest to compare the MRI scans of brains or brain connectivity patterns between the subjects with neurological disorder and the healthy controls. Motivated from these applications, we propose a new regression model with a tensor response and a vector predictor. The model embeds two key low dimensional structures: sparsity and low rankness. It can handle both a general and a symmetric tensor response, and thus it is applicable to both structural and functional neuroimaging data. We formulate the model parameter estimation as a non-convex optimization problem, and develop an efficient alternating updating algorithm. We establish a non-asymptotic estimation error bound for the actual estimator obtained from the proposed algorithm. This error bound reveals an interesting interaction between the computational efficiency and the statistical rate of convergence. Based on this general error bound, we further obtain an optimal estimation error rate when the distribution of the error tensor is Gaussian. We illustrate the efficacy of the new model through intensive simulations and an analysis an Autism spectrum disorder neuroimaging data.

EO0608: **Network cross-validation by edge sampling**
Presenter: **Cheng Qian**, University of Michigan, Ann Arbor, United States
Co-authors: Tianxi Li, Liza Levina, Ji Zhu
Many models and methods are now available for network analysis, but model selection and tuning remain challenging. Cross-validation is a useful general tool for these tasks in many settings, but is not directly applicable to networks since splitting network nodes into groups requires deleting edges and destroys some of the network structure. We propose a new network cross-validation strategy based on splitting edges rather than nodes, which avoids losing information and is applicable to a wide range of network problems. We provide a theoretical justification for our method in a general setting, and in particular show that the method has good asymptotic properties under the stochastic block model. Numerical results on both simulated and real networks show that our approach performs well for a number of model selection and parameter tuning tasks.

EO0858: **An efficient framework for model-free variable selection**
Presenter: **Xin He**, City University of Hong Kong, Hong Kong
Co-authors: Junhui Wang
Variable selection plays a crucial role in the analysis of high dimensional datasets. While dealing with an ultra-high dimensional dataset, most existing methods require different model assumptions. We propose a model-free variable selection method which can exactly identify informative variables that are related to the conditional mean function by measuring the corresponding gradients. More importantly, this method allows the number of predictors grows exponentially with the sample size, which is opposed to most existing model-free methods which can only handle finite number of predictors. Some applications of the proposed method under various model assumptions are also considered in this paper. The proposed method is implemented via an efficient and fast computing algorithm. The asymptotic estimation and variable selection consistencies are established in the model-free framework, which assures that the truly informative variables are correctly identified with high probability. The effectiveness of the proposed method is also supported by a variety of simulated and real-life examples.

EO020 Room LSK1010 KSS SESSION: STATISTICAL LEARNING Chair: Yongdai Kim

EO0266: **Quantile-slicing estimation for dimension reduction in regression**
Presenter: **Seung Jun Shin**, Korea University, Korea, South
Co-authors: Yichao Wu
In high-dimensional data analysis, it is often of primary interest to reduce dimensionality of the data without loss of information. In this regard, sufficient dimension reduction (SDR) recently has got much attentions due to its promising performance under less stringent model assumptions. We propose a new class of SDR approaches based on slicing conditional quantiles. We first show that the central subspace of the conditional quantiles is the same as that of the response. We then propose two versions of SDR estimation schemes: quantile-slicing mean estimation (QUME) and quantile-slicing variance estimation (QUVE). Quantile-slicing is particularly useful when quantile function is more efficient to capture underlying model structure than the response, for example, when heteroscedasticity exists in a regression context. Both simulated and real data analysis results demonstrate promising performance of the proposed quantile-slicing SDR estimation methods.

EO0412: **Primal path algorithm for compositional data analysis**
Presenter: **Jong-june Jeon**, University of Seoul, Korea, South
High-dimensional l_1 penalized regression and classification for compositional explanatory covariates is considered. To reflect the structure of the componentwise compositions, zero equality constraints on coefficients are further adopted. With the equality constraints, we develop the entire primal solution path algorithm that extends least angle regression algorithm. In addition, the proposed algorithm is extended to the more general loss function like piecewisely defined quadratic loss function. We apply the proposed algorithm for human gut microbiome data analysis.

EO0432: **Identifying differential networks in brain connectome graphs using a penalized regression model**
Presenter: **Donghyeon Yu**, Inha University, Korea, South
Co-authors: Sang Han Lee, Johan Lim, Guanghua Xiao, Cameron Craddock, Bharat Biswal
A procedure is proposed to find differential networks between two graphs from high-dimensional data. We estimate the precision matrices and their differences by solving a penalized regression-based least square problem. We assume sparsity on differences between two graphs, not graphs themselves. Thus, we impose an ℓ_2 penalty on partial correlations (not the elements of precision matrices) and an ℓ_1 penalty on their differences. We apply the proposed procedure to finding differential functional connectivity between healthy individuals and those suffering from Alzheimer's disease.

EO200 Room LSK1014 HIGH-DIMENSIONAL STATISTICS: TESTING, ESTIMATION AND BEYOND Chair: Weiming Li

EO0298: **On structure testing for component covariance matrices of a high-dimensional mixture**
Presenter: **Weiming Li**, Shanghai University of Finance and Economics, China
By studying the family of p-dimensional scaled mixtures, a non trivial example is shown for the first time where the eigenvalue distribution of the corresponding sample covariance matrix does not converge to the celebrated Marcenko-Pastur law. A different and new limit is found and characterized. The reasons of failure of the Marcenko Pastur limit in this situation are found to be a strong dependence between the p-coordinates of the mixture. Next, we address the problem of testing whether the mixture has a spherical covariance matrix. It is shown that the traditional Johns test and its recent high-dimensional extensions both fail for high-dimensional mixtures, precisely due to the different spectral limit above. In order to find a remedy, we establish a novel and general CLT for linear statistics of eigenvalues of the sample covariance matrix. Anew test using this CLT is constructed afterwards for the sphericity hypothesis.

EO0463: **Spectral analysis of linear time series in moderately high dimensions**
Presenter: **Lili Wang**, Zhejiang Gongshang University, China
Co-authors: Alexander Aue, Debashis Paul
The focus is on the spectral behavior of p-dimensional linear processes in the moderately high-dimensional case when both dimensionality p and sample size n tend to infinity so that $p/n \to 0$. It is shown that, under an appropriate set of assumptions, the empirical spectral distributions of the renormalized and symmetrized sample autocovariance matrices converge almost surely to a nonrandom limit distribution supported on the real line. The key assumption is that the linear process is driven by a sequence of p-dimensional real or complex random vectors with i.i.d. entries possessing zero mean, unit variance and finite fourth moments, and that the $p \times p$ linear process coefficient matrices are Hermitian and simultaneously diagonalizable. Several relaxations of these assumptions are discussed. The results put forth can help facilitate inference on model parameters, model diagnostics and prediction of future values of the linear process.

EO0623: **Sparsity oriented importance learning for high-dimensional linear regression**
Presenter: **Yi Yang**, McGill University, Canada
With now well-recognized non-negligible model selection uncertainty, data analysts should no longer be satisfied with the output of a single final model from a model selection process, regardless of its sophistication. To improve reliability and reproducibility in model choice, one constructive approach is to make good use of a sound variable importance measure. Although interesting importance measures are available and increasingly used in data analysis, little theoretical justification has been done. We propose a new variable importance measure, sparsity oriented importance learning (SOIL), for high-dimensional regression from a sparse linear modeling perspective by taking into account the variable selection uncertainty via the use of a sensible model weighting. The SOIL method is theoretically shown to have the inclusion/exclusion property: When the model weights are properly around the true model, the SOIL importance can well separate the variables in the true model from the rest. In particular, even if the signal is weak, SOIL rarely gives variables not in the true model significantly higher important values than those in the true model. Extensive simulations in several illustrative settings and real data examples with guided simulations show desirable properties of the SOIL importance in contrast to other importance measures.

EO252 Room LSK1003 RECENT ADVANCES IN COMPLEXLY-STRUCTURED TIME SERIES ANALYSIS Chair: Zhou Zhou

EO0304: **Optimal estimation of change-point in time series**
Presenter: **Chun Yip Yau**, Chinese University of Hong Kong, Hong Kong
Asymptotic theory is established for optimal estimation of change-points in time series. We show that the Bayes estimator is asymptotically efficient for change-point estimation under mean-squared error loss. Three subsampling procedures are developed to construct confidence intervals for the change-points. Simulations and real data applications are presented to investigate the finite sample performance of the Bayes estimator and the three subsampling procedures.

EO0383: **Time-varying network estimation from high-dimensional time series**
Presenter: **Mengyu Xu**, University of Central Florida, United States
Co-authors: Xiaohui Chen, Wei Biao Wu
The estimation of time-varying networks from high-dimensional time series is considered. Two types of non-stationarity are investigated: structural breaks and smooth changes. Our approach can achieve consistent detection of the change points and simultaneous estimation of the piece-wisely smooth-varying networks. Rates of convergence for estimating change points and networks are obtained under mild moment and dependence conditions.The method is applied to the analysis of network structure of the S&P 500 index.

EO0520: **Change point analysis of correlation in non-stationary time series**
Presenter: **Weichi Wu**, University College London, United Kingdom
Co-authors: Holger Dette, Weichi Wu, Zhou Zhou
A restrictive assumption in change point analysis is "stationarity under the null hypothesis of no change-point", which is crucial for asymptotic theory but not very realistic from a practical point of view. For example, if change point analysis for correlations is performed, it is not necessarily clear that the mean, marginal variance or higher order moments are constant, even if there is no change in the correlation. Change point analysis is developed for the correlation structures under less restrictive assumptions. In contrast to previous work, our approach does not require that the mean, variance and fourth order joint cumulants are constant under the null hypothesis. Moreover, we also address the problem of detecting relevant change points.

EO132 Room LSK1007 FINDING GROUP STRUCTURES IN BIOMEDICAL AND HEALTH DATA Chair: Shu-Kay Ng

EO0431: **PCA based clustering for ultrahigh-dimensional data**
Presenter: **Makoto Aoshima**, University of Tsukuba, Japan
Co-authors: Kazuyoshi Yata
High-dimension, low-sample-size (HDLSS) data situations occur in many areas of modern science such as genetic microarrays, medical imaging, text recognition, finance, chemometrics, and so on. We consider clustering based on principal component analysis (PCA) for HDLSS data. We give theoretical reasons why PCA is effective for clustering HDLSS data. First, we derive a geometric representation of HDLSS data taken from a two-class mixture model. With the help of the geometric representation, we give geometric consistency properties of sample principal component scores in the HDLSS context. We develop the idea of geometric representations and provide geometric consistency properties for multiclass mixture models. We show that PCA can classify HDLSS data under certain conditions in a surprisingly explicit way. Finally, we demonstrate the performance of the clustering by using microarray data sets. We show that HDLSS data sets consisting of several lung carcinomas types hold the geometric consistency properties.

EO0530: **Testing for group structure**
Presenter: **Geoffrey McLachlan**, University of Queensland, Australia
Cluster analysis is commonly applied to explore a data set for the presence of groups. Having found some group structure in the form of a number of clusters, there is the question of whether the clusters represent a genuine as opposed to a spurious grouping in the data. We consider this problem for high-dimensional data where the dimension of each observation vector p (that is, the number of variables) may greatly exceed the number of observations n. We consider a mixture model approach to this problem whereby a probabilistic clustering of the data into g clusters is effected by fitting a g-component mixture distribution. In this mixture model framework, we can approach the question of how many clusters there are in the data in terms of an assessment of the smallest number of components g in the mixture compatible with the data. The focus is on tests on g based on the likelihood ratio statistic. Initially, consideration is given to a reduction in the number of variables to allow the fitting of multivariate normal component distributions and also asymmetric/robust variants to handle possible skewness and long-tailedness in the clusters.

EO0630: **SNP-set clustering and testing with Hamming distance for association studies**
Presenter: **Chuhsing Kate Hsiao**, National Taiwan University, Taiwan
Co-authors: Charlotte Wang
The computational complexity due to the large number of genomic markers collected in association studies has raised challenges in statistical inference. One common approach is to reduce the number of markers via clustering techniques. These clusters are often termed as marker-set or SNP-set in genome-wise association studies. We employ the Hamming distance to measure the similarity between strings of SNP genotypes and evaluate whether the given SNPs or SNP-sets should be clustered. We next propose a test to examine if the resulting SNP-set is associated with the disease of interest. This test measures if the similarity, based on Hamming distance again, between two individuals of different disease status differs significantly from two individuals in the same disease group. The proposed methods are illustrated with applications and simulation studies. The results show that the clustering algorithm can effectively improve the power of any association test and the test outperforms other tests especially when the number of neutral SNPs is large.

EO122 Room LSK1005 RECENT DEVELOPMENTS ON DYNAMIC TREATMENT REGIMES Chair: Binyan Jiang

EO0657: **Statistical and dynamical systems modeling of m-intervention for pain**
Presenter: **Chae Ryon Kang**, University of Pittsburgh, United States
Co-authors: Daniel Abrams, Jingyi Li, Qi Long, Nirmish Shah
With the growing popularity of mobile phone technology, new opportunities have arisen for real-time adaptive medical intervention. The simultaneous growth of multiple big data sources (eg. mobile health data, electronic health records, lab test results, genomic data) allows for the development of personalized recommendations. We develop a new mathematical model for changes in subjective pain over time in patients with chronic conditions. The proposed model consists of a dynamical systems approach using differential equations to forecast future pain levels, as well as a statistical approach tying system parameters to patient data (including reported pain levels, medication history, personal characteristics and other health records). The model is combined with statistical techniques to ultimately obtain optimized, continuously-updated treatment plans balancing competing demands of pain reduction and medication minimization. Application of the resulting personalized treatment plans to a currently active pilot study on mobile intervention in patients living with chronic pain due to sickle cell disease (SCD) will be presented.

EO0656: **Value function inference in high-dimensional Q-learning for dynamic treatment regimes**
Presenter: **Wensheng Zhu**, Northeast Normal University, China
Co-authors: Donglin Zeng, Rui Song
A dynamic treatment regime is a set of decision rules, in which treatment decisions often need to be tailored over time according to patients responses to previous treatments as well as covariate history. There is a growing interest in development of statistical inference for optimal dynamic treatment regimens, the difficulty of which is the non-regularity problem in the presence of positive zero-treatment effect probabilities, especially when the dimension of covariates is very high. We propose a high-dimensional Q-learning (HQ-learning) to facilitate the inference of optimal values and parameters, which allows us to simultaneously estimate the optimal dynamic treatment regimes and select the important variables that truly contribute to the individual reward. To reduce the effect of non-regularity on the statistical inference of HQ-learning, we remove those subjects with very small treatment effects by using truncated pseudo-outcome in the first-stage penalized regression. The asymptotic properties for the parameter estimators as well as the estimated optimal value function are also established by adjusting for the bias due to the truncation. The simulation studies and real data analysis show that our HQ-learning can simultaneously estimate the optimal dynamic treatment regimens and selected the important variables very well in the presence of high-dimensional case and non-regularity.

EO0741: **Non-inferiority and equivalence tests in a sequential multiple-assignment randomized trial (SMART) design**
Presenter: **Bibhas Chakraborty**, Duke-NUS Medical School, National University of Singapore, Singapore
Co-authors: Palash Ghosh
In a sequential multiple-assignment randomized trial (SMART), a patient is given a sequence of treatments over the trial period. SMART designs are applicable in various branches of medical science like depression, behavioral science, cancer etc. The treatments/drugs used in a SMART could be from multiple pharmaceutical companies. Generally, pharmaceutical companies are reluctant to do a comparison of different marketed drugs from different companies by a traditional one/two-sided test, because a negative result may damage their business severely. In this context, a general perception is that a SMART design may not be a suitable choice for pharmaceutical companies. Contrary to that general perception, we argue that a SMART design could be profitable for pharmaceutical companies. It may be a win-win situation for all the companies involved in a SMART. We develop the non-inferiority and the equivalence test procedures to compare treatment-sequences in a SMART, to facilitate the uptake of SMART designs in the pharma-sponsored trials. We show that the trade-off between the efficacy and the quality of life (QOL) from a treatment-sequence can be addressed by a non-inferiority/equivalence test. Sample size and power issues have been discussed with supporting simulations. We show the applicability of the developed test procedures in the SMART context using the STAR*D data.

EO126 Room LSKG003 SOME NEW DEVELOPMENT IN COMPLEX SURVIVAL DATA Chair: Catherine Liu

EO0676: **Robust feature screening for ultra-high dimensional right censored data via distance correlation**
Presenter: **Xiaolin Chen**, Qufu Normal University, China
Two robust feature screening procedures for ultra-high dimensional right censored data will be introduced based on distance correlation. The first approach performs feature screening through replacing response and covariate by their cumulative distribution functions' Kaplan-Meier estimator and empirical distribution function respectively in the distance correlation, while the second one modifies the distance correlation via an idea of composite quantile regression. The proposed methods have some desirable characters. The advantages of the proposed methods could be seen in our simulation studies. A real data example analysis will also be presented.

EO0825: **Semiparametric transformation models with multi-Level random effects for family survival data**
Presenter: **Baosheng Liang**, The University of Hong Kong, Hong Kong
Co-authors: Yuanjia Wang, Donglin Zeng
For familial disease, one main task is to estimate the probability distribution of genetic mutation in carrier and non-carrier groups using pedigree's history information. This task becomes extremely difficult when only a single member's genotypes or mutation statuses in a family is known. We propose a class of transformation models with multi-level random effects for one kind of family data which arises from a population-based case-control family study with family members' genotypes missing. The class of transformation model could handle both the homogeneity within family and the heterogeneity of each family member. Specific conditions are developed to guarantee the identifiability of proposed model. Nonparametric maximum likelihood estimation approach is applied for parameters' estimation. An expectation-maximization algorithm based on multi-dimensional adaptive quadrature is developed. The nonparametric maximum likelihood estimator is shown to be consistent, asymptotically normal, and semiparametric efficient. Simulation studies show that the proposed method performs well and provides good inferential properties under proper finite sample sizes. The proposed methods are applied to an Alzheimer's disease study.

EC0704: **Transformation model for sparse functional data**
Presenter: **Guochang Wang**, Jinan University, China
Functional linear models are usually used to model the relationship between a scalar response and some functional predictors, because of its simplicity and easy interpretation. However, the linear form limits its application to data with more complicated structures and the nonparametric model suffers its low convergence rate. To combine the advantages both of the liner model and nonparametric model we consider the functional transformation model (FTM) and propose an innovative procedure called mixed data singular component analysis (MDSCA) to estimate the transformation function and the functional regression parameter, simultaneously. Compared with the existing methods, the proposed method is numerical stability and is insensitive to the selection of smoothing parameters and can be applied in sparse functional data. Furthermore, we also apply MDSCA to measure the correlation between a multivariate sample and a set of functional data. The asymptotic properties of MDSCA, the correlation coefficient between the sparse functional data and multivariate, the convergence rate for the transformation function and the functional regression parameter. Lastly, real and simulation data examples are further presented to show the value of these approaches.

EO262 Room LSK1001 NEW DEVELOPMENTS IN BIOMEDICAL RESEARCH I Chair: Jinfeng Xu

EO0725: **Additive partially linear models for massive heterogeneous data**
Presenter: **Yixin Fang**, New Jersey Institute of Technology, United States
Co-authors: Binhuan Wang, Heng Lian, Hua Liang
An additive partially linear framework is considered for modelling massive heterogeneous data. The major goal is to extract multiple common features simultaneously across all sub-populations while exploring heterogeneity of each sub-population. A previous partially linear framework, which considers only one common feature, is generalized. We propose an aggregation type of estimators for the commonality parameters that possess the asymptotic optimal bounds and the asymptotic distributions as if there were no heterogeneity. This oracle result holds when the number of sub-populations does not grow too fast and the tuning parameters are selected carefully. A plug-in estimator for the heterogeneity parameter is further constructed, and shown to possess the asymptotic distribution as if the commonality information were available. The performance of the proposed methods is evaluated via simulation studies and an application to the Medicare Provider Utilization and Payment data.

EO0731: **A simple test for dependent censoring**
Presenter: **Antai Wang**, New Jersey Institute of Technology, United States
A simple test for dependence between a failure time and a censoring time is proposed when there exists an independent censoring time. The procedure is demonstrated using a leukemia data set.

EO0740: **Penalized estimation of sparse concentration matrices based on prior knowledge with applications to placenta metal data**
Presenter: **Jiang Gui**, Dartmouth College, United States
Essential elements (P, S, K, Ca, Mg, Mn, Co, Fe, Cu, Zn, Se,) play critical cellular roles as structural components of bio-molecules, signalling molecules, catalytic co-factors and regulators of protein expression, and in biological systems their concentrations are homeostatically regulated. Altered metal homeostasis occurs in neurological diseases and cancer. The chemical architecture for metals may be complex, and advanced biostatistical methods are needed to infer the dependency structures of the metals. We introduce a weighted sparse Gaussian graphical model that can incorporate prior knowledge to infer the structure of the network of metal concentrations measured in the human term placenta. We present the L1 penalized regularization procedure for estimating the sparse precision matrix in the setting of Gaussian graphical models. Simulation results indicate that the proposed method yields a better estimate of the precision matrix than the procedures that fail to account for the prior knowledge of the network structure. We also applied this method to a New Hampshire Birth Cohort Study for inference of the chemical network in placenta biopsies.

EO100 Room LSK1011 NEW DEVELOPMENTS IN EXPERIMENTAL DESIGNS AND INDUSTRIAL STATISTICS Chair: Chang-Yun Lin

EO0687: **Model-robustly D- and A-optimal designs for log-contrast models in mixture experiments**
Presenter: **Hsiang-Ling Hsu**, Institute of Statistics, National University of Kaohsiung, Taiwan
The issue of the model robust optimal designs in a mixture experiment is investigated. A mixture experiment in $(q-1)$-dimensional probability simplex $S^(q-1)$ is an experiment in which the q factors are nonnegative and subject to the simplex restriction, i.e., the factors represent relative proportions of q ingredients in the experiment. We develop model-robustly D- and A-optimal designs for mixture experiments with consideration of uncertainties between the first-order and second-order models. The corresponding optimal designs and complete classes for each of the log contrast models are adopted to carry out the structure and the algorithms for obtaining the model robustly D- and A-optimal designs, verified through the well-known Kiefer-Wolfowitz equivalence theorem. Numerical model-robustly optimal designs are illustrated in our study and we also compare the behaviors of the efficiencies for the linear combination of the D- and A-optimal designs from the individual models.

EO0771: **A head-to-head comparative study of the conditional performance of control charts based on estimated parameters**
Presenter: **Inez Zwetsloot**, University of Amsterdam, Netherlands
It is often of interest to detect changes in data. Statistical process monitoring (SPM) provides tools to monitor data and signal changes in the data. Three well-known tools are the Shewhart control chart, the cumulative sum (CUSUM) control chart, and the exponentially weighted moving average (EWMA) control chart. Before a control charts can be set up, estimates of the process parameters are needed. To this end, an initial data set is collected. Many researchers have shown that the resulting estimation error strongly influences the performance of these charts. However, a given amount of estimation error may differ in effect across the three control charts. Therefore, we perform a pairwise comparison of the effect of estimation error across these charts. We conclude that the Shewhart chart is more strongly affected by estimation error than the CUSUM and EWMA charts. Furthermore, we show that the general belief that the CUSUM and EWMA charts have similar performance no longer holds under estimated parameters.

EO0723: **Inference for response adaptive designs in multi-center clinical trials**
Presenter: **Alwell Oyet**, Memorial University, Canada
Co-authors: Selvakkadunko Selvaratnam
The implementation of response adaptive designs in multi-center clinical trials is discussed and a generalized linear mixed model (GLMM) is developed for analyzing data obtained from the trials. The asymptotic properties of maximum likelihood estimators of the model parameters are derived using influence function techniques. The advantage of using the influence function approach is that it leads to a closed form expression for the asymptotic covariance of the estimated parameters. To our knowledge, such a closed form expression does not currently exist in the literature. The performance of the maximum likelihood estimator under various response adaptive designs is examined through simulation studies. We also use our simulation studies to compare the asymptotic covariance matrix based on the influence function to that based on the inverse of the Hessian matrix obtained from the likelihood function of the observations. The techniques are applied to real data obtained from a multi-center clinical trial designed to compare two cream preparations (active drug/control) for treating an infection.

EG165 Room LSK1034 CONTRIBUTIONS IN SEMI-PARAMETRIC METHODS IN ECONOMETRICS Chair: Claudio Morana

EC0179: **A new semiparametric approach for nonstandard econometric problems**
Presenter: **Bo Zhou**, Tilburg University, Netherlands

A new approach is developed to derive efficiency bounds and efficient tests for non-standard problems in semiparametric econometric models where the innovation density is treated as an infinite-dimensional nuisance parameter. The approach is based on an explicit nonparametric modelling of the innovation density and a structural version of the limit experiment. The structural limit experiment exhibits an invariance restriction that is employed to eliminate the nuisance parameter. The associated maximal invariant gives the efficiency bound using the Neyman-Pearson lemma. Moreover, the invariance structure naturally leads to statistical inference procedures. For example, the appearance of Brownian Bridges suggests the use of rank statistics. A simple linear regression model is used as a running example, leading to the same efficiency bounds and tests as the traditional least-favorable parametric submodel approach. We apply our results to two nonstandard econometric problems: testing cointegration rank and testing hypothesis with weak instruments. In both cases, we derive the semiparametric power envelopes and the induced rank-based inference procedures.

EC0796: **Semiparametric Bayesian forecasting with an application to stochastic volatility**
Presenter: **Fabian Goessling**, University of Muenster, Germany
Co-authors: Martina Danielova Zaharieva

A new Bayesian sampling algorithm is proposed for nonlinear state space models under nonparametric distributions. In particular, the estimation framework combines particle filtering and smoothing techniques for the unobserved latent process with a Dirichlet process mixture model based on the Chinese restaurant process representation for the error term. Compared to the existing sampling algorithms for such models, the proposed one offers a more general approach as neither conjugacy of the prior distributions nor any model transformations are necessary in order to sample the trajectories of the unobserved variables. Moreover, our approach straightforwardly nests non-latent or parametric models, such that it is easily adopted to a variety of models. Thus, several examples for the nested models are provided and subsequently the approach is applied to a Bayesian semiparametric stochastic volatility model for simulated and real financial data.

EC0188: **Semiparametric estimation of multivariate GARCH models**
Presenter: **Claudio Morana**, Universita di Milano-Bicocca, Italy

A new estimator is introduced for the conditional variance covariance and correlation matrix (SP-DCC). While sharing a similar sequential approach to existing dynamic conditional correlation (DCC) methods, SP-DCC has the advantage of not requiring the direct parameterization of the conditional covariance or correlation processes. In the proposed framework, conditional variances are estimated by univariate GARCH models, for actual and suitably transformed series, in the first step; the latter are then nonlinearly combined in the second step, according to basic properties of the covariance and correlation operator, to yield nonparametric estimates of the various conditional covariances and correlations. Moreover, in contrast to available DCC methods, SP-DCC allows for straightforward estimation also for the non-simultaneous case, i.e., for the estimation of conditional cross-covariances and correlations, displaced at any time horizon of interest. A simple ex-post procedure, to ensure well behaved conditional covariance and correlation matrices, grounded on nonlinear shrinkage, is finally proposed. Due to its sequential implementation and scant computational burden, SP-DCC is very simple to apply and suitable for the modeling of vast sets of conditionally heteroskedastic time series. Monte Carlo evidence strongly support the proposed methodology, performing better than standard DCC approaches under various DGPs.

EG003 Room LSK1032 CONTRIBUTIONS IN APPLIED ECONOMETRICS Chair: Jan Hagemejer

EC0713: **The nonlinear nature of country risk and its implications for DSGE models**
Presenter: **Jacek Kotlowski**, Narodowy Bank Polski, Poland
Co-authors: Michal Brzoza-Brzezina

Country risk premia can substantially affect macroeconomic dynamics. We concentrate on one of their most important determinants - a country's net foreign asset position and - in contrast to the existing research - investigate its nonlinear link to risk premia. The importance of this particular nonlinearity is twofold. First, it allows us to identify the NFA level above which the elasticity becomes much (possibly dangerously) higher. Second, such a nonlinear relationship is a standard ingredient of DSGE models, but its proper calibration/ estimation is missing. We estimate a nonlinear panel model based on data from 41 advanced and emerging economies. First we test the general nonlinearity. While the linear relationship between risk premium and NFA position has been strongly rejected we estimate the nonlinear panel smooth transition regression (PSTR) model with the logistic transition function, as preferred by the data. Our estimation shows that indeed the link is highly nonlinear and helps to identify the NFA position where the nonlinearity kicks in at -70% to -80% of GDP. We also provide a proper calibration of the risk premium - NFA relationship used in DSGE models and demonstrate that its slope matters significantly for economic dynamics in such a model.

EC0824: **Political instability and its effect on GDP growth and foreign direct investments**
Presenter: **Stefanos Dimitrakopoulos**, Oxford Brookes University, United Kingdom
Co-authors: Dimitrios Asteriou, Mike Tsionas

Using a panel of 32 countries covering the period 1980-2012, we first study the effects of political instability on GDP per capita growth. To this end, we propose a dynamic panel data factor model with stochastic volatility. The proposed model specification controls for growth volatility, systemic economic uncertainty and international business cycle effects. To estimate the model parameters we use particle Markov Chain Monte Carlo methods. We also extend the analysis to a two-equation simultaneous panel data model in order to examine the effects of political instability on both the GDP per capita growth and foreign direct investments.

EC0761: **On rushed privatizations**
Presenter: **Jan Hagemejer**, Narodowy Bank Polski, Poland
Co-authors: Joanna Tyrowicz, Jan Svejnar

The aim is to provide the first analysis of whether rushed privatizations, usually carried out under fiscal duress, increase or decrease efficiency and scale of operation (size) of firms. Using a large panel of firm-level data from Poland over 1995-2015, we show that rushed privatization has negative efficiency and scale effects relative to non-rush privation. The negative effect of rushed privatization on the scale of operations is even stronger than its negative effect on TFP. We also find that the negative relative effect of rushed privatization grows over time. A variety of modern econometric methods is used in order to deal with selection problems and endogeneity. We apply propensity score matching in order to account for selection of ex ante better firms for privatization as well as to account for the differences between firms privatized in rush and non-rush years. We use a double difference in difference identification strategy to disentangle both the treatment effect of privatization as well as the difference between effects of privatization in rush versus non-rush years. Last but not least, we also use a regional political instrumental variable to account for endogeneity of the privatization process. Results are robust and stable across specifications.

Friday 16.06.2017 08:30 - 09:50 Parallel Session F – EcoSta2017

EO066 Room LSKG003 RECENT ADVANCE IN TIME SERIES ECONOMETRICS Chair: Tingting Cheng

EO0211: **Strong trending time series models with endogeneity: A bias-correction method**
Presenter: **Li Chen**, Monash University, Australia

Endogeneity problems are studied in strong trending time series regression models. We explore the properties of the simple OLS estimator under difference cases of trending magnitudes for the nonstationary time series regressors. A bias corrected estimator is proposed to adjust for the bias in the simple OLS estimator. For the hypothesis tests, numerical simulations show that compared to the OLS estimator, the probability of making the type I error is significantly reduced based on the bias corrected estimator. We also apply our method to estimate the linear regression model of aggregate disposable personal income on the aggregate personal consumption and the real interest rate as an empirical example.

EO0255: **Estimating a large system of seemingly unrelated regressions using penalized quasi-maximum likelihood estimation**
Presenter: **Qingliang Fan**, Xiamen University, China

Using a shrinkage estimator, a penalized quasi-maximum likelihood estimator (PQMLE) is proposed to estimate a large system of equations in seemingly unrelated regressions models where the number of equations is large w.r.t. the sample size. We develop asymptotic properties of the penalized quasi-maximum likelihood estimator for both the error covariance matrix and the model coefficients. In particular, we derive the asymptotic distribution of the coefficients' estimator and the convergence rate of the estimated covariance matrix in terms of the Frobenius norm. The oracle property of the covariance matrix estimator is also established. Simulation results show that when the number of equations is large w.r.t. the sample size and the error covariance matrix is sparse, the penalized likelihood estimator performs much better than conventional estimators. As an illustration, we apply the PQMLE to the study of state level public capital returns in the United States.

EO0257: **Adaptive estimation of functional-coefficients regressions with time-varying variance**
Presenter: **Yundong Tu**, Peking University, China

The focus is on the efficient estimation of the functional-coefficient stationary regression model in presence of heteroskedasticity of a multiplicative form. The local least square (LLS) estimator is found to be subject to a potential efficiency loss. To address the inefficiency problem, we propose an adaptive version of the LLS estimator based on consistent estimates of the heteroskedasticity generating functions. The proposed adaptive local least square estimator (ALLS) is shown to be asymptotically as efficient as the infeasible generalized estimator. Simulations confirm the asymptotic efficiency loss of the LLS estimator and demonstrate the superiority of our proposed adaptive estimator concerning the finite sample estimation of the functional coefficient. The effect of fiscal policy to the US 10-year treasury bond yield is investigated as an illustration of the heteroskedasticity considered and the proposed adaptive estimation procedure.

EO0316: **Multi-step non- and semi-parametric predictive regressions for stock return prediction**
Presenter: **Tingting Cheng**, Nankai University, China
Co-authors: Jiti Gao, Oliver Linton

Three new predictive models are proposed: the multi-step nonparametric predictive regression model (NPR) and the multi-step additive predictive regression model (APR), in which the predictive variables are locally stationary time series; and the multi- step time-varying coefficient predictive regression model (TVCPR), in which the predictive variables are nonstationary. We also establish the estimation theory and asymptotic properties for these models. To evaluate the effectiveness of these models, we investigate their capability of stock return prediction. The empirical results show that all of these models can substantially outperform the traditional linear predictive regression model in terms of both in-sample and out-of-sample performance. In addition, we find that these models can always beat the historical mean model in terms of in-sample fitting, and also for some cases in terms of the out-of-sample forecasting.

EO196 Room LSK1014 BIG DATA AND ITS APPLICATIONS Chair: Benson Shu Yan Lam

EO0554: **Binary quadratic program algorithms for large scale problems**
Presenter: **Benson Shu Yan Lam**, Hang Seng Management College, China

Many image processing, pattern recognition, and computer vision problems can be formulated as binary quadratic programming (BQP) problems. Due to the NP-hard nature of these problems, the calculation of an extremely large BQP problem with a very good quality solution in real time is still a challenging and unsolved problem for which many different methods have been developed. For a typical image processing problem, it can involve an image with size $600x700$. Then, the corresponding BQP problem needs to deal with at least an nxn matrix with $n = 600x700$. In the current stage of development, spectral and semidefinite relaxation techniques are widely used to solve image processing, pattern recognition, and computer vision problems. The spectral approach is simple and fast, but its error bounds are loose with low solution qualities. The semidefinite approach has tighter error bounds with high solution qualities, but its computational complexity is high. We first show how real-world large-scale problems be formulated as BQP problems with complicated linear and/or quadratic constraints. Then, we present algorithms that can find good solutions in real time.

EO0666: **Parallel sentiment analysis for sales forecasting with big data**
Presenter: **Raymond Lau**, City University of Hong Kong, Hong Kong

While much research work has been devoted to demand forecast, research on designing big data analytics methodologies to enhance sales forecasting is seldom reported in existing literature. The big data of consumer-contributed product comments on online social media provide management with unprecedented opportunities to leverage collective consumer intelligence for enhancing supply chain management in general and sales forecasting in particular. The main contribution is the design of a novel big data analytics methodology that is underpinned by a parallel aspect-oriented sentiment analysis algorithm for mining consumer intelligence from a huge number of online product comments for enhancing sales forecasting. Based on real-world big datasets, our experimental results confirm that consumer sentiments mined from big data can improve the accuracy of sales forecasting across predictive models and datasets. To our best knowledge, this is the first successful research of developing a parallel aspect-oriented sentiment analysis method for big data, and the application of such a method to enhance sales forecasting. The managerial implication of our work is that firms can apply the proposed big data analytics methodology to enhance sales forecasting performance. Thereby, the problem of under/over-stocking is alleviated and customer satisfaction is improved.

EO0680: **On the moderation effects in social media stock opinions on stock return**
Presenter: **Kar Kei Lo**, The University of Hong Kong, Hong Kong
Co-authors: Michael Chau

Social media are becoming more prevalent for investors to seek information for making investment decisions. Although previous research has shown that sentiment analysis in social media may help predict future stock returns, the results could be inconsistent, and there is a lack of research in applying data analytics on understanding the effect of moderators in stock opinions on future stock performance. Investigating such an effect can be crucial in understanding why inconsistent results exist and creating a strategic value for investors and shareholders. Using social media data

from an investment forum, we explore the effect of moderators in stock opinions on stock returns in terms of three aspects involving the quality of articles, author's position, and author's popularity. We find that a subset of good-quality articles can be helpful as investment advice in social media. In terms of the author's position, we discover that the stated long and short position by the author has a direct effect on stock returns irrespective of the sentiment level in the related articles. In addition, articles written by authors with more followers have a better predictive power on future stock performance.

EO0694: **News co-mention and stock returns correlation**
Presenter: **Alvin Leung**, City University of Hong Kong, Hong Kong
News is a major source of information where investors gather to make investment decision. When two firms are frequently mentioned together in news, investors are likely to pay keen attention to both companies. The frequently co-mentioned companies may form an information habitat. According to information diffusion theory in finance, firms in the same habitat are likely to exhibit returns comovement. Taking into account of coverage and persistency of co-mentioned news, we construct a dynamic industry classification scheme with the hope that it can outperform traditional approaches to classify firms into relevant industries.

EG297 Room LSK1007 CONTRIBUTIONS IN STATISTICAL MODELS WITH APPLICATIONS Chair: Won Chang

EC0328: **Improving ice sheet model calibration using paleoclimate and modern data**
Presenter: **Won Chang**, University of Cincinnati, United States
Co-authors: Murali Haran, Patrick Applegate, David Pollard
Human-induced climate change may cause significant ice volume loss from the West Antarctic Ice Sheet (WAIS). Projections of ice volume change from ice-sheet models and corresponding future sea-level rise have large uncertainties due to poorly constrained input parameters. In most future applications to date, model calibration has utilized only modern or recent (decadal) observations, leaving input parameters that control the long-term behavior of WAIS largely unconstrained. Many paleo-observations are in the form of localized time series, while modern observations are non-Gaussian spatial data; combining information across these types poses non-trivial statistical challenges. We introduce a computationally efficient calibration approach that utilizes both modern and paleo-observations to generate better-constrained ice volume projections. Using fast emulators built upon principal component analysis and a reduced dimension calibration model, we can efficiently handle high-dimensional and non-Gaussian data. We apply our calibration approach to the PSU3D-ICE model which can realistically simulate long-term behavior of WAIS. Our results show that using paleo observations in calibration significantly reduces parametric uncertainty, resulting in sharper projections about the future state of WAIS. One benefit of using paleo observations is found to be that unrealistic simulations with overshoots in past ice retreat and projected future regrowth are eliminated

EC0728: **The Poisson-reciprocal inverse Gaussian distribution: Properties and applications**
Presenter: **Enrique Calderin**, University of Melbourne, Australia
The barely known continuous reciprocal inverse Gaussian distribution is used to introduce the Poisson-reciprocal inverse Gaussian discrete distribution. Many of its most relevant statistical properties are examined, some of them directly inherited from the reciprocal of the inverse Gaussian distribution. Furthermore, a mixed Poisson regression model that uses the reciprocal inverse Gaussian as mixing distribution is presented. Parameters estimation in this regression model is performed via an EM algorithm. In light of the numerical results, the distributions introduced are competitive with the negative binomial and Poisson–inverse Gaussian distributions.

EC0729: **A statistical modelling strategy for non-ignorable non-response in household surveys**
Presenter: **Oya Kalaycioglu**, Abant Izzet Baysal University, Bolu, Turkey
In household surveys, respondents often refuse to answer the questions related to their income levels due to their low or high levels of income. This results in non-ignorable non-response, as the missing values of the income level depend on the respondents' missing income values. Therefore, in practice, the analysts are more inclined to discard the respondents with missing data from the analysis or impute the missing values using the methods that assume missing data is ignorable. However, any statistical analysis that ignores non-ignorable missing data may result in biased parameter estimates and degrade the performance of confidence intervals. The aim is to emphasise the importance of accounting for non-ignorable missing data, review the modelling strategies which can be used in the case of non-ignorable outcome and present a case study using a data set obtained from a household survey. In the case study, a regression model is fitted to assess the factors that affect household income level and the coefficient estimates of this regression model is assessed under different assumptions of missing data through a sensitivity analysis. For conducting the sensitivity analysis a selection modelling framework is used and the results showed that the coefficient estimates of regression model can substantially change if non-ignorable non-response occurs in the outcome variable.

EC0748: **A new approach for estimating split point and confidence set in MOB trees using change point framework**
Presenter: **Suneel Babu Chatla**, National Tsing Hua University, Taiwan
Model-based (MOB) trees are popular for their ability to incorporate parametric models into classification and regression trees. The MOB recursive partitioning algorithm enables constructing a tree with estimated models in each terminal node. While the classification and regression trees are primarily data driven and exploratory, MOB brings together the advantages of both data-driven and theory driven. The existing implementation of MOB uses an exhaustive search to identify the split points and it is computationally very intensive. In addition, there is no way to assess the variability of the identified split points, which is crucial for the generalizability of the results. To solve this problem, we propose a new split point estimation method which uses a change point framework. Our proposed estimation is not only computationally much faster than the default greedy search, but it also provides the confidence sets for each split point using bootstrap methodology. We consider two popular bootstrap methods: regular and smoothed. Through a simulation study, we compare the regular and smoothed bootstrap methods and outline their merits in determining confidence sets. We illustrate the usefulness of our approach using real world datasets.

EG029 Room LSK1034 CONTRIBUTIONS IN FORECASTING ECONOMIC AND FINANCIAL TIME SERIES Chair: Peter Exterkate

EC0603: **Bayesian identification of support and resistance levels in financial time series**
Presenter: **Milan Ficura**, University of Economics in Prague, Czech Republic
Co-authors: Jiri Witzany
A methodology is presented for the Bayesian identification of Support & Resistance (SR) levels in financial time series of asset prices. SR levels, commonly used by stock, forex and commodity traders, represent price levels at which large volumes of buy (or sell) orders are concentrated. This causes the asset price to get pushed away from these levels whenever it approaches them and the orders get executed. A methodology is proposed to test for the presence of SR levels in the asset price evolution, by using the Stochastic-Volatility Jump-Diffusion (SVJD) model framework. Specifically a modified SVJD model is proposed that allows for the presence of price levels at whose proximity the conditional drift of the price process becomes significantly different from the unconditional drift. The posterior distributions of the latent state variables and the parameters of the model, including the locations and the strength of the potential SR levels, are then estimated with the Bayesian Markov-Chain Monte-Carlo

method. The analysis is performed on multiple assets and time periods with the results confirming the usefulness of SR levels for asset price modelling.

EC0655: **Testing forecast accuracy of expectiles and quantiles with the extremal consistent loss functions**
Presenter: **Yu-Min Yen**, National Chengchi University, Taiwan

Statistical tests are developed for comparing performances of forecasting expectiles and quantiles of a random variable under consistent loss (scoring) functions. Kolmogorov-Smirnov's type test statistics are constructed by using the extremal consistent loss functions. The null hypothesis of the tests is that a benchmark forecast performs at least equally well as a competitive one under all extremal consistent loss functions. It can be shown that if such a null hypothesis holds, the benchmark will also perform at least equally well as the competitor under all consistent loss functions. Thus, under the null hypothesis, when different consistent loss functions are used, the result that the competitor does not outperform the benchmark will not be altered. We propose to use the re-centered bootstrap to construct empirical distributions of the proposed test statistics. Through simulations, we show the proposed test statistics perform reasonably well. We apply the proposed test on re-examining abilities of some predictors on forecasting risk premium of the S&P500 index.

EC0768: **A regime-switching stochastic volatility model for forecasting electricity prices**
Presenter: **Peter Exterkate**, University of Sydney, Australia
Co-authors: Oskar Knapik

In a recent review paper, several crucial challenges outstanding in the area of electricity price forecasting are pinpointed. The aim is to address all of them by (i) showing the importance of considering fundamental price drivers in modelling, (ii) developing new techniques for density forecasting of electricity prices, and (iii) introducing a universal Bayesian technique for model comparison to this literature. We propose a regime-switching stochastic volatility model with three regimes (negative jump or "drop", normal price or "base", positive jump or "spike") where the transition matrix depends on explanatory variables in a novel way, using an underlying ordered probit model. The main focus is on one-day-ahead density forecasting of hourly prices in the Nord Pool market, where all Northern European electricity is traded. We show that the proposed model outperforms several benchmark models at this task, as measured by their predictive Bayes factors.

EC0773: **Reduced rank regression in large VARs**
Presenter: **Kelly Trinh**, University of Queensland, Australia

The over-parameterization issue in large vector autoregressive (VARs) models is addressed by using reduced rank regression. Our aim is to provide a model specification which is invariant to the variable ordering. To achieve a computable and efficient algorithm for estimating such large VAR models, we adopt the parameter expansion approach, and a computation procedure exploiting a certain Kronecker structure of covariance matrices. We carry out an extensive Monte Carlo simulation to examine the performance of marginal likelihood approximated by using cross entropy, predictive likelihood, and Laplace approximation in selecting the number of ranks of the VAR matrices. The results reveal that these approaches underestimate the number of ranks when the dimension of VAR systems grows, and when the singular values of the VAR matrices are small (close to zero). We go further to examine the forecast performance of the models with misspecified ranks relative to the model with the correct rank (the benchmark model) using the measures of point forecast and density forecast. Our results suggest that the models with lower rank perform worse than the benchmark for short forecast horizons, however, they perform as well as or even beat the benchmark for long forecast horizons. These patterns are more evident when the magnitudes of the singular values of the VAR coefficient are small.

EG069 Room LSK1003 CONTRIBUTIONS IN VOLATILITY MODELLING AND FORECASTING Chair: Gianluca Cubadda

EC0667: **Bayesian analysis of generalized long-memory stochastic volatility**
Presenter: **Alex Gonzaga**, University of the Philippines Manila, Philippines

A Bayesian approach is proposed for estimating the parameters and future values of the Generalized Long-memory Stochastic Volatility (GLMSV) model will be predicted by utilizing the approximate likelihood function of discrete wavelet packet coefficients. This provides an alternative method incorporating prior information about the model parameters and utilizing the decorrelating property of wavelet packet transform of the model. This simplifies the variance-covariance matrix of the model by approximately decorrelating wavelet coefficients within and across scales. This approximation does not depend on the signal, but the length of the wavelet filter, which is under the control of the analyst. It allows for a computationally efficient sampling from the posterior density of the parameters in a Bayesian approach for estimating and predicting future values. The proposed method is then applied in the analysis of volatility of financial time series, such as the intraday volatility of stock prices.

EC0782: **Global volatility transmission with overlapping trading zones**
Presenter: **Andreas Masuhr**, Westfalische Wilhelmsuniversitat Munster, Germany

The transmission of volatility in global financial markets is considered by using a multivariate copula GARCH model with uncorrelated but dependent errors. The specification allows geographic trading zones to overlap in terms of trading times and thus can be adapted to the actual trading hours of principal exchanges. Additionally the model can flexibly account for non normal marginal distributions as well as non normal dependency structures between the returns of different trading zones and therefore adapts stylized facts as heavy tails and tail dependence. The model is estimated by MCMC methods and supports the hypothesis that volatility is transmitted from both same day news from preceding trading zones and news of previous days in the same trading zone. Adding the same day volatility of the immediately preceding zone as explanatory variable reveals even stronger volatility transmission patterns between trading zones.

EC0375: **Testing for jumps in prices under jump-driven leverage effect in volatility: A simulation study**
Presenter: **Ping Chen Tsai**, Southern Taiwan University of Science and Technology, Taiwan

Recent non-parametric tests for jumps in prices are robust against a diffusion-driven leverage effect in volatility. When the leverage effect is channelled via correlated jumps, it is less clear whether the existing jump detection methods give consistent results. To evaluate the impact of jump-driven leverage effect on the properties of jump tests, a simulation study is conducted using two stochastic volatility models in which one is purely diffusion-driven and another can be motivated by jumps. The two volatility models are specified and their parameters calibrated in a way that their marginal distributions are approximately identical. The simulation results indicate that a backward-looking test sees its detection power impaired by jump-driven leverage effect in volatility.

EC0797: **Bayesian semiparametric multivariate stochastic volatility**
Presenter: **Martina Danielova Zaharieva**, University of Muenster, Germany
Co-authors: Mark Trede, Wilfling Bernd

The proposed framework is a Cholesky-type multivariate stochastic volatility model, in which the multivariate distribution of the error term is modeled as an infinite scale mixture of Gaussian distributions. A Bayesian non-parametric approach, in particular a Dirichlet process mixture, is adopted. This allows for high flexibility of the return distribution with respect to the kurtosis. Furthermore, the Cholesky decomposition allows for parallel univariate process modeling, creating potential for the estimation of higher dimensional models. Markov Chain Monte Carlo methods are applied for the posterior simulation and the computation of the predictive density. Finally, a five-dimensional stock market application is presented.

EC288 Room LSK1005 CONTRIBUTIONS IN ROBUST METHODS Chair: Germain Van Bever

EC0722: **Total variation depth for functional data**
Presenter: **Ying Sun**, KAUST, Saudi Arabia
Co-authors: Huang Huang
There has been extensive work on data depth-based methods for robust multivariate data analysis. Recent developments have moved to infinite-dimensional objects such as functional data. We propose a new notion of depth, the total variation depth, for functional data. As a measure of depth, its properties are studied theoretically, and the associated outlier detection performance is investigated through simulations. Compared to magnitude outliers, shape outliers are often masked among the rest of samples and harder to identify. We show that the proposed total variation depth has many desirable features and is well suited for outlier detection. In particular, we propose to decompose the total variation depth into two components that are associated with shape and magnitude outlyingness, respectively. This decomposition allows us to develop an effective procedure for outlier detection and useful visualization tools, while naturally accounting for the correlation in functional data. Finally, the proposed methodology is demonstrated using real datasets of curves, images, and video frames.

EC0763: **Halfspace depth for scatter, concentration and shape matrices**
Presenter: **Germain Van Bever**, Universite libre de Bruxelles, Belgium
Co-authors: Davy Paindaveine
Half-space depth concepts are proposed for scatter, concentration and shape matrices. For scatter matrices, our concept extends previous ones to the non-centered case. Rather than focusing, as in earlier works, on deepest scatter matrices, we thoroughly investigate the properties of the proposed depth and of the corresponding depth regions. We do so under minimal assumptions and, in particular, we do not restrict to elliptical distributions nor to absolutely continuous distributions. Interestingly, fully understanding scatter halfspace depth requires considering different geometries/topologies on the space of scatter matrices. We also discuss the structural properties a scatter depth should satisfy, and investigate whether or not these are met by the proposed depth. As mentioned above, companion concepts of depth for concentration matrices and shape matrices are also proposed and studied.

EC0272: **Heteroskedasticity and autocorrelation robust inference for a system of regression equations**
Presenter: **Robert Anderson**, Newcastle Univeristy, United Kingdom
Co-authors: Denise Osborn, Ralf Becker
Standard single equation heteroskedasticity and autocorrelation (HAC) robust inference methods are extended to allow consistent inference for a system of vector moving-average correlated equations also accommodating contemporaneous correlations. This is of particular relevance to the examination of inflation forecast errors, as forecasts for different groups are contemporaneously correlated, while any proposed forecasting model utilising a time-series of multi-period forward-looking expectations data will suffer from overlapping errors inducing a moving-average error structure. The proposed methodology is a generalisation of previous work. Monte Carlo simulations confirm that the method performs well in large samples. Applications testing the rationality of male versus female inflation forecasts, and those of defined educated groups, are also included.

EC0738: **Robust inference in a heteroskedastic multilevel linear model with structural change**
Presenter: **Ronrick Da-ano**, University of the Philippines-Diliman, Philippines
Co-authors: Erniel Barrios, Joseph Ryan Lansangan
The aim is to estimate a multilevel model with cross-sectional interactions in higher and lower levels with structural change by a hybrid procedure of the forward search algorithm preceding bootstrap method. The simulation study exhibits the ability of the hybrid procedure to produce narrower confidence interval even when there is model misspecification error and structural change. Moreover, it has a comparable predictive ability with the classical restricted maximum likelihood (REML) estimation. However, the hybrid method yield estimates of the parameters with lower bias relative to REML. The hybrid of forward search and bootstrap method can further robustify estimates of fixed and random coefficients under various levels of interclass correlation or in the presence of structural change in a multilevel model.

EG013 Room LSKG001 CONTRIBUTIONS IN HIGH DIMENSIONAL AND COMPLEX DATA ANALYSIS Chair: Tommaso Proietti

EC0758: **A classification procedure based on eigenstructures in the high-dimension, low-sample-size context**
Presenter: **Aki Ishii**, Tokyo University of Science, Japan
One of the features of modern data is that the data dimension is extremely high, however, the sample size is relatively low. We call such data HDLSS data. We consider two-class classification for HDLSS data. We note that eigenvalues of high-dimensional data grow very rapidly depending on the dimension. There are two types of high-dimensional eigenvalue models: the strongly spiked eigenvalue (SSE) model and the non-SSE (NSSE) model. A lot of works have been done under the NSSE model. In particular, there was the misclassification rate adjusted classifier that can ensure accuracies in classification rates for two classes simultaneously under the NSSE model. We consider the two-class classification under the SSE model. Based on the eigenstructures of the two-classes, we give a classification rule that can ensure high accuracy in classification rates for the two classes simultaneously. Note that one can check the validity of assumptions by using our method. We compare the performance of our classifier to other previous classifiers. Finally, we show that our classification procedure works well by using actual microarray data sets.

EC0756: **Estimating multivariate and conditional density functions using local Gaussian approximations**
Presenter: **Haakon Otneim**, Norwegian School of Economics, Norway
Co-authors: Dag Tjoestheim
It is well known that the nonparametric kernel density estimator breaks down quickly as the dimension increases. We will present a procedure for multivariate density estimation that makes use of the local likelihood framework, but with crucial adaptations so that we circumvent the curse of dimensionality without introducing too much modeling error. This method, that is based on local Gaussian approximations, can be used for conditional density estimation in a natural way. Simulation experiments and applications to real data will illustrate the finite sample performance of the estimator, and asymptotic results will be presented as well.

EC0363: **A mechanistic nonlinear model for censored and mis-measured covariates in longitudinal modeles**
Presenter: **Hongbin Zhang**, CUNY (SPH), United States
When modeling longitudinal data, the true values of time-varying covariates may be unknown due to detection-limit censoring or measurement error. A common approach in the literature is to empirically model the covariate process based on observed data, and then predict the censored values or mis-measured values based on this empirical model. Such an empirical model can be misleading, especially for censored values since the (unobserved) censored values may behave very differently than observed values due to the underlying data-generation mechanisms or disease status. We propose a mechanistic nonlinear covariate model based on the underlying data-generation mechanisms to address censored values and mis-measured values. Such a mechanistic model is based on solid scientific or biological arguments, so the predicted censored or mis-measured values are more reasonable. We use a Monte Carlo EM algorithm for likelihood inference, and apply the methods to an AIDS dataset, where viral

load is censored by a lower detection limit. Simulation results confirm that the proposed models and methods offer substantial advantages over existing empirical covariate models for censored and mis-measured covariates.

EC0851: **A Durbin-Levinson regularized estimator of high-dimensional autocovariance matrices**
Presenter: **Tommaso Proietti**, University of Roma Tor Vergata, Italy
Co-authors: Alessandro Giovannelli
The problem of estimating the high-dimensional autocovariance matrix of a stationary random process is considered, with the purpose of out of sample prediction and feature extraction. This problem has received several solutions. In the nonparametric framework, the literature has concentrated on banding and tapering the sample autocovariance matrix. We propose and evaluate an alternative approach based on regularizing the sample partial autocorrelation function, via a modified Durbin-Levinson algorithm that receives as input the banded and tapered partial autocorrelations and returns a sample autocovariance sequence which is positive definite. We show that the regularized estimator of the autocovariance matrix is consistent and its convergence rates is established. We then focus on constructing the optimal linear predictor and we assess its properties. The computational complexity of the estimator is of the order of the square of the banding parameter, which renders it feasible for high-dimensional time series. We evaluate the performance of the autocovariance estimator and the corresponding linear predictor by simulation and empirical applications.

EC295 Room LSK1010 CONTRIBUTIONS IN METHODOLOGICAL STATISTICS AND ECONOMETRICS Chair: Juwon Seo

EC0801: **Sparse Poisson regression with a penalized weighted score function**
Presenter: **Fang Xie**, University of Macau, China
Co-authors: Lihu Xu, Jinzhu Jia
A new penalized method is proposed to solve sparse Poisson Regression problems. Being different from l1 penalized log-likelihood estimation, our new method can be viewed as a penalized weighted score function method. We show that under mild conditions, our estimator is l1 consistent and the tuning parameter can be pre-specified, which owns the same good property of the square-root Lasso. The simulations show that our proposed method is much more robust than traditional sparse Poisson models using l1 penalized log-likelihood method.

EC0760: **A rank test for linear hypothesis**
Presenter: **Ah Hin Pooi**, Sunway University, Malaysia
Consider a multiple linear regression model with n observations, p regression parameters, and independent random errors which have constant variance and an otherwise unknown distribution. The null hypothesis that the last k regression parameters are zero is tested by using a statistic constructed from the ranks of the last $n-p+k$ rotated observations obtained by pre-multiplying the vector of observations with an orthogonal matrix. A test based on ranks is used to test the significance of the parameters in a three-way classification model. The numerical results on the probability of the rejection region indicate that the test has a size which deviates from the target value 0.05 by less than 0.01 for varying values of the measures of skewness and kurtosis of the random errors. In the case of normal errors, the powers of the test are found to be only slightly smaller than those of the classical test.

EC0592: **Quasi-randomization tests of copula symmetry**
Presenter: **Juwon Seo**, NUS, Singapore
Co-authors: Brendan Beare
New nonparametric tests of bivariate copula exchangeability and radial symmetry are proposed. The tests are simple to compute, control size asymptotically, consistently detect arbitrary forms of asymmetry, and do not require the specification of a tuning parameter. Simulations indicate excellent small sample properties. The novel aspect of the tests is a resampling procedure that exploits group invariance conditions associated with the relevant symmetry hypothesis. They may be viewed as modified versions of exact randomization tests, the latter being inapplicable due to the unobservability of margins.

EC0420: **On the statistical inference for a class of observation-driven time series models for count data**
Presenter: **Yunwei Cui**, Towson University, United States
Co-authors: Qi Zheng
The purpose is to investigate the statistical inference for a class of observation-driven time series models of count data where the conditional distribution of the observed count given a state process is from the one-parameter exponential family. Under certain regularity conditions, the strong consistency and asymptotic normality of the maximum likelihood estimator are established.

EC293 Room LSK1001 CONTRIBUTIONS IN ECONOMETRICS MODELS Chair: Maria Kyriacou

EC0733: **Additive time-dependent hazard model with doubly truncated data**
Presenter: **Gordon Frank**, University of Rostock, Germany
Co-authors: Minwoo Chae, Yongdai Kim
For doubly truncated data, i.e. the variables of interest are only observable if they lie in a certain random interval, an additive hazard model with time-dependent regression coefficients is investigated. Consistency and asymptotic normality are proven under mild assumptions. A simulation study investigates the finite sample properties and the influence of the truncation distribution on the estimation error. Finally, the method is applied to a doubly truncated data set of German companies, where the age at insolvency is of interest.

EC0739: **Continuously updated indirect inference in SAR models with unobserved heterogeneity**
Presenter: **Maria Kyriacou**, University of Southampton, United Kingdom
Co-authors: Peter CB Phillips, Francesca Rossi
Spatial units are often heterogeneous as they vary in many of their observed characteristics such as income and so the assumption homoskedasticity may not hold. In the presence of unobserved heterogeneity the (Q)LE of both the spatial parameter and the exogenous regressors coefficients become, in general, inconsistent. There is an evident lack of estimation methods that account for the presence of hetersokedasticity while allowing for a wider class of heteroskedastic designs and also more realistic weight matrix designs. A Robust Generalized Methods of Moments (RGMM) estimator has been previously proposed which is consistent in heteroskeskadic situations. Also a GMM method robust to heteroskedasticity has been considered. Modifying the QMLE/MLE estimator has been proposed to restore consistency under mild forms of heteroskedasticity. There is yet a method that provides finite sample refinements for moderate sample sizes for general forms of hetersoskedasticy without being restrictive in the design of the exogenously given weights matrix. We propose an indirect inference based method robust to unobserved heterogeneity, the Continuously Updated Indirect Inference (CUII) which is derived using a binding function with continuously updated diagonal variance/covariance matrix. Simulation results reveal that our proposed CUII estimator is effective in reducing both bias and MSE compared to QML/ML and the RGMM estimator.

EC0662: **Efficient propensity score regression estimators of multivalued treatment effects for the treated**
Presenter: **Ying-Ying Lee**, University of California, Irvine, United States

Matching is a widely-used program evaluation estimation method when treatment is as- signed at random conditional on observable characteristics. When a multivalued treatment takes on more than two values, valid causal comparisons for a subpopulation who is treated a particular treatment level are based on two propensity scores one for the treated level and one for the counterfactual level. The main contribution is to provide propensity score regression estimators for a class of treatment effects for the treated that achieve the semiparametric efficiency bounds under the cases when the propensity scores are unknown and when they are known. We derive the large sample distribution that reveals how first step estimation of the propensity score as generated regressors affects asymptotic efficiency. We contribute to the binary treatment literature by a new propensity score regression estimator for the average/quantile treatment effect for the treated: the proposed efficient estimator matches on a normalized propensity score that is a combination of the true propensity score and its nonparametric estimate. Moreover, we formally show that the semiparametric efficiency bound is reduced by knowledge of the propensity scores for the treated levels, but is not affected by knowledge of the propensity score for the counterfactual level. A Monte-Carlo experiment supports our theoretical findings.

EC0790: **Modeling and optimization designs of SPRT-based control schemes for individual observations**
Presenter: **Chenglong Li**, Xián Jiaotong University & City University of Hong Kong, Hong Kong

In some practical situations, it may not be feasible to take samples larger than one, so control charts must be based on individual observations ($n = 1$) rather than samples of $n > 1$. The sequential probability ratio test (SPRT) chart is particularly relevant for individual observations and is very effective in terms of detection speed compared with other charts. The existing research conducts the design or optimization algorithm for a SPRT chart from a statistical point of view only. A Markov chain approach is used and an economic optimization model is proposed for the SPRT chart. The results based on an extensive performance comparison, indicate that under various process distributions and various process scenarios the SPRT chart uniformly outperforms some other competing charts on a cost basis. Further study reveals that the optimal designed SPRT chart is rather robust to most errors incurred in parameter estimation. A case study is provided to illustrate the application of the SPRT chart.

EG011 Room LSKG007 CONTRIBUTIONS IN MODELLING FINANCIAL AND INSURANCE RISKS Chair: Bertrand Maillet

EC0779: **Statistical analysis of the exchange rate of Bitcoin**
Presenter: **Stephen Chan**, University of Manchester, United Kingdom

Bitcoin, the first electronic payment system, is becoming a popular currency. We provide a statistical analysis of the log-returns of the exchange rate of Bitcoin versus the United States Dollar. Fifteen of the most popular parametric distributions in finance are fitted to the log-returns. The generalized hyperbolic distribution is shown to give the best fit. Predictions are given for future values of the exchange rate.

EC0767: **Forecasting time-varying conditional skewness with asymmetric Laplace distribution**
Presenter: **Qian Chen**, Peking University Shenzhen Campus, China

Higher moment risk attracts increasing interest recently, partly due to its ability of accounting for those asset risk that can not be explained by the second moment of asset returns, e.g. some researchers found there could be critical loss even during low volatility period. As the conditional skewness distinguishes negative and positive returns, as well as describes the corresponding distributional features, it naturally carries important information about asset risk. Compared to an exhaustive list of time-evolving mean and volatility models for the asset returns, there are relatively limited and indirect models for time-varying conditional skewness. A GARCH skewness model with asymmetric Laplace conditional return distribution is proposed to capture the time-varying and persistent features lying in the conditional skewness of financial return series. The model is realized with a number of real stock index returns. The empirical results find there is less persistence in the conditional skewness compared to the persistence of conditional volatility. Furthermore, the high negative conditional skewness during low volatility period could be of potential risk as it is easily be neglected by the investors and decision makers. The improved comprehension of the time-varying features of the skewness is supposed to offer implication for the investors, financial institutions as well as the policy makers in risk control and decision making.

EC0816: **Multivariate ordinal regression models: An analysis of corporate credit ratings**
Presenter: **Rainer Hirk**, Vienna University of Economics and Business, Austria
Co-authors: Laura Vana, Kurt Hornik

Correlated ordinal data typically arise from multiple measurements on a collection of subjects. Motivated by an application in credit risk, where multiple credit rating agencies assess the creditworthiness of a firm on an ordinal scale, we consider multivariate ordinal models with a latent variable specification and correlated error terms. Two different link functions are employed, by assuming a multivariate normal and a multivariate logistic distribution for the latent variables underlying the ordinal outcomes. Composite likelihood methods are applied for estimating the model parameters. We investigate how sensitive the pairwise likelihood estimates are to the number of subjects and to the presence of observations missing completely at random, and find that these estimates are robust for both link functions and reasonable sample size. The empirical application consists of an analysis of corporate credit ratings from the big three credit rating agencies (Standard & Poor's, Moody's and Fitch). Firm-level and stock price data for publicly traded US companies as well as an incomplete panel of issuer credit ratings are collected and analyzed to illustrate the proposed framework.

EC0852: **SIFI: On the Systemic Importance of Financial Institutions as determined by an extended CAPM with systemic risk**
Presenter: **Bertrand Maillet**, EMLyon Business School (CEFRA), France

The proposal is to test an extension of the traditional extended Capital Asset Pricing Model (three-factor CAPM), in which a factor is added represented by an Index of Systemic Risk Measures (ISRM), built thanks to a Sparse Principal Component Analysis (SPCA) of a large set of systemic risk measures. The empirical tests of the CAPM with Systemic risk (CAPMS) we run on the American market, show that the new systemic risk factor is highly significant when pricing assets. We lastly propose an original application of the CAPMS related to a new methodology for designating and ranking Systemically Important Financial Institutions (SIFI), based on ordered significant sensitivities to the ISRM: the more sensitive, the most important.

EC282 Room LSK1009 CONTRIBUTIONS IN COMPUTATIONAL AND NUMERICAL METHODS Chair: Florian Heiss

EC0820: **Robust diagnostics for the negative binomial regression model**
Presenter: **Tsung-Chi Cheng**, National Chengchi University, Taiwan

Modeling count variables is a common task in econometrics, social and medical sciences. The negative binomial (NB) regression model is one of the popular approaches to the fitting of overdispersed count data. However, outliers may have some effects on the maximum likelihood estimates of the regression coefficients for NB regression model. We apply the maximum trimming likelihood estimation to deal with outlier problem for the count regression model. Real data examples are used to illustrate the performance of the proposed approach.

EC0809: **Bellman-Harris multi-type decomposable branching processes as models of cancer evolution**
Presenter: **Maroussia Slavtchova-Bojkova**, Sofia University, Bulgaria

A novel approach based on the multi-type Bellman-Harris reducible branching process is proposed which describes dynamics of cell population consisting of one supercritical type of cells, called mutant type and several subcritical ones. The subcritical type cells could mutate and produce supercritical type cells which have stronger potential to reproduce. Based on this model, we present an algorithm of a numerical method to estimate: the probability of escaping extinction of the mutant type cells, the waiting time to the appearance of such mutant which will give the beginning of indefinite path of survival and the modified hazard rates, used to characterize the immediate risk of escaping extinction, i.e., the probability of producing a "successful" mutant given that it has not been produced yet. Our leading example will be the appearance of cancer cells after chemotherapy, which reduces the capacity of division of cancer cells and should lead to the destruction of tumors. However, sometimes mutations in these cells provide resistance to the therapy. The new type of cells has a higher reproduction rate and can escape extinction. We expect that the methodology could support decision making in personalized medicine.

EC0207: **Extremum estimators with quasi-simulated objective functions**
Presenter: **Florian Heiss**, Heinrich-Heine University Duesseldorf, Germany
Co-authors: Michael Griebel, Jens Oettershagen, Constantin Weiser

Flexible nonlinear models often imply analytically intractable likelihood functions or moment conditions. Important examples include multinomial or panel data models of discrete outcomes. The literature has suggested estimators based on Monte-Carlo simulation such as maximum simulated likelihood, the method of simulated scores or the method of simulated moments. The properties of these estimators are well understood. However, the Monte-Carlo simulation error translates into noisy and biased estimators and that it is computationally costly or sometimes prohibitive to sufficiently reduce the simulation error using a brute force approach with a large number of simulation draws. Variance reduction algorithms and quasi-Monte-Carlo simulation methods are often found to work well in empirical applications and simulation exercises in the sense that they substantially reduce the computational burden. So far, little theoretical results are known for these estimators, so many applied papers refer to the properties of simulation estimators instead. We show under which conditions quasi-simulation estimators are consistent and derive the speed of convergence in terms of the computational burden. We also show conditions under which these estimators are asymptotically normally distributed and efficient. We also demonstrate these findings and small-sample properties in a simulation exercise for the multinomial probit model.

EC0765: **Pros and cons of Laplace approximations in mixed hidden Markov models**
Presenter: **Geir Drage Berentsen**, University of Bergen, Norway

Hidden Markov models (HMMs) are a popular tool for modelling temporal data and has been successfully applied to many problems. Using HMMs for analysing multiple processes, such as longitudinal data, have not yet gained the same popularity. This is partly due to the computational complexity involved in adjusting for heterogeneity in the data, which typically involves numerical or Monte Carlo approximations of multiple integrals. Implications of different model formulations for mixed Hidden Markov models (mHMMs) and pros and cons of using Laplace approximations to obtain the likelihood of the data are discussed. Applications of the method to both real and simulated data will be presented.

Friday 16.06.2017 10:20 - 12:25 Parallel Session G – EcoSta2017

EO104 Room LSK1014 INFERENCE FOR CORRELATED DATA Chair: Samuel Mueller

EO0151: **Sparse pairwise likelihood estimation for multivariate longitudinal mixed models**
Presenter: **Francis Hui**, The Australian National University, Australia
Co-authors: Samuel Mueller, Alan Welsh
In most longitudinal studies, the question of interest often involves studying not one but multiple response variables. For instance, in the social sciences we may be interested in uncovering the personal and environmental factors affecting the mental health of individuals over time, where mental health is measured using a set of questionnaire items. To analyze such data, we can extend the standard univariate mixed model approach to handle multiple outcomes. Estimating such multivariate mixed models presents a considerable challenge however, let alone performing variable selection to uncover which covariates are important in driving each of the outcomes. Motivated by composite likelihood ideas, we propose a new method for fixed effects selection in multivariate mixed models called Approximate Pairwise Likelihood Estimation and Shrinkage (APLES). The approach works by constructing a quadratic approximation to each component of the pairwise likelihood function, and augmenting this with a penalty that encourages individual and group coefficient sparsity. We show how the full regularization path for the APLES estimator can be constructed efficiently, and present a data-driven approach to selecting the tuning parameter. Asymptotic properties of the APLES estimator are discussed, with simulations studies demonstrating it can outperform univariate approaches based on analyzing each outcome separately.

EO0254: **Time-varying proportional odds models for mega-analysis of clustered event times**
Presenter: **Tanya Garcia**, Texas A&M University, United States
Co-authors: Karen Marder, Yuanjia Wang
Mega-analysis, or the meta-analysis of individual patient data, is the gold standard for synthesizing data from multiple studies when individual level data are available. It borrows information across studies to attain reliable and more precise estimation and to analyze multiple outcomes. An important aspect in the mega-analysis of time-to-event data is estimating the distribution function while accounting for the data hierarchy and potential right-censoring. An often encountered challenge is when the time-to-event data are clustered as in biomedical studies where multiple, life-impacting events on each subject are often measured (e.g., first loss of cognitive ability, first lost of motor-function). The multiple events measured on the same subject lead to clusters of correlated event times, and the intraclass dependency needs to be properly modeled to ensure correct inference on regression parameters.

EO0343: **Semiparametric observation-driven models for time-series of count**
Presenter: **Thomas Fung**, Macquarie University, Australia
A semiparametric class of generalised linear models for integer-valued time series, that contains the integer-valued generalised autoregressive conditional heteroskedasticity (INGARCH) framework as special case, was considered. Instead of specifying conditional distributions for the responses, the underlying response distribution is treated as an infinite-dimensional parameter which can be estimated simultaneously with the usual finite-dimensional parameters via a maximum empirical likelihood approach. Simulations suggests that both estimation and inferences using the approach can perform as well as correctly-specified parametric models, but can be more robust than parametric methods under model misspecification.

EO0500: **Inference on growth curves using monotone polynomials**
Presenter: **Berwin Turlach**, The University of Western Australia, Australia
Co-authors: Andrew Manderson, Edward Cripps, Kevin Murray
The problem of fitting monotone polynomials in a Bayesian framework using the probabilistic programming language Stan is discussed. The fitting of such polynomials to data is discussed in the context of growth curves. Considering the growth curves of several individuals simultaneously leads to the extension of our method to a longitudinal mixed model setting. We will examine this extension to hierarchical modeling in our Bayesian framework and illustrate how various question of interest about growth curves can be effectively addressed in this framework.

EO0237: **Modelling the conditional distribution of durations via mixture distributions**
Presenter: **Jennifer Chan**, The University of Sydney, Australia
Co-authors: Rasika Yatigammana
Conventionally, positive valued financial time series such as trade duration are generally unimodal, positive skewed with excessive near-zero values occurring during period of active trading from optimistic traders. Drawing from financial theory on market microstructure that divides the traders into distinct categories, primarily based on access to information as informed and uninformed traders, presumably the distribution of durations is derived from a mixture of distributions. We adopt this mixture approach using mixtures of exponential and generalized Beta type II distribution with Weibull as a special case to capture the distinct empirical traits including the high proportion of near-zero durations. In addition, the flexible and often heavier tails capture extreme outliers. The model performance is evaluated based on predictive log likelihood of the forecasts. The tail-risk measures such at time-at-risk (TaR) and conditional TaR are assessed via several criteria such as violation rates, quantile loss function and mean and maximum absolute deviations.

EO218 Room LSK1001 ECONOMETRIC METHODS FOR MACROECONOMIC ANALYSIS AND FORECASTING Chair: Yohei Yamamoto

EC0182: **Estimation and inference of structural changes in high dimensional factor models**
Presenter: **Xu Han**, City University of Hong Kong, Hong Kong
Co-authors: Jushan Bai, Yutang Shi
The estimation of a break point in high dimensional factor models is considered where the unobserved factors are estimated by Principal Component Analysis (PCA). The factor loading matrix is assumed to have a structural break at a certain time. We establish the conditions under which the least squares (LS) estimator is consistent for the break date. Our consistency result holds for smaller breaks than those in the existing literature. We also find the LS estimators asymptotic distribution, which depends on the data generating process of the unobserved factors. Simulation results confirm that the break date can be accurately estimated by the LS even if the breaks are small. In two empirical applications, we implement our method to estimate the break points in the U.S. stock market and U.S. macroeconomy, respectively.

EO0313: **Some properties of the modified CUSUM Tests**
Presenter: **Peiyun Jiang**, Hitotsubashi University, China
Co-authors: Eiji Kurozumi
The CUSUM tests played an important role in theory and applications related to structural change. There is a well-known result that the power of the CUSUM test in linear regression models crucially depends on the angle between the mean regressor and the structural shifts. A major drawback, as noted in the literature, is that the tests will lose power when the structural break is orthogonal to the mean regressor. We propose two modified

CUSUM tests to detect such structural changes and derive the limiting properties of the tests in a static regression with or without serially correlated errors. Then we investigate their power for parameter stability and demonstrate that the resulting tests are superior to the standard procedure when detecting the orthogonal structural shifts.

EO0436: **Evaluating the inefficiency of CBOs budgetary forecasts**
Presenter: **Natsuki Arai**, National Chengchi University, Taiwan
The aim is to evaluate the efficiency of the Congressional Budget Office's (CBO) budgetary and economic forecasts and to investigate the causes of the inefficiency in its forecasts. While the efficiency is accepted for the CBO's forecasts for outlays, it is strongly rejected for its revenue and economic forecasts. The conditional forecast evaluation suggests that efficiency is accepted once we control the differences in the underlying economic forecasts. By replacing the CBO's underlying economic forecasts with another subjective forecast in an out-of-sample forecasting exercise, we find that this adjustment could achieve modest improvements in the accuracy of the CBO's budgetary forecasts. Though the CBO tends to make positive forecast errors for the budget deficit, there is no definitive evidence for the asymmetry in its loss function.

EO0513: **An alternative estimation of a time-varying parameter model**
Presenter: **Tatsuma Wada**, Keio University, Japan
Co-authors: Mikio Ito, Akihiko Noda
A non-Bayesian, generalized least squares (GLS) based approach is proposed to estimate a class of time-varying AR parameter models. This approach is proven very efficient because unlike conventional methods, it does not require the Kalman filtering and smoothing procedures, but yields the smoothed estimate that is identical to the Kalman smoothed estimate. Since the GLS based approach can compute the smoothed estimate at once, the time required to compute the smoothed estimate is shorter than the time required by alternative approaches. In addition, this approach enables us to deal with stochastic volatility models and models having a time-dependent variance-covariance matrix. Other features, such as the possibility of the pile-up problem are assessed through simulations.

EO0493: **Determinants of world business cycles: Some insights from a flexible dynamic factor model**
Presenter: **Peng Wang**, Hong Kong University of Science and Technology, Hong Kong
Co-authors: Ka Yui Leung
Using a dynamic factor model, we characterize the business cycles co-movement among a panel of countries using a world factor, a country-specific factor, and an idiosyncratic component. We allow correlated factors and then study the contribution of each type of factor to the business cycle dynamics. We find that three world factors are necessary to characterize the co-movement on the world level. Our analysis shows that the three world factors correspond to the price dynamics, the consumption dynamics, and the GDP growth dynamics respectively. We then compare the importance of the world factor with that of the spill-over effects from the country-specific factor. Then we study how these factors are related to country shocks as well as other economic variables. Our data set covers 23 countries, which include the G7, Four Little Dragons (except Taiwan), Four Little Tigers (except Indonesia), and large developing economies such as India and China, covering more than 75% of world total output.

EO128 Room LSK1010 NEW DEVELOPMENTS IN SURVIVAL ANALYSIS Chair: Xingqiu Zhao

EO0186: **Inference for net survival models with application to cancer data**
Presenter: **Laurent Bordes**, University of Pau, France
Co-authors: Olayide Boussari, Valerie Jooste
A general methodology is proposed for testing the null hypothesis that an excess hazard rate model, with or without covariates, belongs to a parametric family. Estimating the excess hazard rate function parametrically through the maximum likelihood method and non-parametrically (or semi-parametrically) we build a discrepancy process which is shown to be asymptotically Gaussian under the null hypothesis. Based on this result, we are able to build some statistical tests in order to decide whether or not the null hypothesis is acceptable. We illustrate our results by the construction of chi-square tests. Their properties are studied through a Monte-Carlo study, then the testing procedure is applied to colon cancer data provided by population based specialized cancer registry: Registre Bourguignon des Cancers Digestifs.

EO0469: **A unified semi-empirical likelihood ratio confidence interval for treatment effects based on length-biased data**
Presenter: **Tao Li**, Shanghai University of Finance and Economics, China
In two sample studies, the treatment effects that we are interested in may have different types, such as mean difference, the difference of probabilities, and etc. We apply semi-parametric empirical likelihood principle to length biased data and derived a unified empirical likelihood ratio confidence interval for various types of treatment effects. The empirical likelihood ratio is shown to be asymptotically distributed as chi-squared. Both simulation study and real data analysis are conducted and illustrate that the confidence interval derived from empirical likelihood ratio is better than its counterpart from the estimating equation.

EO0584: **On subgroup-specific treatment effects analysis**
Presenter: **Wenxin Liu**, The Chinese University of Hong Kong, Hong Kong
Co-authors: Yuanyuan Lin
Identifying subgroups of patients with similar features is important in individualized treatment strategies. A general heterogeneous model with coefficients of treatment effects that are subject-dependent is considered. We propose a method to estimate the number of subgroups, identify the subgroups and estimate the underlying coefficients of treatment effects within each subgroup. The oracle estimator of subgroup coefficients with the subgroup structure known is showed to be consistent, approximately normal and a local minimizer of the objective function with a high probability. Besides, simulation studies and applications are presented to evaluate our method.

EO0472: **Sampling and regression methods for the Cox model**
Presenter: **Yuan Yao**, Hong Kong Baptist University, Hong Kong
A new retrospective sampling design is considered that improves the efficiency of the case-cohort and case-control designs. The regression analysis is conducted using the Cox model. The maximum likelihood approach and the inverse probability weighting approach are developed to estimate the regression parameters. The resulting estimators are proved to be consistent and asymptotically normal. Simulation and real data studies are presented.

EO0561: **Nonparametric statistical inference for case one interval censored data**
Presenter: **Kin Yat Liu**, The Hong Kong Polytechnic University, Hong Kong
Co-authors: Meiling Hao, Xingqiu Zhao, Yuanyuan Lin
Nonparametric maximum likelihood has been derived for the survival functions with current status data. The nonparametric maximum likelihood estimators of current status data, the estimator's self-consistency property and the confidence intervals have been studied. However, as the estimator is discrete, it is not suitable to study the density and the hazard. As hazard can give more insight about the event of interest than the survival function, numerous articles have been developed based on the hazard. Similar to the case of right censored data, kernel-based approaches were given. In order to avoid selecting the sensitive bandwidth for the estimation, spline methods have also been extended to the interval censored data. Methods

with the splines did not consider the asymptotic properties of the estimators. We use the penalized likelihood method to get the estimator of the cumulative hazard function. A functional Bahadur representation will be established. Based on the technical tool, we show the estimator enjoys the pointwise asymptotic normality and the global asymptotic Gaussianity. Furthermore, the likelihood ratio test is shown, which reveals some efficient properties of the test.

EO240 Room LSK1027 RECENT ADVANCES IN MIXTURE MODELS AND LATENT VARIABLE MODELS Chair: Chi Tim Ng

EO0194: **Robust deviation information criterion for latent variable models**
Presenter: **Yong Li**, Renmin university of China, China

The data augmentation technique is shown to undermine the theoretical underpinnings of the deviance information criterion (DIC), a widely used information criterion for Bayesian model comparison, although it facilitates parameter estimation for latent variable models via Markov chain Monte Carlo (MCMC) simulation. Data augmentation makes the likelihood function non-regular and hence invalidates the standard asymptotic arguments. A new information criterion, robust DIC (RDIC), is proposed for Bayesian comparison of latent variable models. RDIC is shown to be a good approximation to DIC without data augmentation. While the later quantity is difficult to compute, the expectation maximization (EM) algorithm facilitates the computation of RDIC when the MCMC output is available. Moreover, RDIC is robust to nonlinear transformations of latent variables and distributional representations of model specification. The proposed approach is illustrated using several popular models in economics and finance.

EO0333: **Model-based peak alignment of metabolomic profiling from comprehensive two-dimensional GCxGC/TOF-MS**
Presenter: **Jaesik Jeong**, Chonnam National University, Korea, South

Comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry (GCxGC/TOF-MS) has been used for metabolite profiling in metabolomics. However, there is still much experimental variation to be controlled including both within-experiment and between-experiment variation. For efficient analysis, an ideal peak alignment method to deal with such variations is in great need. Using experimental data of a mixture of metabolite standards, we demonstrated that our method has better performance than other existing method which is not model-based. We then applied our method to the data generated from the plasma of a rat, which also demonstrates good performance of our model. Summarizing, we developed a model-based peak alignment method to process both homogeneous and heterogeneous experimental data. The unique feature of our method is the only model-based peak alignment method coupled with metabolite identification in an unified framework. Through the comparison with other existing method, we demonstrated that our method has better performance.

EO0441: **A new understanding of LASSO in the presence of measurement errors**
Presenter: **Woojoo Lee**, Inha University, Korea, South
Co-authors: Chi Tim Ng, Youngjo Lee

LASSO has been criticized for selecting too many covariates. We show that the LASSO method achieves an optimal prediction when the covariates are subjected to measurement errors. Up to now, all criticisms assume that the covariates are observed without measurement errors, which is not likely to be true in many practical situations. Under measurement errors, the meaning of relevant covariates can be ambiguous. In such a situation, we illustrate that some covariates without an association with the response can potentially be relevant. Neglecting such covariates results in bias in prediction and misinterpretation. To understand the subset of covariates that should be included, a factor model of the covariates is introduced. Two real data examples are provided to illustrate the behaviors of the LASSO method under measurement errors.

EO0701: **High dimensional LASSO variable selection for correlated covariates**
Presenter: **Kaimeng Zhang**, Chonnam National University, Korea, South
Co-authors: Chi Tim Ng

A regression model is considered that regress the response against the idiosyncratic factors obtained from the factor analysis of the covariates. Such a model is particularly useful in the high-dimensional variable selection problems in certain econometrics applications where all covariates are correlated due to systematic economic factors. In such cases, it is shown both theoretically and empirically that the usual penalized regression of the response against the covariates tends to select either all or none of the covariates. On the contrary, the proposed hybrid approach of factor analysis and penalized regression can select relevant covariates consistently under $p = O(e^n)$ and other mild conditions on the factor loading matrix, where p and n are the number of covariates and the sample size respectively. To illustrate the ideas, two empirical data analysis examples are considered, (i) the gross domestic production of a chosen country against capital inputs and labor inputs of all countries and (ii) the stock returns of a chosen stock against the trading volumes of all stocks in the financial market.

EO0705: **Regression diagnostic in Hurdle-generalized Poisson regression models**
Presenter: **Guo Xin Zuo**, Central China Normal University, China
Co-authors: Li Wei Zhang

Count data with extra zeros encountered usually in Biology, Medicine, and Economics study. Parametric inference is discussed for Hurdle-generalized Poisson regression model with count data. Then, statistical diagnostics are developed based on the case-deletion method and local influence analysis, and an approximate score test is obtained for significance of the dispersion parameter in the model. Finally, simulation studies are given to illustrate the methodology.

EO158 Room LSK1033 INSURANCE MODELS WITH DEPENDENCE Chair: KC Yuen

EO0228: **Multivariate countermonotonicity and the minimal copulas**
Presenter: **Ka Chun Cheung**, The University of Hong Kong, Hong Kong
Co-authors: Woojoo Lee, Jae Youn Ahn

Frechet-Hoeffding upper and lower bounds play an important role in various bivariate optimization problems because they are the maximum and minimum of bivariate copulas in concordance order, respectively. However, while the Frechet-Hoeffding upper bound is the maximum of any multivariate copulas, there is no minimum copula available for dimensions higher than 3. Therefore, multivariate minimization problems with respect to a copula are not straightforward as the corresponding maximization problems. When the minimum copula is absent, minimal copulas are useful for multivariate minimization problems. We illustrate the motivation of generalizing the joint mixability to d-countermonotonicity through variance minimization problems and show that d-countermonotonic copulas are minimal copulas.

EO0222: **Correlated default models driven by a multivariate regime-switching shot noise process**
Presenter: **Yinghui Dong**, Suzhou University of Science and Technology, China

A reduced-form credit risk model with regime-switching intensities is developed to investigate the pricing of a CDS contract. We assume that the defaults of all the names are driven by some shock events. The arrivals of the shock events and the interest rate are modeled by a multivariate regime-switching shot noise process. We provide the flexibility that the model parameters, including the intensities and the jump sizes of the jump component can switch over time according to a continuous-time Markov chain. The states of the chain may be interpreted as different states of an

economy or different stages of a business cycle. Based on the joint Laplace transform of the regime-switching shot noise processes, we derive the explicit formulas for the spreads of CDS contract with and without counterparty risk.

EO0552: **Some optimality results for insurers under mean-variance criterion**
Presenter: **Junyi Guo**, Nankai University, China
Some recent results on optimal investment and optimal reinsurance problems will be presented. Especially, we consider the optimal reinsurance problem with stochastic volatility of stocks and jump-diffusion financial market under the mean-variance criterion. By solving a backward stochastic differential equation or HJB equation , the closed-form expressions of the efficient frontiers and efficient strategies are obtained and the economic behavior of the efficient frontiers are analyzed numerically.

EO0604: **On the effect of reinsurance on an insurer: A cost-benefit analysis incorporating default risk**
Presenter: **Ambrose Lo**, University of Iowa, United States
Reinsurance is often empirically hailed as a value-adding risk management strategy which an insurer can utilize to achieve various business objectives. In the context of a distortion-risk-measure-based three-party model incorporating a policyholder, insurer and reinsurer, we formulate explicitly the optimal insurance-reinsurance strategies from the perspective of the insurer. Our analytic solutions are complemented by intuitive but scientifically rigorous explanations on the marginal cost and benefit considerations underlying the optimal insurance-reinsurance decisions. These cost-benefit discussions not only cast light on the economic motivations for an insurer to engage in insurance with the policyholder and in reinsurance with the reinsurer, but also mathematically formalize the value created by reinsurance with respect to stabilizing the loss portfolio and enlarging the underwriting capacity of an insurer. Our model also allows for the reinsurer's failure to deliver on its promised indemnity when the regulatory capital of the reinsurer is depleted by the reinsured loss. The reduction in the benefits of reinsurance to the insurer as a result of the reinsurers default is quantified, and its influence on the optimal insurance-reinsurance policies analyzed.

EO0447: **Optimal dividends and reinsurance for a risk model with dependence**
Presenter: **KC Yuen**, HKU, China
The variance premium principle is adopted to investigate the problem of optimal dividends and reinsurance in a diffusion approximation risk model with thinning-dependence structure. We first study the optimal problem without capital injection. We then consider the incorporation of forced capital injection into the model whenever the reserve level drops below zero. We finally turn to the general problem in which capital injection is allowed but not compulsory. For the three optimal problems, we apply the technique of stochastic control theory to obtain closed-form expressions for the optimal strategies and the corresponding value functions for two classes of insurance business with thinning dependence. We also present some numerical examples to show the effect of parameter values on the optimal policies.

EO052 Room LSK1007 RECENT ADVANCES IN HIGH DIMENSIONAL STATISTICAL INFERENCE Chair: Ping-Shou Zhong

EO0262: **Scalable interpretable multi-response regression via SEED**
Presenter: **Zemin Zheng**, University of Science and Technology of China, China
Sparse reduced-rank regression is an important tool to uncover meaningful dependence structure between large numbers of predictors and responses in many big data applications such as genome-wide association studies and social media analysis. Despite the recent theoretical and algorithmic advances, scalable estimation of sparse reduced-rank regression remains largely unexplored. We suggest a scalable procedure called sequential estimation with eigen-decomposition (SEED) which needs only a single top-r singular value decomposition to find the optimal low-rank and sparse matrix by solving a sparse generalized eigenvalue problem. Our suggested method is not only scalable but also performs simultaneous dimensionality reduction and variable selection. Under some mild regularity conditions, we show that SEED enjoys nice sampling properties including consistency in estimation, rank selection, prediction, and model selection. Numerical studies on synthetic and real data sets show that SEED outperforms the state-of-the-art approaches for large-scale matrix estimation problem.

EO0553: **Optimal uniform change point tests in high-dimension**
Presenter: **Moritz Jirak**, TU Braunschweig, Germany
Consider d dependent change point tests, each based on a CUSUM-statistic. We provide an asymptotic theory that allows us to deal with the maximum over all test statistics as both the sample size n and d tend to infinity. We achieve this either by a consistent bootstrap or an appropriate limit distribution. This allows for the construction of minimax optimal tests. In case of nested factor models these tests are even adaptive with respect to the spatial dependence structure, and the corresponding minimax rate can be expressed in terms of underlying quantiles.

EO0528: **Spectral analysis of high-dimensional linear processes with applications**
Presenter: **Debashis Paul**, University of California, Davis, United States
Results are presented for the limiting behavior of the empirical distribution of eigenvalues of a weighted integral of the sample periodogram for a class of high-dimensional linear processes. The processes under consideration are characterized by having simultaneously diagonalizable coefficient matrices. We make use of these asymptotic results, derived under the setting where the dimension and sample size are comparable, to formulate an estimation strategy for the distribution of eigenvalues of the coefficients of the linear process. This approach generalizes existing works on estimation of the spectrum of an unknown covariance matrix for high-dimensional i.i.d. observations. We discuss various applications of the proposed methodology including in the context of estimation of mean variance frontier in the Markowitz portfolio optimization problem.

EO0812: **CLT for the k largest eigenvalues and unit root tests for high dimensional nonstationary time series**
Presenter: **Guangming Pan**, Nanyang Technological University, Singapore
The focus is on both the convergence in probability and the asymptotic joint distribution of the first k largest eigenvalues of sample covariance matrices when data are nonstationary. As an application, a new unit root test for a vector of high-dimensional time series is proposed, and then studied both theoretically and numerically.

EO0523: **Statistical inference for large precision matrices with applications to brain connectivity**
Presenter: **Yumou Qiu**, University of Nebraska Lincoln, United States
Motivated by the importance of precision matrix in many practical applications, statistical inference is considered for precision matrices with ultra high-dimensional time dependent observations. We propose a novel data-driven procedure to construct confidence regions for the precision matrix with application in hypothesis testing and support recovery. Comparing with the existing methods, the proposed procedure imposes weaker structural assumptions on the observed data, which broadens the application scope of this new method. A computationally efficient algorithm is developed to implement the proposed procedure. Applications to brain connectivity by FMRI data is also discussed.

EO234 Room LSKG003 NEW CHALLENGES IN COMPLEX DATA ANALYSIS Chair: Yichuan Zhao

EO0271: **Cusp catastrophe linear regression model and its applications**
Presenter: **Din Chen**, University of North Carolina, United States
Co-authors: Xinguang Chen
A novel cusp catastrophe linear regression model is introduced which can be used to model both rational and irrational behaviors simultaneously in social and health behavior research. We propose a maximum likelihood estimation to estimate the model parameters and their associated statistical inference. Simulation studies will be used to validate this novel model and the estimation procedure. A real example is used to illustrate this new methods.

EO0335: **Concordance measure-based feature screening and variable selection**
Presenter: **Yunbei Ma**, Southwestern University of Finance and Economics, China
The C-statistic, measuring the rank concordance between predictors and outcomes, has become a standard metric of predictive accuracy and is therefore a natural criterion for variable screening and selection. However, as the C-statistic is a step function, its optimization requires brute-force search, prohibiting its direct usage in the presence of high-dimensional predictors. We develop a smoothed version of the C-statistic to facilitate variable screening and selection. Specifically, we propose a smoothed C-statistic sure screening (C-SS) method for screening ultrahigh-dimensional data, and a penalized C-statistic (PSC) variable selection method for regularized modeling based on the screening results. We have shown that these two coherent procedures form an integrated framework for screening and variable selection: the C-SS possesses the sure screening property, and the PSC possesses the oracle property. Specifically, the PSC achieves the oracle property if $m_n = O(n^{1/4})$, where is the cardinality of the set of predictors captured by the C-SS. Our extensive simulations reveal that, compared to existing procedures, our proposal is more robust and efficient. Our procedure has been applied to analyze a multiple myeloma study, and has identified several novel genes that can predict patients response to treatment.

EO0805: **Nonparametric estimation in a multi-armed bandit problem with covariates**
Presenter: **Wei Qian**, University of Delaware, United States
Co-authors: Yuhong Yang
The multi-armed bandit problem is a popular online decision-making problem with a wide range of modern applications in data science. Its classical setting consists of multiple slot machines with unknown expected payoffs, and the goal of the optimization game is to design a sequential arm allocation algorithm to maximize the total payoff. Motivated by promising applications in personalized medical and online web service, we consider a setting where the mean rewards of bandit machines are associated with covariates. With the key tradeoff between exploring new information and exploiting history information, we propose a kernel estimation based sequential allocation algorithm, and investigate its finite-time optimality under a nonparametric framework. In addition, since many nonparametric and parametric methods in supervised learning may be applied to estimating the mean reward functions, we integrate a model combining strategy into the allocation algorithm for adaptive performance. Simulations and real data evaluation are conducted to illustrate the algorithm performance and support the necessary consideration of covariates.

EO0457: **Combing multiple adherence measures in HIV prevention trials**
Presenter: **Yifan Zhu**, Fred Hutchinson Cancer Research Center, United States
Co-authors: Takumi Saegusa, Chong-Zhi Di, Ying Qing Chen
Good adherence, in addition to drug efficacy, social interaction and behavior, is important for the effectiveness of HIV pre-exposure prophylaxis(PrEP). Multiple adherence measures were collected in HPTN 069, a phase II randomized, double-blind PrEP study in men who have sex with men and at-risk women. These included self-reports via computer-assisted self-interview, drug concentration measurements in blood sample, and Wisepill, an electronic drug monitoring device. Wisepill signals were reported daily, self-reports were conducted every two months and drug concentration tests were performed least frequently. Both parametric and non-parametric procedures were performed for clustering subject adherence pattern from Wisepill. We found stronger correlation between Wisepill and drug concentration than correlation between self-report and drug concentration. We also fitted latent class models to combine the adherence measurements. Regression analysis indicated several predictive covariates of adherence pattern.

EO0337: **New nonparametric procedures for partial areas under ROC curves**
Presenter: **Yichuan Zhao**, Georgia State University, United States
The receiver operating characteristic (ROC) curve is a well-known measure of the performance of a classification method. Interest may only pertain toa specific region of the curve and, in this case, the partial area under the ROC curve (pAUC) provides a useful summary measure. Related measures such as the ordinal dominance curve (ODC) and the partial area under the ODC (pODC) are frequently of interest as well. Based on a novel estimator of pAUC, we develop nonparametric approaches to the pAUC and pODC using normal approximation, the jackknife and the jackknife empirical likelihood. A simulation study demonstrates the flaws of the existing method and shows proposed methods perform well. Simulations also substantiate the consistency of our jackknife variance estimator. The Pancreatic Cancer Serum Biomarker data set is used to illustrate the proposed methods.

EO140 Room LSK1032 FINANCIAL VOLATILITY Chair: Toshiaki Watanabe

EO0322: **Tail risk premium and predicting credit spreads**
Presenter: **Masato Ubukata**, Kushiro Public University of Economics, Japan
The aim is to investigate the role of time-varying jump tail risk component of market variance risk premium for predicting credit spreads in US and Japanese corporate bond markets. Based on a recent model-free estimation procedure, we find that the jump risk premium could strongly predict lower-rated credit spreads and default spreads in Japan, even with control for traditional predictors and lagged credit spreads. The predictive pattern on forecasting horizons ranging from one month to one year differs from that of diffusive component of variance risk premium, and thus the variance risk premium decomposition increases the forecasting power of standard predictability regressions. Unlike in Japan, the jump risk premium might be a weaker predictor for US credit spreads, because it becomes insignificant after controlling volatility measures and lagged credit spreads.

EO0325: **Prediction, filtering, and smoothing for the integrated variance with intraday returns**
Presenter: **Daisuke Nagakura**, Keio University, Japan
A state space method is proposed to calculate the linear projection of the integrated variance (IV) on squares of high frequency returns in the presence of market microstructure noise (MMN). Unlike most of the previous literature, we allow MMNs to be serially correlated. We also incorporate the leverage effects into the model. The method is based on a state space representation of squares of high frequency returns, in which the IV is one of the state variables. Our simple numerical experiment shows that our method performs better than other existing state space methods, in particular, in terms of forecasting.

EO0499: **Bayesian emulation for high-dimensional portfolio problem**
Presenter: **Kaoru Irie**, University of Tokyo, Japan
The portfolio problem to select a small number of promising financial assets from all the assets available in the market is discussed. Based on the equivalence between the Bayesian expected utility optimization and the posterior analysis of synthetic statistical models, we "mode" the portfolio decision problem as the state space models, introducing the priors with different levels of shrinkage effects, such as Laplace, horseshoe and their combination, for the purpose of portfolio selection. This approach also allows for the use of computational methodology in statistics customized for the posterior mode, including the EM methods and other stochastic model search technique. The advantage of these decision models are exemplified by the sequential portfolio analysis of Nikkei 225 stocks, by alternatively computing the multiple step ahead predictions provided by the multivariate volatility models and making the portfolio decision.

EO0533: **Frequency wise decomposition of variance risk premium**
Presenter: **Kosuke Oya**, Osaka University, Japan
The predictability of variance risk premium (VRP) has attracted much attention after the confirmation of its ability for the equity market in the United State. On the other hand, the lack of predictability is documented for several countries, including Japan. The empirical evidences shown in previous studies are based on the overlapping long-horizon return regression with VRP as the predictor variable. They conclude that the US equity market return is predictable over relatively short 3- to 5-month horizon. The times series of VRP and equity market return exhibit non-negligible temporal variations and the variations often tend to be large. Then the interrelationship between two series is hard to capture in time domain. Although the regression analyses used in previous studies are conducted in time domain, the analysis in frequency domain is more suitable to gauge in what horizon the VRP has the most significant predictive power. We introduce the frequency wise decompositions of VRP and equity return and examine in which frequency band the VRP has the largest causality to the equity return. We found that the hump shaped causality of VRP to equity return in frequency domain and the largest causality is attained in one month horizon.

EO0556: **Realized variance driven by order flow imbalance**
Presenter: **Makoto Takahashi**, Osaka University, Japan
The order flow imbalance (OFI), defined as the imbalance between supply and demand at the best bid and ask prices, is a main driver of price changes over a short interval and potentially a signal of private information. Based on a structural vector autoregressive model with identification through heteroskedasticity, a realized variance (RV) is decomposed into OFI-correlated and -uncorrelated components. The OFI-correlated component is constructed from returns driven by OFI innovations, that is, it represents the price impact of OFIs. The ratio of this component to the total RV is a relative informativeness of OFIs with respect to total public information and can be viewed as a measure of illiquidity. For E-mini S&P 500 futures, the OFI-correlated component increases with the total RV but its ratio decreases when the total RV increases. Additionally, OFIs are slightly less informative when major macroeconomic announcements are released. Moreover, heterogeneous autoregression results show that the decomposition improves the in-sample forecast of volatility.

EO042 Room LSKG007 LARGE-SCALE, NON-ELLIPTIC PORTFOLIO OPTIMIZATION Chair: Marc Paolella

EO0332: **The univariate collapsing method for portfolio optimization**
Presenter: **Marc Paolella**, University of Zurich, Switzerland
The univariate collapsing method (UCM) for portfolio optimization is based on obtaining the predictive mean and a risk measure such as variance or expected shortfall of the univariate pseudo-return series generated from a given set of portfolio weights and multivariate set of assets under interest and, via simulation or optimization, repeating this process until the desired portfolio weight vector is obtained. The UCM is well-known conceptually, straightforward to implement, and possesses several advantages over use of multivariate models, but, among other things, has been criticized for being too slow. As such, it does not play prominently in asset allocation and receives little attention in the academic literature. Fast model estimation methods combined with new heuristics for sampling are proposed to be used, based on easily-determined characteristics of the data, to accelerate and optimize the simulation search. An extensive empirical analysis confirms the viability of the method.

EO0597: **Getting out of the COMFORT zone: The MEXI distribution for asset returns**
Presenter: **Jeffrey Naef**, University of Zurich, Switzerland
Co-authors: Marc Paolella, Ronald W Butler, Pawel Polak
A new class of multivariate distributions with an elegant probabilistic structure is proposed that nests numerous models currently in use for modeling financial asset returns, including the recently proposed so-called COMFORT model. The primary innovation is that each margin can have its own tail shape parameter. Use of models such that each margin has the same tail index can result in non-optimal asset allocation and have potentially detrimental implications for risk management. The new model supports use of time-varying conditional heteroskedasticity and asymmetry. Fast, reliable estimation methods of some of the new structures are presented. An empirical exercise with real data demonstrates the benefit to asset allocation in terms of total wealth and Sharpe ratio performance.

EO0536: **A flexible regime switching model for asset returns**
Presenter: **Patrick Walker**, University of Zurich, Switzerland
Co-authors: Marc Paolella, Pawel Polak
A Markov regime switching dynamic correlation (RSDC) model for a multivariate set of asset returns is proposed. The univariate series are endowed with the usual GARCH structure, and the underlying innovations are multivariate generalized hyperbolic distribution (MGHyp). The multivariate conditional predictive distribution is MGHyp, hence weighted sums of marginals are themselves GHyp and thus tractable, enabling, e.g., portfolio optimization. To accomplish joint likelihood estimation of all the model parameters, a new, fast and efficient two-stage EM-algorithm is developed for estimation. This is coupled with shrinkage estimation of the correlation matrices via a quasi-Bayesian prior, enhancing both estimation ease and forecast quality. Based on several data sets of stock returns and foreign exchange rates, the new model is demonstrated to outperform all special cases in terms of in-sample fit and out-of-sample density forecasts. An application to portfolio optimization shows the importance of dynamical correlations for optimal asset allocation by providing consistently higher Sharpe ratios for all RSDC models, compared to their CCC counterparts.

EO0671: **Optimal shrinkage and selection methods**
Presenter: **Pawel Polak**, Columbia University, United States
An estimator is introduced for a multiple linear regression model with asymptotically optimal shrinkage and selection penalty. Optimality is meant with respect to a quadratic loss function, asymptotically as the number of observations and the number of variables go to infinity together. The forecasting performance of the new estimator is demonstrated by simulations and in large-scale portfolio application.

EO0806: **MCD meets MCD: Minimum Covariance Determinant Methods with Mixed Correlation Dynamics**
Presenter: **Marco Gambacciani**, University of Zurich and Swiss Financial Institute, Switzerland
Co-authors: Marc Paolella
A new model class for large-scale multivariate financial asset returns is proposed and studied. It expands upon recent work using the multivariate

normal mixture model, estimated using the minimum covariance determinant (MCD), in two ways. First, the multivariate Laplace is used in place of the normal, and its benefit is shown both in terms of density forecasting and portfolio performance. Second, the model is extended to support a DCC-GARCH type of structure, whereby both components in the mixture are endowed with a DCC law of motion for the two time-varying covariance matrices. This is accomplished by using the MCD fit to generate posterior probabilities associated with each component, at each point in time.

EO280 Room LSK1003 MODELLING FINANCIAL MARKET DYNAMICS Chair: Ruijun Bu

EO0349: **Modelling mortality dynamics with nonlinear income effect**
Presenter: **Yongok Choi**, Chung-Ang University, Korea, South
A nonparametric panel method is introduced to model dynamics of mortality rates. Age-specific mortality rates are separately modelled as a function of time and income-level. Income has different nonlinear effect on mortality improvements by age. Especially, mortality improvement for the elderly accelerates along with income only up to certain level. Also, we find that country-specific mortality rates by age decrease nonlinearly with respect time, but income level-specific mortality rates by age improve linearly along with time. This property can be used in mortality forecasting model.

EO0165: **Testing for normal copulas**
Presenter: **Dante Amengual**, CEMFI, Spain
Co-authors: Enrique Sentana
Computationally simple and intuitive expressions are derived for score tests of Gaussian copulas against Generalised Hyperbolic alternatives, including symmetric and asymmetric Student t, and many other distributions. We decompose our tests into third and fourth moment components, and obtain one-sided Likelihood Ratio analogues, whose asymptotic distribution we provide. We conduct Monte Carlo exercises to assess the finite sample properties of asymptotic and bootstrap versions of our tests. In an empirical application to CRSP stocks, we find that short-term reversals and momentum effects are better captured by non-Gaussian copulas. We estimate their parameters by indirect inference, and devise successful trading strategies.

EO0171: **Oil price volatility and macroeconomic fundamentals: A regime switching GARCH-MIDAS model**
Presenter: **Zhiyuan Pan**, Southwestern University of Finance and Economics, China
Co-authors: Yudong Wang, Li Liu
A regime switching GARCH-MIDAS model is introduced to investigate the relationships between oil price volatility and its macroeconomic fundamentals. Our model takes into account both effects of long-term macroeconomic factors and short-term structural breaks on oil volatility. The in-sample and out-of-sample results show that macroeconomic fundamentals and provide useful information regarding future oil volatility beyond the historical volatility. We also find that evidence that the structural breaks cause higher degree of GARCH-implied volatility persistence. Two-regime GARCH-MIDAS models can significantly beat their single-regime counterparts in forecasting oil volatility out-of-sample.

EO0579: **The joint credit risk of UK global-systemically important banks**
Presenter: **Yang Zhao**, Jiangxi University of Finance and Economics, China
Co-authors: Minjoo Kim, Mario Cerrato, John Crosby
The joint credit risk in the UK banking sector is studied by using the weekly CDS spreads of global systemically important banks over 2007-2015. We show that the time-varying and asymmetric dependence structure of the CDS spread changes is closely related to the joint default probability that two or more banks simultaneously default. We are able to flexibly measure the joint credit risk at the high-frequency level by applying the combination of the reduced-form model and the GAS-based dynamic asymmetric copula model to the CDS spreads. We also verify that much of the dependence structure of the CDS spread changes are driven by the market factors. Overall, our study demonstrates that the market factors are key inputs for the effective management of the systemic credit risk in the banking sector.

EO0289: **Semiparametric MLE for diffusion models**
Presenter: **Bin Wang**, Shanghai Jiao Tong University, China
Co-authors: Ruijun Bu, Jihyun Kim
A new semiparametric approach has been recently proposed for modelling stationary nonlinear continuous-time diffusions as fully nonparametric transformations of stationary underlying parametric diffusions (UPDs). We extend that by considering nonparametrically transformed recurrent diffusions which can be either stationary or nonstationary. We provide a simple identification argument and propose a new pseudo-ML estimation method for the parameters of the UPDs from discrete samples. Our asymptotic theory assumes that the sampling interval shrinks to zero and the sampling horizon goes to infinity. We show that in the stationary case our PMLE is first-order asymptotically equivalent to the infeasible MLE and thus more efficient than the previous PMLE. Our simulation study examines the finite sample properties of alternative estimators, and an application to financial data is also presented.

EO250 Room LSK1026 HIGH DIMENSIONAL PROBLEMS IN ECONOMETRICS Chair: Juhyun Park

EO0488: **Stochastic evolution of distributions: Applications to CDS indices**
Presenter: **Nicolas Brunel**, ENSIIE, France
Co-authors: Simone Scotti, Guillaume Bernis, Antoine Kornprobst
A mixture of percentile functions is used to model credit spread evolution, which allows us to obtain a flexible description of credit indices and their components at the same time. We show regularity results in order to extend mixture percentile to the dynamic case. We characterise the stochastic differential equation of the flow of cumulative distribution function and we link it with the ordered list of the components of the credit index. The main application is to introduce a functional version of Bollinger bands. The crossing of bands by the spread is associated with a trading signal. Finally, we show the richness of the signals produced by functional Bollinger bands compared with standard one with a practical example.

EO0543: **Estimating spillovers using panel data with a factor structure**
Presenter: **Hao Dong**, London School of Economics, United Kingdom
Co-authors: Qiwei Yao
The estimation of spillover effects using panel data is considered. In contrast with existing approaches, we propose a new method taking advantage of the latent factor structure. Precisely, the cross-sectional variations of the variables that generate spillovers are assumed to be driven by a finite number of factors. Such a latent factor structure implies constraints on spillovers and can be used to improve the performance of existing estimator like the Lasso. As the components of the constraints may not be observed, we propose an estimates based constrained Lasso (Elasso) by substituting in corresponding estimators. When all factors are strong, Elasso asymptotically shares the same error bound with the infeasible constrained Lasso (Classo) where the constraints are assumed to be fully observable, which is known to have sharper error bound than the Lasso. In the presence of

weak factors, the error bound of the Elasso is not as sharp as that of the Classo, but still can be sharper than that of the Lasso when the number of factors is small. Simulation results demonstrate our findings.

EO0618: **Inference on social effects when the network is sparse and unknown**
Presenter: **Eric Gautier**, Toulouse School of Economics, France
Co-authors: Christiern Rose
Models are considered for social interaction when the underlying networks are unobserved but sparse and there are endogenous, contextual, and correlated effects. We accommodate prior knowledge on the sparsity pattern and restrictions on the parameters. We provide results on identification, rates of convergence, model selection, and inference for the parameters and linear functionals in the high-dimensional paradigm. Inference is robust to identification and uniform over large classes of sparse identifiable parameters and data generating processes. Some results hold in finite samples. For computational convenience, we only rely on convex programs.

EO0602: **Functional sparsity in nonparametric varying coefficient models**
Presenter: **Juhyun Park**, Lancaster University, United Kingdom
Co-authors: Catherine Tu, Haonan Wang
Nonparametric estimation is studied under a certain type of sparsity consideration for varying coefficient models in analyzing longitudinal data. The interest in varying coefficient models as concurrent functional linear regression models lies in understanding the time varying effects of the covariates. In the presence of many redundant variables, however, nonparametric inference of such model is not trivial. Some existing works address the problem as a nonparametric variable selection. We directly focus on the problem of selection of active domains of the functional coefficients, locally as well as globally, which we refer to functional sparsity. Our formulation thus generalizes the nonparametric variable selection problem in varying coefficient models. We develop a penalised regression method that addresses both types of selection problems in a unified framework. We establish asymptotic properties of the proposed method. Our method is illustrated in simulation study and real data analysis.

EO0204: **Homogeneity pursuit in panel data models: Theory and applications**
Presenter: **Wuyi Wang**, Singapore Management University, Singapore
Co-authors: Peter CB Phillips, Liangjun Su
The estimation of a panel data model with latent structures is studied. Individuals can be classified into different groups where slope parameters are homogeneous within the same group but heterogeneous across groups. To identify the unknown group structure of vector parameters, we design an algorithm called Panel-CARDS which is a systematic extension of the CARDS. The extension addresses the problem of comparing vector coefficients in a panel model for homogeneity and introduces a new concept of controlled classification of multidimensional quantities called the segmentation net. We show that the Panel-CARDS method identifies group structure asymptotically and consistently estimates model parameters at the same time. External information on the minimum number of elements within each group is not required but can be used to improve the accuracy of classification and estimation in finite samples. Simulations evaluate performance and corroborate the asymptotic theory in several practical design settings. Two empirical economic applications are considered: one explores the effect of income on democracy by using cross-country data over the period 1961-2000; the other examines the effect of minimum wage legislation on unemployment in 50 states of the United States over the period 1988-2014.

EO264 Room LSK1005 DESIGN AND ANALYSIS OF COMPLEX EXPERIMENTS: THEORY AND APPLICATIONS Chair: MingHung Kao

EO0550: **Nature-inspired meta-heuristic algorithms for generating optimal experimental designs for nonlinear models**
Presenter: **Weng Kee Wong**, UCLA, United States
Nature-inspired meta-heuristic algorithms are increasingly studied and used in computer science and engineering disciplines to solve high-dimensional complex optimization problems in the real world. It appears relatively few of these algorithms are used in mainstream statistics even though they are simple to implement, very flexible and frequently able to find an optimal or a nearly optimal solution quickly. These general optimization methods usually do not require any assumption on the function to be optimized and the user only needs to input a few tuning parameters. We provide an overview of such algorithms and demonstrate the usefulness of one of these algorithms for finding different types of optimal designs for nonlinear models, including multiple-objective optimal designs.

EO0601: **Thoughts on optimal statistical design for environmental risk assessment**
Presenter: **John Stufken**, Arizona State University, United States
Co-authors: Wanchunzi Yu
A problem that comes up in environmental risk studies is the estimation of a benchmark dose for polluting agents under a dose-response model. A complication is that there is a great deal of model uncertainty. For a study one needs to select doses for the pollutants at which measurements are to be made, which is a design problem. This design problem has not been studied in great detail. We describe the problem and propose a method that helps to identify efficient designs for estimation of the benchmark dose under model uncertainty. The proposed methodology depends on efficient computations, for which a particle swarm optimization (PSO) algorithm can be considered.

EO0595: **Some connections between multivariate Bernoulli distributions and fractional factorial designs for 2-level factors**
Presenter: **Roberto Fontana**, Politecnico di Torino, Italy
The characterization of multivariate binary Bernoulli distributions with given margins is presented. Each class of multivariate Bernoulli distributions with given margins is expressed as the convex hull generated by some special densities which belong to the class. The problem of finding distributions in a given class and with given moments is studied and solved. Such a characterization is later used for studying fractional factorial designs for 2-level factors. Some comparisons with existing methods for fractional factorial design generation are shown. An algorithm, implemented in SAS, and its application to some examples are also discussed.

EO0466: **Optimizing two-level orthogonal arrays for estimating main effects and pre-specified two-factor interactions**
Presenter: **Ray-Bing Chen**, National Cheng Kung University, Taiwan
Co-authors: Ping-Yang Chen, Devon Lin
The purpose is to consider the construction of D-optimal two-level orthogonal arrays that allow for the joint estimation of all main effects and a specified set of two-factor interactions. A sharper upper bound on the determinant of the related matrix is derived. To numerically obtain D-optimal and nearly D-optimal orthogonal arrays of large run sizes, an efficient search procedure is proposed based on a discrete optimization algorithm. Results on designs of 20, 24, 28, 36, 44 and 52 runs with three or fewer two-factor interactions are presented.

EO0287: **Optimal experimental designs for mixed categorical and continuous responses**
Presenter: **MingHung Kao**, Arizona State University, United States
Experiments with mixed continuous and categorical responses that are possibly correlated are not uncommon in engineering, medical studies and other fields. We develop optimal experimental designs for such experiments for which a joint model for both continuous and categorical responses is considered. We develop results that help to significantly reduce the number of candidate designs, and implement a convex optimization algorithm

to search for optimal designs. The optimality of our designs is verified with the generalized equivalence theorem.

EO148 Room LSKG001 BAYESIAN NONPARAMETRICS Chair: Igor Pruenster

EO0551: **Probabilistic community detection with unknown number of communities**
Presenter: **Anirban Bhattacharya**, , United States
Co-authors: Debdeep Pati, Junxian Geng
A fundamental problem in network analysis is clustering the nodes into groups which share a similar connectivity pattern. Existing algorithms for community detection typically assume the knowledge of the number of clusters or estimate it a priori using various selection criteria and subsequently estimate the community structure. We instead propose a coherent probabilistic framework (MFM-SBM) for simultaneous estimation of the number of communities and the community structure, adapting recently developed Bayesian nonparametric techniques to network models. An efficient Markov chain Monte Carlo (MCMC) algorithm is proposed which obviates the need to perform reversible jump MCMC on the number of clusters. The methodology is shown to outperform recently developed community detection algorithms in a variety of synthetic data examples and in benchmark real-datasets. We derive non-asymptotic bounds on the marginal posterior probability of the true configuration, and subsequently use it to prove clustering consistency results.

EO0341: **Stratified survival regression with Bayesian nonparametric mixtures**
Presenter: **Bernardo Nipoti**, Trinity College Dublin, Ireland
Co-authors: Luis E Nieto-Barajas
The stratified proportional hazards model accounts for heterogeneity in the baseline hazard rate among different groups or strata. Usually strata are determined by means of some categorical covariate. We consider a stratification that takes into account all regression variables and is induced by a Bayesian nonparametric mixture model. We specialise our approach to the frameworks of the proportional hazards and the accelerated life models. Conditionally on the stratification induced by the model, inference is produced for each stratum. We illustrate the performance of our approach by analysing both synthetic and real data.

EO0426: **Nonparametric Bayesian test of independence**
Presenter: **Weining Shen**, uc irvine, United States
The independence testing problem is considered from a Bayesian perspective. We use nonparametric priors such as Dirichlet mixture and random series priors and study the theoretical properties of the resulting Bayes factor. We obtain a consistency result and a non-asymptotic bound under certain regularity conditions.

EO0361: **On distributed Bayesian computation**
Presenter: **Botond Szabo**, Eindhoven University of Technology, Netherlands
Co-authors: Harry Zanten
Due to the rapidly increasing amount of available information computer scientists and statisticians are facing new challenges to deal with "big data" problems. One of the most popular and frequently applied approach to solve this problem are the distributed methods where the data is split into multiple local servers and computations are done in the local computers parallel to each other. This reduces computational time and memory requirement. Then the local machines transmit the outcome of their computations to a global server which aggregates the local results into a global one. In the Bayesian literature, various distributed computational methods were proposed with seemingly good practical performance, but with limited theoretical underpinning. In our work, we investigate the existing distributed methods in a standard non-parametric setting (the Gaussian white noise model) and compare their theoretical performance, i.e. posterior contraction rate and coverage of credible sets. Then we turn our attention to adaptive techniques, which do not use any information about the regularity of the true functional parameter (which is in practice usually not available), but estimate it from the data. We show that the proposed standard methods provide sub-optimal contraction rate and coverage tending to zero for certain true functional parameters. We then propose a new approach which is rate adaptive on (a limited scale of) regularity classes, over performing the standard techniques.

EO0856: **Recursive non-parametric predictive for a discrete regression model**
Presenter: **Lorenzo Cappello**, Bocconi University, Italy
Co-authors: Stephen Walker
A recursive algorithm is proposed to estimate a finite set of conditional distributions. The procedure is fully nonparametric and has a Bayesian interpretation. Indeed the recursive updates follow a Bayesian update to a certain extent. We prove weak convergence of the distribution estimates, and demonstrate numerical accuracy via simulations. A novel fixed-point argument is used to prove the result, along with a novel assumption on the user-supplied sequence of weights that govern the iterations. The estimate is very fast and requires limited computing power; being also parallelizable. We show that it is competitive with both frequentist and Bayesian nonparametric models.

EO096 Room LSK1034 STATISTICAL COMPUTING FOR LARGE PANEL DATA Chair: Feng Li

EO0669: **Bayesian rank selection in multivariate regression**
Presenter: **Anastasios Panagiotelis**, Monash University, Australia
Co-authors: George Athanasopoulos, Rob Hyndman, Bin Jiang, Farshid Vahid
Estimating the rank of the coefficient matrix is a major challenge in multivariate regression, including vector autoregression (VAR). We develop a novel fully Bayesian approach that allows for rank estimation. The key to our approach is reparameterizing the coefficient matrix using its singular value decomposition and conducting Bayesian inference on the decomposed parameters. By implementing a stochastic search variable selection on the singular values of the coefficient matrix, the ultimate selected rank can be identified as the number of nonzero singular values. Our approach is appropriate for small multivariate regressions as well as for higher dimensional models with up to about 40 predictors. In macroeconomic forecasting using VARs, the advantages of shrinkage through proper Bayesian priors are well documented. Consequently, the shrinkage approach proposed here that selects or averages over low rank coefficient matrices is evaluated in a forecasting environment. We show in both simulations and empirical studies that our Bayesian approach provides forecasts that are highly competitive against two of the most promising benchmark methods, dynamic factor models and factor augmented VARs.

EO0491: **Exact subsampling MCMC**
Presenter: **Matias Quiroz**, University of New South Wales, Australia
Co-authors: Minh-Ngoc Tran, Mattias Villani, Robert Kohn
Speeding up Markov Chain Monte Carlo (MCMC) for data sets with many observations by data subsampling has recently received considerable attention in the literature. Most of the proposed methods are approximate, and the only exact solution has been documented to be highly inefficient. We propose a simulation consistent subsampling method for estimating expectations of any function of the parameters using a combination of MCMC with data subsampling and an importance sampling correction for occasionally negative likelihood estimates. Our algorithm is based on

first obtaining an unbiased but not necessarily positive estimate of the likelihood. The estimator uses a soft lower bound such that the likelihood estimate is positive with a high probability, and computationally cheap control variables to lower variability. Second, we carry out a correlated pseudo marginal MCMC on the absolute value of the likelihood estimate. Third, the sign of the likelihood is corrected using an importance sampling step that has low variance by construction. We illustrate the usefulness of the method with two examples.

EO0263: **Conditional Akaike information criteria for a class of Poisson mixture models with random effects**
Presenter: **Dalei Yu**, Yunnan University of Finance and Economics, China
Focusing on the model selection problems in the family of Poisson mixture models (including the Poisson mixture regression model with random effects and zero-inflated Poisson regression model with random effects), the current study derives two conditional Akaike information criteria based on the perturbation techniques. The criteria are the unbiased estimators of the conditional Akaike information based on the conditional log-likelihood and the conditional Akaike information corresponding to the joint log-likelihood, respectively. The derivation is free from the specific parametric assumptions about the conditional mean of the true data-generating model and applies to different types of estimation methods. Additionally, the derivation is not based on the asymptotic argument. Simulations show that the proposed criteria have promising estimation accuracy. In addition, it is found that the criterion based on the conditional log-likelihood demonstrates good model selection performance under different scenarios. Two sets of real data are used to illustrate the proposed method.

EO0700: **Forecasting performance evaluation by time series visualisation and generation**
Presenter: **Yanfei Kang**, Beihang University, China
Co-authors: Rob Hyndman, Kate Smith-Miles
It is common practice to evaluate the strength of forecasting methods using collections of well-studied time series datasets, such as the M3 data. But not so common to determine how diverse these time series are, how challenging, and whether they enable us to study the unique strengths and weaknesses of different forecasting methods. We propose a visualisation method for a collection of time series that enables a time series to be represented as a point in a 2-dimensional instance space. The effectiveness of different forecasting methods can be visualised easily across this space, and the diversity of the time series in an existing collection can be assessed. Noting that the diversity of the M3 dataset has been questioned, a method is also proposed for generating new time series with controllable characteristics to fill in and spread out the instance space, making generalisations of forecasting method performance as robust as possible.

EO0702: **Modeling tail-dependence of stock returns and news sentiments with copulas**
Presenter: **Feng Li**, Central University of Finance and Economics, China
Co-authors: Anastasios Panagiotelis, Yanfei Kang
Tail-dependence modeling based on copula with flexible marginal distributions is widely used in financial time series. Most of the available copula approaches for estimating tail-dependence are restricted within certain types of bivariate copulas due to computational complexity. We propose a general Bayesian approach for jointly modeling high-dimensional tail-dependence for financial returns and related news information.Our method allows for variable selection among the key words in news in the copula tail-dependence parameters. We apply an efficient sampling technique into the posterior inference where the likelihood function is estimated from a random subset of the data, resulting in substantially fewer density MCMC evaluations.

EO190 Room LSK1009 QUANTILE REGRESSION AND ROBUST METHODS Chair: Nan Lin

EO0798: **Testing for marginal effects in quantile regression**
Presenter: **Ian McKeague**, Columbia University, United States
Co-authors: Huixia Judy Wang, Min Qian
The aim is to discuss new marginal testing procedure to detect the presence of significant predictors associated with the quantiles of a scalar response. The idea is to fit the marginal quantile regression on each predictor separately, and then, base the test on the t-statistic associated with the chosen most informative predictor at the quantiles of interest. A resampling method is devised to calibrate this test statistic, which has non-regular limiting behavior due to the variable selection. Asymptotic validity of the procedure is established in a general quantile regression setting in which the marginal quantile regression models can be misspecified. Even though a fixed dimension is assumed to derive the asymptotic results, the proposed test is applicable and computationally feasible for large-dimensional predictors. The method is more flexible than existing marginal screening test methods based on mean regression, and has the added advantage of being robust against outliers in the response. The approach is illustrated using an application to an HIV drug resistance dataset.

EO0429: **Quantile regression for functional partial linear models in high dimensions**
Presenter: **Zhongyi Zhu**, Fudan University, China
A functional partial linear quantile model in high dimensional scenario is considered , where response is a scalar and predictors include multiple random processes and high-dimensional scalar covariates. A framework of regularization with two nonconvex penalty functions in the context of functional quantile regression are proposed formally, and the selection and estimation of important variables can be achieved by minimizing an double penalized functional quantile objective function. The approach proposed takes advantage of functional principal components which greatly facilitates implementation. We introduce a two-step technique for selecting tuning parameters, and establish the asymptotic oracle properties of the proposed estimators based on the difference convex analysis under some regularity conditions. The empirical performance and the usefulness of our approach are demonstrated through a large number of simulation studies and an application to air pollution data.

EO0392: **Estimation and testing for time-varying quantile single-index models with functional/longitudinal data**
Presenter: **Heng Lian**, City University of Hong Kong, Hong Kong
For semiparametric quantile regression, the existing literature is largely focused on independent observations. We propose a time-varying quantile single-index model that is suitable for more complex data where the responses and covariates are longitudinal/functional, with measurements taken at discrete time points. We also develop a statistic for testing whether the time effect is significant. The proposed methodology is illustrated using Monte Carlo simulations and an empirical data analysis.

EO0650: **Flexible composite quantile regression**
Presenter: **Yuanyuan Lin**, The Chinese University of Hong Kong, Hong Kong
Quantile regression is an important tool to analyze heterogeneous data. Composite quantile regression (CQR) is an effective approach for the estimation of quantile regression coefficients, under the assumption that the slope parameters are identical across multiple quantile levels. We propose a new regression method called flexible composite quantile regression (FCQR), which has several advantages: it avoids the global common slope coefficients assumption of CQR; it is able to combine commonality shared by multiple quantile levels nearby for the estimation of the regression coefficient for certain given quantile level; the estimation of the conditional density function is easily available and straightforward in the proposed framework; the crossing problem, which could be rather delicate in the context of quantile regression, can be naturally handled by

a direct correction to the FCQR through inequality constraints to ensure non-crossing quantile curves. We prove the asymptotic properties of the resulting estimator and use extensive simulation studies to examine our theoretical findings.

EO0300: **Stock return autocorrelations in the Chinese stockmarket from threshold quantile autoregressive models**
Presenter: **Liwen Zhang**, School of Economics, Shanghai University, China
The aim is to apply the threshold quantile autoregressive model to study stock return autocorrelations in the Chinese stock market from 2005 to 2014.

EO124 Room LSK1011 RECENT ADVANCES IN LATENT VARIABLE MODELS Chair: Xinyuan Song

EO0450: **Bayesian semiparametric measurement error models with application to dose response analysis**
Presenter: **Taeryon Choi**, Korea University, Korea, South
The aim is to consider Bayesian semiparametric measurement error models, or errors-in-variable regression models based on Fourier series and Dirichlet process mixture, in which the true covariate is not observable, but the surrogate of the true covariate, is only observed. The proposed methodology is compared with other existing approaches to Bayesian measurement error models in simulation studies and bench mark data example. More importantly, we consider a real data application for meta analysis with dose-response analysis, in which measurement errors and shape constraints in the regression functions need to be incorporated with inter-study variability.

EO0572: **Nonparametric estimation of a latent variable model**
Presenter: **Tim Fabian Schaffland**, Eberhard Karls Universitaet Tuebingen, Germany
Co-authors: Augustin Kelava, Michael Kohler, Adam Krzyzak
A new nonparametric latent variable approach is presented. In this approach the model is estimated without specifying the underlying distributions of the latent variables. In a first step, we fit a common factor analysis model to the observed variables. The main trick in estimation of the common factor analysis model is to estimate the values of the latent variables in such a way that the corresponding empirical distribution asymptotically satisfies the conditions that characterize the distribution of the latent variables uniquely. In a second step, we apply suitable nonparametric regression techniques to analyze the relation between the latent variables in this model. Theoretical results (e.g., concerning consistency of the estimates) are briefly presented. Furthermore, the finite sample size performance of the proposed approach is illustrated by applying it to simulated data in three simulation studies.

EO0573: **Batch effects correction with unknown subtypes**
Presenter: **Yingying Wei**, The Chinese University of Hong Kong, Hong Kong
Co-authors: Xiangyu Luo
High-throughput experimental data are accumulating exponentially in public databases. Unfortunately, however, mining valid scientific discoveries from these abundant resources is hampered by technical artifacts and inherent biological heterogeneity. The former are usually termed "batch effects," and the latter is often modeled by subtypes. Existing methods either tackle batch effects provided that subtypes are known or cluster subtypes assuming that batch effects are absent. Consequently, there is a lack of research on the correction of batch effects with the presence of unknown subtypes. We combine a location-and-scale adjustment model and model-based clustering into a novel hybrid one, the Batch-effects-correction-with-Unknown-Subtypes model (BUS). BUS is capable of (a) correcting batch effects explicitly, (b) grouping samples that share similar characteristics into subtypes, (c) identifying features that distinguish subtypes, (d) allowing the number of subtypes to vary from batch to batch, and (e) enjoying a linear-order computational complexity. We prove the identifiability of BUS and provide conditions for study designs under which batch effects can be corrected. BUS is evaluated by simulation studies and a real breast cancer dataset combined from three batches measured on two platforms. Results from the breast cancer dataset offer much better biological insights.

EO0788: **Bayesian analysis of semiparametric hidden Markov models with latent variables**
Presenter: **Jingheng Cai**, Sun Yat-sen University, China
Co-authors: Xinyuan Song, Ming Ouyang, Kai Kang
In psychological, social, behavioral, and medical studies, hidden Markov models (HMMs) have been extensively applied to the simultaneous modeling of heterogeneous observation and hidden transition in analyzing longitudinal data. However, the majority of the existing HMMs are developed in a parametric framework without latent variables. A novel semiparametric HMM is studied, which comprises a semiparametric latent variable model to investigate the complex interrelationships among latent variables and a nonparametric transition model to examine the linear and nonlinear effects of potential predictors on hidden transition. The Bayesian P-splines approach and Markov chain Monte Carlo methods are developed to estimate the unknown functions and parameters. Penalized expected deviance, a Bayesian model comparison statistic, is employed to conduct model comparison. The empirical performance of the proposed methodology is evaluated through simulation studies. An application to a data set derived from the National Longitudinal Survey of Youth is presented.

EO0836: **Bayesian two-level model for partially ordered repeated responses**
Presenter: **Xiaoqing Wang**, The Chinese University of Hong Kong, Hong Kong
Co-authors: Xiangnan Feng, Xinyuan Song
Owing to the questionnaire design and problem nature, partially ordered data that are neither completely ordered nor completely unordered are frequently encountered in behavioral and medical researches. However, among literature, little attention has focused on longitudinal observations with partially ordered data structure. We propose a Bayesian two-level regression model for analyzing longitudinal data with partially ordered responses. The first-level model is defined for partially ordered observations taken at each time point nested within individual and the second-level model is defined for individuals to assess the effect of their characteristics on longitudinal responses. A full Bayesian approach with Markov Chain Monte Carlo (MCMC) algorithm is developed for statistical inference. A simulation study demonstrates that the developed methodology performs satisfactorily. An application to a longitudinal study concerning adolescent substance use is presented.

EP001 Room Ground Floor Hall POSTER SESSION Chair: Panagiotis Paoullis

EP0168: **Confidence intervals for the coefficient of variation with known mean and a bounded standard deviation**
Presenter: **Wararit Panichkitkosolkul**, Thammasat University, Thailand
The natural parameter space is known to be bounded in many real applications such as engineering, sciences and social sciences. The standard confidence interval derived from the classical Neyman procedure is unsatisfactory in the case of a bounded parameter space. New confidence intervals for the coefficient of variation associated with a normal distribution with a known population mean and a bounded standard deviation are proposed. A simulation study has been conducted to compare the performance of the proposed confidence intervals.

EP0542: **Analysis of the effect of the carbon emission reduction in EU ETS Phase 1 on PSM and system dynamics panel DID model**
Presenter: **Honghee Kim**, Statistics Korea, Korea, South
Co-authors: Hyun Kim
The Kyoto Protocol went into effect in 2005. Thus, the EU ETS (Emissions Trading Scheme) was introduced in 2005 as the regulation of market inducement aimed at efficiently reducing the carbon emissions, the cause of climate change. Thereafter, various studies on the performance evaluation of the EU ETS Phase 1 have been carried out, but a number of studies have reported that there is no effective reduction of carbon emissions. In the preceding research, only the participating and non-participating countries that have already been selected are evaluated. In the empirical analysis, the selection of the panel DID model has a limitation in that the estimation coefficient is distorted due to the autocorrelation problem of the error term. Participating and non-participating countries were selected as a sample based on GDP and population density characteristics variable to classify similar countries using PSM (Propensity Scoring Method). After applying the system dynamics panel DID (Difference in Differences) model, the DID coefficient estimates showed a statistically more (-) value than the existing panel DID model (fixed-effects model).In other words, the existing panel DID model does not show statistically significant reductions in carbon emissions from participating countries through the implementation of ETS, whereas the System Dynamics Panel shows that participating countries achieve statistically significant reductions in emissions compared to non-participating countries.

EP0571: **Large shocks and the business cycle: The effect of outlier adjustments**
Presenter: **Yoshihiro Ohtsuka**, Tohoku Gakuin University, Japan
The impact of outlier-adjusted data on business cycle inferences is examined using coincident indicators of the composite index (CI) in Japan. To estimate the CI and business cycles, a Markov switching dynamic factor model incorporating Student's t-distribution in both the idiosyncratic noise and the factor equation is proposed. Furthermore, the model includes a stochastic volatility process to identify whether a large shock is associated with a business cycle. From the empirical analysis using unadjusted data, both the factor and the idiosyncratic component have fat-tail error distributions, and the estimated CI and recession probabilities are close to those published by the Economic and Social Research Institute. Compared with the estimated CI using the adjusted data set, the outlier adjustment reduces the depth of the recession. Moreover, the results of the shock decomposition show that the financial crisis in mid-2008 was caused by unexpected shocks. In contrast, the Great East Japan Earthquake in 2011 was derived from idiosyncratic noise and did not cause a recession. When analyzing whether to use a sample that includes outliers associated with the business cycle, it is not desirable to use the outlier-adjusted data set.

EP0807: **Frequency response analysis of monetary policy transmission**
Presenter: **Lubos Hanus**, Charles University, Czech Republic
Co-authors: Lukas Vacha
A new approach is considered to look at the effects of shocks to dynamics of economic systems. We analyse the widely known phenomenon of price puzzle in a time-varying environment using the frequency decomposition. We use the frequency response function to measure the power of a shock transferred to different economic cycles. Considering both the time-variation of the system and frequency analysis, we can quantify the dynamics of shocks at given time and over different frequencies (horizons) and thus reveal policy implications the system can provide. While studying the monetary policy transmission of the U.S., the empirical evidence shows that low-frequency cycles are prevalent, however, their amplitudes vary significantly in time.

EP0804: **On sums of independent generalized Pareto random variables with applications to insurance and CAT bonds**
Presenter: **Yuanyuan Zhang**, University of Manchester, United Kingdom
Single integral representations are derived for the exact distribution of the sum of independent generalized Pareto random variables. The integrands involve the incomplete and complementary incomplete gamma functions. Applications to insurance and catastrophe bonds are described.

EP0828: **Bayesian analysis of financial volatilities addressing long memory, conditional heteroscedasticity and skewness**
Presenter: **Rosy Oh**, Ewha Womans University, Korea, South
Co-authors: DongWan Shin, Man-Suk Oh
A combined model of ARFIMA, GRACH, and skewed t error distribution is proposed in order to accommodate important features of volatility data; long memory, heteroscedasticity, and asymmetric error distribution. A Bayesian approach is proposed to estimate the parameters of the model, which yields parameter estimates satisfying necessary constraints and can be easily implemented using a free and user-friendly software JAGS. The method is illustrated using a daily volatility index from CBOE (Chicago Board Options Exchange).

EP0853: **Pricey puts and return predictability**
Presenter: **Alex Kontoghiorghes**, Queen Mary University of London, United Kingdom
Significant predictability for the S&P 500 using information extracted from the monthly empirical pricing kernel across a recent ten year sample is documented. These relations persist after controlling for the variance risk premium, risk neutral skewness and kurtosis and the other common behavioural and accounting measures which are used in the predictability literature. The pricing kernel's slope is seen to be highly correlated to business cycle variables, offering suggestions to where the predictability stems from. Cumulative prospect theory preferences is the advocated driver after time varying risk aversion, sentiment measures and divergences in opinion are ruled out.

EP0688: **Application of artificial neural networks to variable annuities**
Presenter: **Longhai Li**, University of Saskatchewan, Canada
The brute-force Monte Carlo (BFMC) simulation is a straightforward method that has long been used to evaluate the risk of financial portfolios. Although the BFMC method works reasonably well for the risk management purposes, it usually involves very intensive computation, due to the nested simulation structure and the needs to simulate a large number of inner loops. To reduce the computation cost, the least-squares Monte Carlo (LSMC) method is used to simulate the liability values. However, the LSMC methods do not work very well in estimating the performance of hedging strategies and the Greeks of variable annuities, and additional research needs to be done to improve the performance of the LSMC method. Potential improvement can be achieved by applying the neural network curve fitting method to the fitting of the proxy functions of liabilities and Greeks, so that better proxy functions can be obtained by allowing more curve fitting flexibilities.

Friday 16.06.2017 15:00 - 16:40 Parallel Session I – EcoSta2017

EI006 Room LSKG001 ADVANCES IN SPATIAL STATISTICS Chair: Mike So

EI0540: **Transformed Gaussian random fields**
Presenter: **Marc Genton**, KAUST, Saudi Arabia
Various transformed Gaussian random fields are studied. In particular, we propose a new class named Tukey g-and-h (TGH) random fields to model non-Gaussian spatial data. The proposed TGH random fields have extremely flexible marginal distributions, possibly skewed and/or heavy-tailed, and, therefore, have a wide range of applications. The special formulation of the TGH random field enables an automatic search for the most suitable transformation for the dataset of interest while estimating model parameters. Asymptotic properties of the maximum likelihood estimator and the probabilistic properties of the TGH random fields are investigated. An efficient estimation procedure, based on maximum approximated likelihood, is proposed and an extreme spatial outlier detection algorithm is formulated. Kriging and probabilistic prediction with TGH random fields are developed along with prediction confidence intervals. The predictive performance of TGH random fields is demonstrated through extensive simulation studies and an application to a dataset of total precipitation in the south east of the United States. Extensions of these ideas to the construction of new spatial max-stable processes and autoregressive time series are presented as well.

EI0647: **High-dimensional Bayesian geostatistics**
Presenter: **Sudipto Banerjee**, UCLA, United States
With the growing capabilities of Geographic Information Systems (GIS) and related software, statisticians today routinely encounter spatial data containing observations from a massive number of locations and time points. Important areas of application include environmental exposure assessment and construction of risk maps based upon massive amounts of spatiotemporal data. Spatiotemporal process models have been, and continue to be, widely deployed by researchers to better understand the complex nature of spatial and temporal variability. However, fitting hierarchical spatiotemporal models is computationally onerous with complexity increasing in cubic order for the number of spatial locations and temporal points. Massively scalable Gaussian process models, such as the Nearest-Neighbor Gaussian Process (NNGP), that can be estimated using algorithms requiring floating point operations (flops) and storage linear in the number of spatiotemporal points. The focus will be on a variety of modeling and computational strategies to implement massively scalable Gaussian process models and conduct Bayesian inference in settings involving massive amounts of spatial data.

EI0653: **Spatial confounding in hierarchical models**
Presenter: **Alexandra Schmidt**, McGill University, Canada
Usually, in spatial generalized linear mixed models, covariates that are spatially smooth are collinear with spatial random effects. This affects the bias and precision of the regression coefficients. This is known in the spatial statistics literature as spatial confounding. One remedy is to assume a restricted spatial regression model wherein the spatial random effects are constrained to be orthogonal to the fixed effects of the model. We discuss the problem of spatial confounding in the case of spatial hierarchical models. In particular, we investigate the possible effect of spatial confounding in the analysis of the performance of students in the 2013 edition of the Brazilian Mathematical Olympiads for Public Schools (OBMEP).

EO022 Room LSK1007 RECENT ADVANCES ON THE ANALYSIS OF EVENT HISTORY STUDIES Chair: Jianguo Sun

EO0164: **A semiparametric likelihood-based method for regression analysis of mixed panel-count data**
Presenter: **Liang Zhu**, UTHealth at Houston, United States
Co-authors: Ying Zhang, Yimei Li, Jianguo Sun, Leslie Robison
Panel-count data arise when each study subject is observed only at discrete time points in a recurrent event study. Observed data include only the numbers of occurrences of the event of interest between observation time points. This type of data frequently occurs in medical follow-up studies, reliability studies and social science. However, sometimes even the number of the event is unknown and we only know whether the event of interest has occurred during that interval or not. We call this type of data panel-binary data. The panel-count data and panel-binary data could co-exist in some recurrent event studies, such as when the count of events is observed for some observation intervals but only binary outcome is observed for other observation intervals. We call this new type of data mixed panel-count data and propose a semiparametric likelihood-based method for its regression analysis. Both finite sample performance and asymptotic properties of the resulting estimators are thoroughly studied. The numerical studies suggest that the proposed method works well in real situations. The approach is applied to a Childhood Cancer Survivor Study that motivated this study.

EO0195: **Semiparametric varying-coefficient regression for recurrent events**
Presenter: **Yang Li**, University of North Carolina-Charlotte, United States
Co-authors: Yanqing Sun, Li Qi
A generalized semiparametric varying-coefficient model for recurrent event data is investigated. It can flexibly model three types of covariate effects: time-constant effects, nonparametric time-varying effects, and varying-coefficient effects that are parametric functions of an exposure variable specified up to a finite number of unknown parameters. Various link functions can be selected to provide a rich family of models for recurrent event analysis. The estimation procedure is developed using profile weighted least squares estimation techniques and the resulting estimators are shown to be consistent and asymptotically normal. An extensive simulation study is conducted to assess the finite sample performance and the methods are applied to data from an acyclovir study.

EO0414: **Regression analysis of case K interval-censored failure time data in the presence of informative censoring**
Presenter: **Peijie Wang**, Jilin University, China
Interval-censored failure time data occur in many fields such as demography, economics, medical research and reliability and many inference procedures on them have been developed. However, most of the existing approaches assume that the mechanism that yields interval censoring is independent of the failure time of interest and it is clear that this may not be true in practice. We consider regression analysis of case K interval-censored failure time data when the censoring mechanism may be related to the failure time of interest. For the problem, an estimated sieve maximum likelihood approach is proposed for the data arising from the proportional hazards frailty model and for estimation, a two-step procedure is presented. In the addition, the asymptotic properties of the proposed estimators of regression parameters are established and an extensive simulation study suggests that the method works well. Finally, we apply the method to a set of real interval-censored data.

EO0539: **Subgroup analysis in censored linear regression**
Presenter: **Xingqiu Zhao**, The Hong Kong Polytechnic University, Hong Kong
Treatment heterogeneity problems with censored response variable are considered. The key of developing individualized treatment strategies is to correctly identify subgroups from a heterogeneous population. To solve this problem, we consider a censored and heterogeneous linear regression model, in which the treatment effects are subject-specific and belong to different subgroups without knowing the group membership of the individuals in advance, and propose a concave fusion penalized method for subgroup analysis. The proposed procedure can automatically

specify the subgroup structure and estimate the subgroup-dependent treatment effects simultaneously. Simulation studies and a real data example illustrate the effectiveness of our proposed method and its practical applications.

EO072 Room LSK1005 STATISTICAL METHODS FOR FUNCTIONAL DATA AND COMPLEX DATA OBJECTS Chair: Ying Chen

EO0166: **Variable selection and structure identification for varying coefficient Cox models**
Presenter: **Toshio Honda**, Hitotsubashi Univeristy, Japan
Co-authors: Ryota Yabe
Varying coefficient Cox models with high-dimensional covariates are considered. We apply the group Lasso to these models and propose a variable selection procedure. Our procedure can cope with simultaneous variable selection and structure identification from a high dimensional varying coefficient model to a semivarying coefficient model. We also derive an oracle inequality and closely examine restrictive eigenvalue conditions. We give the details for Cox models with time-varying coefficients. The theoretical results on variable selection can be easily extended to some other important models and we briefly mention those models since those models can be treated in the same way. The considered models are the most popular models among structured nonparametric regression models. The results of numerical studies are also reported.

EO0169: **Parallel singular value decomposition via multiple random sketches for large matrices**
Presenter: **Weichung Wang**, National Taiwan University, Taiwan
The singular value decomposition (SVD) and high-order SVD (HOSVD) of large-scale matrices is a key tool in data analytics and scientific computing. The rapid growth in the size of matrices further increases the need for developing efficient large-scale SVD algorithms. Randomized SVD based on one-time sketching has been studied, and its potential has been demonstrated for computing a low-rank SVD. We consider a Monte Carlo type integrated SVD algorithm based on multiple random sketches. The proposed integration algorithm takes multiple random sketches and then integrates the results obtained from the multiple sketched subspaces. So that the integrated SVD can achieve higher accuracy and lower stochastic variations. The main component of the integration is an optimization problem and an average scheme over a matrix Stiefel manifold. In addition to the theoretical and statistical analyses, we also consider practical algorithms that are suitable for parallel computers. The proposed algorithms can be implemented on the latest multi-core CPU, many-core GPU, and MPI-based cluster. Numerical results suggest that the proposed integrated SVD and HOSVD algorithms are promising.

EO0191: **Factor augmented dynamic expectile model**
Presenter: **Weining Wang**, Humboldt University at Berlin, Germany
A dynamic expectile factor model is proposed for the tail behavior of a high dimensional time series. In the first step, we extract tail event driven latent factors. In the second step, an VAR (Vector autoregression model) is carried out to analyze the interaction between the factors and the macro economic variables. In particular, we apply the PEC (Principal Expectile Component) algorithm to estimate the latent factors in the first step, and the VAR analysis would facilitate us to carry out impulse response analysis and forecast error variance decomposition to construct a network. The dataset we consider involves 100 US financial firms which represent the US financial market, and 8 macroeconomic variables which are applied in Comprehensive Capital Analysis and Review (CCAR) for the stress test. We develop two scenarios which are caused by the financial market shock, and illustrate the path of those macro variables. Our method provides an additional possibility to design the supervisory scenarios, which considers the network effect of the macro variables as a whole. Our major findings are: GDP, Unemployment rate, Ten year interest rate and house price index are sensitive to the shock of financial market factors; In high risk case, financial factors play more important roles in the network than in the average return case; The network is more dense in the high risk case than the mean case; There are more connectedness during financial crisis than in the stable periods.

EO0483: **A multivariate dynamic analysis of third-party funds**
Presenter: **Alona Zharova**, Humboldt-Universitaet zu Berlin, Germany
Co-authors: Wolfgang Karl Haerdle , Stefan Lessmann
Successful research management of universities requires both qualitative and quantitative information about their research units as well as clear understanding of the existing interdependence structure for effective decision making. The quantitative analysis through parametric and non-parametric techniques is provided. For identification of dependence structure we use vector autoregressive model with exogenous variables (VARX). The data provided by the Humboldt-Universitaet zu Berlin and Scopus database include third-party expenses, publication and citation count as well as academic age of professors from 2001 to 2015. While the analysis is implemented on individual level, the results are introduced in aggregated form on the faculty level.

EO303 Room LSK1034 WAVELETS IN ECONOMICS AND FINANCE Chair: Antonio Rua

EO0200: **Forecasting the equity risk premium with frequency-decomposed predictors**
Presenter: **Fabio Verona**, Bank of Finland, Finland
Co-authors: Goncalo Faria
The out-of-sample forecast of the equity risk premium is significantly improved by taking into account the frequency-domain relationship between the equity risk premium and several potential predictors. We consider fifteen predictors from the existing literature, for the out-of-sample forecasting period from January 1990 to December 2014. The best result achieved for individual predictors is a monthly out-of-sample R2 of 2.98% and utility gains of 549 basis points per year for a mean-variance investor. This performance is improved even further when the individual forecasts from the frequency-decomposed predictors are combined. These results are robust for different subsamples, including the great moderation period, the great financial crisis period and, more generically, periods of bad, normal and good economic growth. The strong and robust performance of this method comes from its ability to disentangle the information aggregated in the original time series of each variable, which allows to isolate the frequencies of the predictors with the highest predictive power from the noisy parts.

EO0386: **Zooming the ins and outs of the U.S. unemployment**
Presenter: **Antonio Rua**, Banco de Portugal, Portugal
Co-authors: Pedro Portugal
To better understand unemployment dynamics it is key to assess the role played by job creation and job destruction. Although the U.S. case has been studied extensively, the importance of job finding and employment exit rates to unemployment variability remains unsettled. The aim is to contribute to this debate by adopting a novel lens, wavelet analysis. We resort to wavelet analysis to unveil time- and frequency-varying features regarding the contribution of the job finding and job separation rates for the U.S. unemployment rate dynamics. Drawing on this approach, we are able to reconcile some apparently contradictory findings reported in previous literature. We find that the job finding rate is more influential for the overall unemployment behavior but the job separation rate also plays a critical role, especially during recessions.

EO0377: **Financial cycles in the euro area: A wavelet analysis**
Presenter: **Michael Scharnagl**, Deutsche Bundesbank, Germany
Co-authors: Martin Mandler
The purpose is to study the relationship of loans to nonfinancial corporations, loans to households, house prices and equity prices between countries or within countries for nine euro area countries. Wavelet analysis allows to account for variations in these relationships both over time and across frequencies. We find evidence of strong comovements between the growth rates in loans to nonfinancial corporations over the sample from 1980 to 2015 for all nine countries. For loans to private households the comovement increases with the start of EMU. The cycle length is in general shorter than claimed by the BIS. Equity prices comove at various frequencies. The degree of synchronization is in general higher for specific series across the EMU countries than across financial series within specific countries.

EO0610: **Stochastic and robust designs to model the interaction of fiscal and monetary policies in a wavelet-based framework**
Presenter: **Patrick Crowley**, Texas A&M University - Corpus Christi, United States
Co-authors: David Hudgins
Fiscal and monetary policies are analyzed in a wavelet-based optimal control setting under stochastic and mixed minimax robust optimal control. The state-space model is constructed by applying the Maximal Overlap Discrete Wavelet Transform (MODWT) to U.S. quarterly real GDP data. We derive the theoretical framework for the linear-quadratic Gaussian design and the mixed design where the disturbances in a subset of frequency ranges are stochastic, and the other disturbances are modeled by a multiple-parameter minimax worst-case robust design. This research is the first to integrate robust control and dynamic game theory with wavelet decomposition.

EO012 Room LSK1010 HIGH DIMENSIONAL AND COMPLEX DATA ANALYSIS Chair: Ray-Bing Chen

EO0236: **Variable selection methods in high-dimensional linear regression**
Presenter: **Yi-Chi Chen**, National Cheng Kung University, Taiwan
Models with high-dimensional covariates have commonly emerged in economics and other fields. Recent developments in methodology have focused on the "large p small n" problems where the number of predictors p is much larger than the sample size n. In this setting, sparse modeling has been widely used in the sense that only a small number of covariates is most relevant to the response. A number of methods for variable selection are considered that are computationally feasible to deal with high dimensionality. We conduct an extensive simulation study to compare accuracy in selection and prediction under different scenarios with regard to the signal-to-noise ratio, sample size and the covariance structure of the predictors. By doing so, we highlight the situations where the performance of these approaches could differ in high-dimensional settings and thus provide guidance to an appropriate choice of method. For empirical application, we consider a high-dimensional dataset on the recent global financial crisis whose number of the potential leading indicators to assess country vulnerability is much larger than a cross-section of countries.

EO0229: **Bayesian structure selection for vector autoregression models**
Presenter: **Shih-Feng Huang**, National University of Kaohsiung, Taiwan
Co-authors: Chi-Hsiang Chu, Mong-Na Lo, Ray-Bing Chen
Vector autoregression (VAR) models are powerful in economic data analysis, because they can include several different time series data simultaneously. A weakness of VAR models is related to the huge coefficient dimensionality. In order to reduce the coefficient dimensionality, a Bayesian structure selection is adopted. For different types of VAR structures, different Bayesian selection approaches are proposed. The results of simulations and a real example show the performance of the proposed Bayesian approaches.

EO0369: **Robust principal expectile component analysis**
Presenter: **Liang-Ching Lin**, National Cheng Kung University, Taiwan
Co-authors: Ray-Bing Chen, Mong-Na Lo Huang, Meihui Guo
Principal component analysis (PCA) is widely used in dimensionality reduction for high dimensional data. It finds principal components by sequentially maximizing the component score variance around the mean. However, in many applications, one is interested in capturing the tail variations of the data rather than variation around the center. In order to capture the tail characters, it was previously proposed principle expectile components (PEC) based on an asymmetric L_2 norm. We introduce a new method called Huber-type principal expectile component (HPEC) using an asymmetric Huber norm to produce robust PECs. Statistical properties of the HPEC are derived and a derivative free optimization approach, particle swarm optimization (PSO), is used to find HPECs. As illustrations, HPEC is applied to real and simulated data with encouraging results.

EO0480: **High-dimensional static and dynamic portfolio selection problems via l1 minimization**
Presenter: **Hoi Ying Wong**, The Chinese University of Hong Kong, Hong Kong
Co-authors: Chi Seng Pun
Several portfolio problems are presented under static and dynamic settings when the number of risky assets (p) is large, and even possibly larger than the number of observation times (n). We prove that the classical plug-in estimation seriously distorts the optimized portfolio. We investigate a constrained l1 minimization approach for directly estimating effective parameters appearing in the optimal portfolio solution. Similar to the Dantzig Selector, the estimator is efficiently implemented with linear programming and the resulting portfolio is called the linear programming optimal (LPO) portfolio. We derive the consistency and rate of convergence for the LPO portfolios. The LPO procedure essentially filters out unfavorable assets based on the investment objective of the investor, resulting in a sparse portfolio. The advantages of the LPO portfolio include the computational superiority, its applicability for dynamic portfolio problems and non-Gaussian distributions of asset returns. Simulations validate the theory and its finite-sample properties. Empirical studies show that the LPO-based portfolios outperform the equally weighted portfolio, the portfolios using shrinkage estimators and other competitive estimators.

EO230 Room LSK1011 RECENT CHALLENGES IN GENETIC ASSOCIATION ATUDIES Chair: Taesung Park

EO0259: **Enhancing power of rare variant association test by a zoom-focus algorithm to locate optimal testing region**
Presenter: **Maggie Haitian Wang**, the Chinese University of Hong Kong, Hong Kong
The increasing amount of whole exome or genome sequencing data brings forth the challenge of analyzing the association of rare variants that have extremely small minor allele frequencies. Various statistical tests have been proposed, which are specifically configured to increase power for rare variants by conducting the test within a certain bin, such as a gene or a pathway. However, a gene may contain from several to thousands of markers, and not all of them are related to the phenotype. Combining functional and non-functional variants in arbitrary genomic region could impair the testing power. We propose a Zoom-Focus Algorithm (ZFA) to locate the optimal testing region within a given genomic region. It can be applied as a wrapper function of existing rare variant association tests to increase testing power. The algorithm is very efficient and the complexity is linear to the number of variants. Simulation studies showed that ZFA substantially increased the statistical power of rare variants tests, including the burden test, SKAT, SKAT-O, and the W-test. The algorithm was applied on real exome sequencing data of hypertensive disorder, and identified biologically relevant genetic markers to metabolic disorder that were undiscoverable by gene-based method. The proposed algorithm is an efficient and powerful tool to enhance the power of association study for whole exome or genome sequencing data.

EO0409: **Testing for genetic associations in arbitrarily structured populations**
Presenter: **Minsun Song**, Sookmyung Women's University, Korea, South

A new statistical test of association between a trait and genetic markers is presented, which is theoretically and practically proved to be robust to arbitrarily complex population structure. The statistical test involves a set of parameters that can be directly estimated from large scale genotyping data, such as those measured in genomewide association studies. We also derive a new set of methodologies, called a genotype-conditional association test, shown to provide accurate association tests in populations with complex structures, manifested in both the genetic and non-genetic contributions to the trait. Our proposed framework provides a substantially different approach to the problem from existing methods.

EO0411: **Pathway-based rare variant association tests with multiple phenotypes**
Presenter: **Taesung Park**, Seoul National University, Korea, South
Co-authors: Sungyoung Lee, Sungkyoung Choi

The advent of next generation sequencing technology has allowed rare variant detection and association with common traits, often by investigating specific genomic regions for rare variant effects on a trait. Although multiple correlated phenotypes are often concurrently observed, most association tests for rare variants analyze only single phenotypes, which may lessen statistical power. To increase power, multivariate analyses, which consider correlations between multiple phenotypes, can be used. However, few existing multivariate analyses can identify rare variants for assessing multiple phenotypes. Here, we propose pathway-based rare variant association tests with multiple phenotypes. Our method is based on the general structured component analysis. We applied the proposed method to whole exome sequencing (WES) data from a Korean population of 1,058 subjects to discover genes associated with multiple traits of liver function. We then performed a simulation study to evaluate the performance of the proposed method. This study illustrates a feasible and straightforward approach to identifying pathways of rare variants correlated with multiple phenotypes, with likely relevance to missing heritability.

EO0832: **Incorporating family disease history in risk prediction models with large-scale genetic data**
Presenter: **Sung Ho Won**, Seoul National University, Korea, South

Despite the many successes of genome-wide association studies (GWAS), the known susceptibility variants identified by GWAS have modest effect sizes, leading to notable scepticism about the effectiveness of building a risk prediction model from large-scale genetic data. However, in contrast to genetic variants, the family history of diseases has been largely accepted as an important risk factor in clinical diagnosis and risk prediction. Nevertheless, the complicated structures of the family history of diseases have limited their application in clinical practice. We develop a new method that enables the incorporation of the general family history of diseases with a liability threshold model, and propose a new analysis strategy for risk prediction with penalized regression analysis that incorporates both large numbers of genetic variants and clinical risk factors. Application of our model to type 2 diabetes (T2D) patients in the Korean population (1846 cases and 1846 controls) demonstrated that single nucleotide polymorphisms accounted for 0.28 of the variability of risk in T2D cases, and incorporation of family history led to an additional 0.06 improvement in prediction. Our results illustrate that the family history of diseases represents invaluable information about the variability of complex diseases and improve the prediction performance

EO164 Room LSK1032 NON- AND SEMI-PARAMETRIC METHODS FOR ECONOMICS AND FINANCIAL DATA Chair: Chae Young Lim

EO0362: **Capital asset pricing model: A time-varying volatility approach**
Presenter: **Taejin Kim**, The Chinese University of Hong Kong, Hong Kong
Co-authors: Kun Ho Kim

A methodology is proposed to conduct uniform inference of volatility in the capital asset pricing model (CAPM). To that end, relevant theory is employed to construct the uniform confidence band of the volatility in the CAPM. The methodology is applied to the U.S. stock return data. The empirical results show strong evidence of co-movement among the volatility estimates for six U.S. stocks of large market capitalization. The hypothesis of constant volatility for the CAPM is rejected unanimously, mainly due to the surge in volatility in the early 2000s and during the 2008 financial crisis.

EO0524: **Testing heterogeneous treatment effects based on permutation tests**
Presenter: **EunYi Chung**, University of Illinois at Urbana Champaign, United States

A test procedure is proposed based on permutation tests for testing whether the treatment effect is constant across all observations. Of course, when the constant treatment effect τ is known, the permutation test is exact in finite samples. However, when the parameter τ is unknown and thus needs to be estimated, the permutation test based on the plug-in test statistic may fail to control the Type 1 error. To overcome this so-called Durbin problem, the martingale transformation is proposed, as proposed previously. As a result, the transformed test statistic becomes asymptotically pivotal and thus the permutation test will be asymptotically valid while still providing an exact error control in finite samples when the treatment effect is null.

EO0532: **Forward premium anomaly: New insight through a time-varying-parameter approach**
Presenter: **Kun Ho Kim**, Hanyang University, Korea, South

Recently developed kernel smoothing regression procedures and simultaneous confidence bounds are used to investigate the forward premium puzzle. These new statistical methods estimate the time-varying slope coefficient of the regression of spot returns on the lagged interest rate differential. Simultaneous confidence bands are used to test when uncovered interest parity is violated. The estimated betas in the forward premium smoothed regression are found to vary substantially over time and to be partially explicable in terms of lagged fundamentals and money growth volatility arising from risk premium. Model averaging procedures indicate the relative importance of these variables in terms of predicting the breakdown of UIP condition.

EO0487: **Bootstrap-assisted unit root testing with piecewise locally stationary errors**
Presenter: **Yeonwoo Rho**, Michigan Technological University, United States
Co-authors: Xiaofeng Shao

In the unit root testing, the piecewise locally stationary process is adopted to accommodate nonstationary errors that can have smooth and abrupt changes in second or higher order properties. Under this new framework, the limiting null distributions of the conventional unit root test statistics are derived and shown to contain a number of unknown parameters. To circumvent the difficulty of direct consistent estimation, we propose to use the dependent wild bootstrap to approximate the nonpivotal limiting null distributions and provide a rigorous theoretical justification for the bootstrap consistency. The proposed method is compared with the recolored wild bootstrap procedure, which was developed for the error that follows a heteroscedastic linear process through finite sample simulations. Further, a combination of autoregressive sieve recoloring with the dependent wild bootstrap is shown to perform well. The validity of the dependent wild bootstrap in the nonstationary setting is revealed for the first time, showing the possibility of extensions to other inference problems associated with locally stationary processes.

EO046 Room LSKG007 NEW DEVELOPMENTS IN TIME SERIES ANALYSIS Chair: Cathy W-S Chen

EO0370: **A time series model for realized volatility matrices based on the matrix-F distribtuion**
Presenter: **Wai-Keung Li**, , China
Co-authors: Jiayuan Zhou, Ke Zhu
Realized covariance (RCOV), as a volatility estimator calculated from high frequency time series, is playing an important role in recent statistics and econometrics researches. We focus on modeling the dynamic of the stochastic, symmetric and positive definite RCOV matrices. Unlike some pioneer works in this field, including WAR or CAW models where the Wishart distribution is adopted, we suggest the utilization of a more general matrix-F distribution, to accommodate heavy-tailed data. We managed to provide sufficient condition for the stationarity of the model. Concerning the curse of dimensionality and over-parametrization problems, under stationarity, a parsimonious variance targeted (VT) model is introduced. Under this VT setting, the asymptotic consistency and normality of the maximum likelihood estimator are established. Moreover, this VT model is further reduced to parsimonious forms, with merits on maintaining positive definiteness and computation flexibility. Monte Carlo simulations are used for demonstrating the methodology and real data analysis is conducted on intraday data from S&P500 database.

EO0435: **Machine learning pairs trading**
Presenter: **Meihui Guo**, National Sun Yat-sen University, Taiwan
Co-authors: Yu-Jin Lai
Pairs trading is a market-neutral trading strategy which utilizes a pair of related financial instruments to make profits by exploiting their relations. We select trading pairs using cointegration models. And use covariates such as spreads, money flow, relative strength index and etc. as input features for support vector machine classifier to set up trading signals of entering positions. We investigate the performance of the proposed pairs trading strategy for high frequency transaction data.

EO0367: **On the CUSUM test for location-scale time series models**
Presenter: **Sangyeol Lee**, Seoul National University, Korea, South
The problem of testing for a parameter change in location-scale time series models including ARMA-GARCH models is considered. We study the cumulative sum (CUSUM) tests based on estimates, score vectors, and residuals, and compare their performance. For the score vector-based CUSUM test, we also discuss a bootsrap method. It is shown that under regularity conditions, their limiting null distributions are the sup of Brownian bridges. A simulation study and real data analysis are conducted for illustration.

EO0546: **Approximate maximum likelihood estimation of a threshold diffusion process**
Presenter: **Henghsiu Tsai**, Academia Sinica, Taiwan
An approximate maximum likelihood approach is proposed for estimating both the drift and the diffusion parameters of a two-regime threshold diffusion process. We study the small and large sample properties of the estimator and apply the model to estimate the term structure of a long time series of US interest rates.

EO216 Room LSK1014 STATISTICAL INFERENCE FOR HIGH-DIMENSIONAL DATA Chair: Shaojun Guo

EO0389: **Statistical inference for model parameters with stochastic gradient descent**
Presenter: **Xi Chen**, New York University, United States
The problem of statistical inference of the true model parameters is investigated based on stochastic gradient descent (SGD). To this end, we propose two consistent estimators of the asymptotic covariance of the average iterate from SGD: (1) an intuitive plug-in estimator and (2) a computationally more efficient batch-means estimator, which only uses the iterates from SGD. As the SGD process forms a time-inhomogeneous Markov chain, our batch-means estimator with carefully chosen increasing batch sizes generalizes the classical batch-means estimator designed for time-homogenous Markov chains. Both proposed estimators allow us to construct asymptotically exact confidence intervals and hypothesis tests. We further discuss an extension to conducting inference based on SGD for high-dimensional linear regression.

EO0439: **Dimensionality reduction and variable selection in multivariate varying-coefficient models**
Presenter: **Kejun He**, Institute of Statistics and Big Data, Renmin University of China, China
Co-authors: Shujie Ma, Heng Lian, Jianhua Huang
Motivated by the study of gene and environment interactions, we consider a multivariate response varying-coefficient model with a large number of covariates. The need of non-parametrically estimating a large number of coefficient functions given relatively limited data poses a big challenge for fitting such a model. To overcome the challenge, we develop a method that incorporates three ideas: i. reduce the number of unknown functions to be estimated by using (non-centered) principal components; ii. approximate the unknown functions by polynomial splines; iii. apply sparsity-inducing penalization to select relevant covariates. The three ideas are integrated into a penalized least squares framework. Our asymptotic theory shows that the proposed method can consistently identify the irrelevant covariates and can estimate the non-zero coefficient functions with the same convergence rate as when only the relevant covariates are included in the model. We also develop a novel computational algorithm to solve the penalized least squares problem by combining proximal algorithms and optimization over Stiefel manifolds. Our method is illustrated using data from Framingham Heart Study.

EO0522: **Learning heterogeneity in causal inference using sufficient dimension reduction**
Presenter: **Wenbo Wu**, University of Oregon, United States
Co-authors: Wei Luo, Yeying Zhu
Often the research interest in causal inference is to see how the covariates affect the mean difference in the potential outcomes. We use sufficient dimension reduction to estimate a lower dimensional linear combination of the high dimensional covariates that are sufficient for this purpose. To enhance interpretability of the results, we further modify the estimator using sparse sufficient dimension reduction, which selects an active set of covariates for variable selection as by-product. The estimator can also be used to test the heterogeneity of the causal effect. Compared to the existing methods, our approach is model-free, and avoids regression on each potential outcome. Thus it can be more applicable and effective. These advantages are supported by both simulation studies and a real data example.

EO0585: **Valid inference on semiparametric estimators in high-dimensions**
Presenter: **Shaojun Guo**, Institute of Statistics and Big Data, Renmin Unversity of China, China
Inference after model selection is a fundamental problem in high dimensional settings. A selective overview of post-selection inference in statistical and economic fields will be given, and why statistical inference after model selection is challenging will be discussed. Second, we will focus on a specific topic of post-selection inference. To be specific, we will try to deal with the problem where covariates are generated through high dimensional regularization. It turns out that the regularization step has a very serious effect for valid inference on parameters of interest. Our primary interest is to develop a novel regularized approach to generate covariates. The proposed estimator can be shown to be asymptotically normal. To illustrate that, we provide several examples to demonstrate the superiority of the proposed approach. This approach is also applicable

to linear or nonlinear functionals in other sparse nonparametric high dimensional regression models such as additive or varying coefficient models.

EO309 Room LSK1003 SMART BETA AND QUANTITATIVE INVESTING Chair: Serge Darolles

EO0459: **Smart beta indices with controlled factor exposures**
Presenter: **Serge Darolles**, Paris Dauphine, France
The standard mean-variance approach can imply extreme weights in some assets in the optimal allocation and a lack of stability of this allocation over time. To improve the robustness of the portfolio construction, practitioners often introduce additional constraints on the risk contributions and/or the weights of the portfolio constituents. An incidental outcome of these alternative weighting approaches is a sometimes unintentional and/or uncontrolled exposure to risk factors. We introduce a set of portfolio management approaches which control for the degree of factors exposure of the portfolio and for the turnover.

EO0403: **Smart beta second generation: Evolution of smart beta towards a continuously adapting investment process**
Presenter: **Alain Groshens**, NovumC, China
Based on a recent Smart Beta (SB) investors survey and two decades of practice, we describe and explain the SB investors needs, preferences and requirements governing the evolution of systematic and quantitative factor investing into what we call Smart Beta Second Generation (SB2G). We will focus our analysis and examples within the multi-asset strategy space. First we layout the survey results: the perceived benefits and shortcomings of SB strategy. We acknowledge the requirement to consider the risk adjusted performance as a primary investment strategy objective. We note the rejection for historical data dependent rules and market blind deterministic allocation process of SB strategies. Subsequently, we discuss the principles and processes that lead to the concept of SB2G: 1- A continuously adaptive investment process integrating and linking asset screening, allocation process and portfolio implementation. 2- SB2G includes human supervision and its investment bias within a rule based, yet adaptive investment process. 3- The performance is a primary objective of the process. 4- SB2G implementation relies on an allocation driver application that continuously recalibrates and feeds the investment processes with live market data and qualitative variables such as market sentiments. At last, we describe an actual SB2G multi-asset strategy in production.

EO0626: **Daily price limits and destructive market behavior**
Presenter: **Wenxi Jiang**, Chinese University of Hong Kong, Hong Kong
Account-level data from the Shenzhen Stock Exchange is considered to analyze the effects of daily price limits, a widely adopted market stabilization mechanism across the world. The following findings suggest that daily price limits may lead to unintended, destructive market behavior: 1) after a stock hits the 10% upper price limit, its price continues to rise on the next day and the price increase reverses in the long run; 2) large investors tend to buy on the day of the upper price limit being hit and then sell on the next day; and 3) large investors' net buying on the limit-hitting day significantly predicts stronger long-run price reversal. We also analyze a sample of special treatment (ST) stocks, which face tighter 5% daily price limits, and provide a causal validation from comparing market dynamics before and after they are officially assigned the ST status.

EO0660: **Limit order book microstructure in Asia and applications in trading**
Presenter: **Ngoc Minh Dang**, HSBC, China
Some relevant microstructure aspects of limit order book for Asia markets are investigated, more precisely, the autocorrelation of the order flow and the order book imbalance. We show how to link a structural model of the transaction price to the estimation and explanation of the Arrival Cost of an execution strategy. This approach allows us to understand how the market context would influence the Arrival Cost, in addition to the size and aggression of the order. We also analyze the link between the order book imbalance and the probability of the next trade arrival. It turns out that this is an important variable when executing orders on large spread and long queue stocks.

EO026 Room LSK1001 NETWORKS AND CAUSALITY Chair: Monica Billio

EO0504: **Networks in risk spillovers: A multivariate GARCH perspective**
Presenter: **Lorenzo Frattarolo**, Ca Foscari Venezia, Italy
Co-authors: Monica Billio, Loriana Pelizzon, Massimiliano Caporin
A spatial BEKK model is proposed where time-varying networks represent the spillover channel. We derive covariance stationarity and identification restrictions. Furthermore, we show the model's usefulness in: (i) evaluating the importance of a risk channel, (i) isolating risk spreaders and risk receivers, (ii) computing target exposure able to reduce the volatility of the system. We investigate financial cross-country holdings as a risk spillover channel in the Euro area.Our empirical analysis use banking statistics provided by the Bank of International Settlements (BIS).

EO0685: **Bayesian Markov switching tensor regression**
Presenter: **Matteo Iacopini**, Ca Foscari University of Venice, Italy
Co-authors: Monica Billio, Roberto Casarin
A Bayesian Markov Switching regression model is proposed for jointly modelling a network and vector of observables. The dependence of the two stochastic processes is driven by a unique hidden Markov chain whose regimes are identified by means of relevant network statistics. We exploit the Polya-Gamma data augmentation scheme for logit models in order to provide an efficient Gibbs sampler for posterior inference. The model is firstly tested on simulated data, then it is used in order to classify financial networks and a set of market indices in order to disentangle their common characteristics.

EO0507: **Estimation and model-based combination of causality networks**
Presenter: **Roberto Panzica**, Goethe University House of finance, Italy
Co-authors: Massimiliano Caporin, Giovanni Bonaccolto
Causality is a widely used concept in theoretical and empirical economics. The recent financial economics literature has used Granger causality to detect the presence of contemporaneous links between financial institutions and, in turn, to obtain a network structure. Subsequent studies combined the estimated networks with traditional pricing or risk measurement models to improve their fit to empirical data. We provide two contributions: we show how to use a linear factor model as a device for estimating a combination of several networks that monitor the links across variables from different viewpoints; and we demonstrate that Granger causality should be combined with quantile-based causality when the focus is on risk propagation. The empirical evidence supports the latter claim.

EO0514: **Collective market movements: An econometric approach through Von Neumann entropy**
Presenter: **Monica Billio**, University of Venice, Italy
Co-authors: Roberto Casarin, Michele Costola, Lorenzo Frattarolo
When the market moves in unison, the complexity of its dynamics is minimal but masked by a network of interactions. A single dominant stochastic trend is a possible econometric interpretation of the phenomenon. The aim is to investigate the Von Neumann entropy (VNE) of the sample covariance as a complexity measure of the system. In fact, VNE is asymptotically minimal and equal to zero only in the single stochastic trend

case. Unfortunately, in the same case, its asymptotic distribution is singular forcing us to develop an inference procedure based on cointegrating vector driven stationary bootstrap. An application on national stock indices allows us to search for crisis and bubbles episodes.

EO214 Room LSK1033 INFERENCE AND APPLICATIONS FOR TIME SERIES MODELS Chair: Kun Chen

EO0574: **Efficient inference for nonlinear state space models with an automatic sample size selection rule**
Presenter: **Jing Cheng**, Huaqiao University, China
The maximum likelihood estimation of nonlinear state space models is studied. Particle Markov chain Monte Carlo method is introduced to implement the Monte Carlo Expectation Maximization algorithm for more accurate and robust estimation. Using this framework, an automated sample size selection criterion is constructed using renewal theory to boost the sample size as one is covered by the confidence interval of relative likelihood constructed. The proposed methodology is applied to the stochastic volatility model for illustration.

EO0578: **Factor model with SCM approach**
Presenter: **Meng Meng**, The Chinese University of Hong Kong, Hong Kong
Co-authors: NH Chan
Factor analysis is one of the most frequently used tools to achieve dimension reduction in analyzing high-dimensional time series. Factor model through eigen-analysis is utilized as the first step in our method to extract the common factors, which affect the dynamics of the multiple components simultaneously. The structure of the latent factors is modeled by vector autoregressive-moving average model (VARMA). Using the SCM approach will ease the procedure of investigating the orders and sparse properties of coefficient matrix. For the identified model with some imposed restrictions, we develop maximum likelihood estimation for an alternative representation of VARMA. This general form is desirable because it directly focuses on the original data while sparsity and locations of zeros are reserved. Our method achieves computational gains via estimating only lower-dimensional nonzero parameters through iterative algorithm. Moreover, we derive the convergence rate of the factors and asymptotic normality of the estimated parameters. Simulations and empirical studies are conducted to illustrate the performance of our methodology.

EO0649: **Least squares estimation of threshold cointegration**
Presenter: **Man Wang**, Donghua University, China
Threshold cointegration is an important model in econometrics and statistics. The estimation theory of this model has always been an interesting topic ever since the model was first proposed. Existing estimation procedures are usually based on the threshold error correction model (TECM). With this representation, we have studied the LSE and established the asymptotic distribution. Particularly, it is found that when the error correction model is discontinuous over the threshold, the estimated threshold is $n-$ consistent and converges weakly to the smallest minimizer of a compound Poisson process. To simulate the non-standard limiting distribution of the estimated threshold, a previous theory is revisited and a numerical approach is proposed. In addition, the result is illustrated with an application to the term structure of interest rates of the U.S.

EO0659: **Dynamic spillover effects of RMB exchange rate to stock markets**
Presenter: **Xiaofei Wu**, Donghua University, China
With the acceleration of the economic globalization and Financial liberalization, the linkage between RMB exchange rate and domestic stock market is gradually enhanced. So it is necessary to use the most advanced measurement models for empirical research in order to improve the scientific study and accuracy. There is a non-stationary and nonlinear volatility spillover effects between the frequent fluctuation of the RMB exchange rate and China's stock market, showed the different performance in different periods of time. Using advanced empirical GC-MSV model, through parameter estimation of MCMC method based on Gibbs sampling, found that the exchange rate market and China's stock market (Shanghai and Shenzhen market) existed volatility spillover effect which shows differently in three periods of time.

EO248 Room LSKG003 INTEGRATING BIG AND COMPLEX IMAGING DATA WITH NEW STATISTICAL TOOLS Chair: Guodong Li

EO0593: **Spatially weighted reduced-rank framework for functional MRI data**
Presenter: **Mihye Ahn**, University of Nevada Reno, United States
Recently, much attention has been received on the analysis of functional imaging data to delineate the intrinsic functional connectivity pattern among different brain regions within each subject. However, only few approaches for integrating functional connectivity pattern from multiple subjects have been proposed. The goal is to develop a reduced-rank model framework for analyzing the whole-brain voxel-wise functional images across multiple subjects in the frequency domain. Considering the neighboring voxels with different weights, the frequency and spatial factors can be extracted. Imposing sparsity on the frequency factors enables us to identify the dominant frequencies. In addition, the spatial maps can be used for detecting group difference, when the comparison between different groups is of specific interest. Simulation study shows that the proposed method achieves less spatial variability and better estimates of frequency and spatial factors, compared to some existing methods. Finally, we apply the proposed method to Alzheimer's Disease Neuroimaging Initiative (ADNI) data.

EO0567: **Network of networks: A large scale graphical model for whole brain networks using fMRI**
Presenter: **Xi Luo**, Brown University, United States
Functional MRI (fMRI) has enabled scientists to study brain networks. However, the dimensionality of whole-brain fMRI, usually in hundreds of thousands, challenges the applicability of statistical methods for estimating direct connectivity. A large-scale graphical model method is introduced for modeling direct connections among a large number of nodes. This model introduces two layers of networks. The first layer network models the connections between groups of voxels or ROIs using sparse Gaussian graphical models, and the second layer network models the local connections within each ROI. This two-layer model provides a rigorous and interpretable approach for inferring big networks with hundreds of thousands of nodes. An interesting implication is that classical ROI-based methods for estimating the first level network will lead to degraded estimation accuracy. We develop a fast method and algorithm to compute both layers of networks and voxel partitions, exploiting the conditional convexity of our formulation. The effectiveness of this approach is demonstrated on simulated data and a real dataset from a stop/go fMRI experiment.

EO0670: **Posterior mean screening for big neuroimaging data via massively parallel computing**
Presenter: **Jian Kang**, University of Michigan, United States
Motivated by the needs of selecting important features from big neuroimaging data, a Bayesian variable screening algorithm for ultra-high dimensional data is developed consisting of two steps: Step 1: compute a multivariate variable screening statistic based on marginal posterior moments; Step 2: perform the mixture model-based cluster analysis on screening statistics to identify the unimportant variables. Step 1 only requires a computational complexity on the order of $O(n^2 p_n)$ and it is straightforward to be parallelized. It has a close connection with sure independent screening (SIS) statistics and high-dimensional ordinary least-squares projection (HOLP) methods. Step 2 is an extension of the local false discovery rate (FDR) analysis. We implement our method using massively parallel computing techniques based on the general-purpose computing on graphics processing units (GPGPU), leading to an ultra-fast variable screening procedure. Our simulation studies show that the proposed approach can perform variable screening on one million predictors within seconds and achieve higher selection accuracy compared with existing methods. We

also illustrate our methods on an analysis of resting state functional magnetic resonance imaging (Rs-fMRI) data from the Autism Brain Imaging Data Exchange (ABIDE) study.

EO0684: **Automated brain hematoma and edema segmentation of CT scans using non-local spatial clustering and the level-set method**
Presenter: **Wei Tu**, University of Alberta, Canada
Co-authors: Linglong Kong
Hematoma and edema volume are potential predictors of 30-day mortality rate and functional outcome (degree of disability or dependence in the daily activities after a stroke) of patients with intracerebral hemorrhage (ICH). The manual segmentation of hematoma and edema from computed tomography (CT) scans is the common practice, but also a time-consuming and labor-intensive task. The automated segmentation of hematoma and edema is an appealing alternative, but a challenging task due to the poorly defined boundary of edema and the surrounding healthy brain tissue. There is only limited literature on this problem. We propose a novel framework to fill this gap between the theoretical development of segmentation methods and this practical need. Our framework is fully automated and works in an unsupervised fashion. The method uses non-local regularized spatial fuzzy C-means clustering in the initialization stage and level set method in the refinement stage. To evaluate the proposed method, 30 subjects with different size, shape and location of hematoma and edema were used. Compared with manual segmentation results of two independent raters, our method shows an excellent matching in hematoma with an average dice score coefficient 0.92, which is significantly better than the previous methods.

EO034 Room LSK1009 RECENT ADVANCES ON HYPOTHESIS TESTING Chair: Xuehu Zhu

EO0673: **Heteroscedasticity tests for nonlinear regression models**
Presenter: **Xu Guo**, Beijing Normal University, China
An efficient test statistic is developed for heteroscedasticity check for nonlinear regression model. Our proposed test statistic can avoid the so called curse of dimensionality. The proposed method requires no bandwidth selection, is simple to compute, based merely on pairwise distances between points in the sample. Asymptotic results are shown. It is proven that our proposed test can converge to finite limit at the rate of $1/n$ under the null hypothesis and can detect any local alternatives which converge to the null hypothesis at the rate of $1/\sqrt{n}$. These results are surprising and interesting. Asymptotically, the dimension of the covariates has no effect on the convergence rate. Some simulation studies are conducted to illustrate the proposed test statistic.

EO0674: **Enhancements of nonparametric generalized likelihood ratio test: Bias-reduction and dimension reduction**
Presenter: **Cuizhen Niu**, Beijing Normal University, China
Co-authors: Xu Guo, Lixing Zhu
Nonparametric generalized likelihood ratio test is a popular method of model checking for regressions. However, there are two issues that may be the barriers for its powerfulness. First, the bias term in its limiting null distribution causes the test not to well control type I error and thus Monte Carlo approximation for critical value determination is required. Second, it severely suffers from the curse of dimensionality due to the use of multivariate nonparametric function estimation. The purpose is thus twofold: a bias-reduction is suggested and a dimension reduction-based adaptive-to-model enhancement is recommended to promote the power performance. The proposed test statistic still possesses the Wilks phenomenon, and behaves like a test with only one covariate. Thus, it converges to its limit at a much faster rate and is much more sensitive to alternative models than the classical nonparametric generalized likelihood ratio test. As a by-product, we also prove that the bias-corrected test can be more efficient than the one without bias-reduction in the sense that its asymptotic variance is smaller. Simulation studies are conducted to evaluate the finite sample performance and to compare with other popularly used tests. A real data analysis is conducted for illustration.

EO0678: **A projection based adaptive to model test for regressions**
Presenter: **Falong Tan**, Hong Kong Baptist University, Hong Kong
A longstanding problem of existing empirical process-based tests for regressions is that when the number of covariates is greater than one, they either have no tractable limiting null distributions or are not omnibus. We propose a projection-based adaptive-to-model approach to attack this problem. When the hypothetical model is parametric single-index, our method can fully utilize the dimension reduction model structure under the null hypothesis as if the covariate were one-dimensional such that the martingale transformation-based test can be asymptotically distribution-free. Further, the test can automatically adapt to the underlying model structure such that the test can be omnibus and thus detect alternative models distinct from the hypothetical model at the fastest possible rate in hypothesis testing. The method is examined through simulation studies and is illustrated by a real data analysis.

EO0672: **Specification testing with mixed discrete and continuous predictors: A projection-based adaptive-to-model approach**
Presenter: **Xuehu Zhu**, Xi'an Jiaotong university, China
A nonparametric projection-based adaptive-to-model specification test is proposed, which can be used to check the regression model with both discrete and continuous predictors. Unlike existing locally smoothing kernel tests, the new test behaves like a locally smoothing test as if the number of continuous predictors was one, and it can detect the local alternative hypotheses distinct from the null hypothesis at the rate that is only related to one continuous predictors. The test statistic is asymptotically normal under the null hypothesis such that critical values are easily determined. Simulation studies and real data analysis are then conducted to illustrate the performance of the new test and compare it with existing tests.

Friday 16.06.2017 17:10 - 18:50 Parallel Session J – EcoSta2017

EI002 Room LSKG001 MODERN METHODS FOR COMPLEX FUNCTIONAL AND LONGITUDINAL DATA Chair: Tsung-I Lin

EI0440: **Heavy-tailed longitudinal regression models for censored data: A likelihood based perspective**
Presenter: **Victor Hugo Lachos Davila**, UNICAMP, Brazil
Co-authors: Mauricio Castro, Tsung-I Lin, Larissa Avila Matos
HIV RNA viral load measures are often subjected to some upper and lower detection limits depending on the quantification assays. Hence, the responses are either left or right censored. Moreover, it is quite common to observe viral load measurements collected irregularly over time. A complication arises when these continuous repeated measures have a heavy-tailed behavior. For such data structures, we propose a robust nonlinear censored regression model based on the scale mixtures of normal (SMN) distributions. To take into account the autocorrelation existing among irregularly observed measures, a damped exponential correlation structure is considered. A stochastic approximation of the EM (SAEM) algorithm is developed to obtain the maximum likelihood estimates of the model parameters. The main advantage of this new procedure allows us to estimate the parameters of interest and evaluate the log-likelihood function in an easy and fast way. Furthermore, the standard errors of the fixed effects and predictions of unobservable values of the response can be obtained as a by-product. The practical utility of the proposed method is exemplified using both simulated and real data.

EI0498: **Identifying multiple changes in a sequence of functional data**
Presenter: **Jeng-Min Chiou**, Academia Sinica, Taiwan
Co-authors: Yu-Ting Chen
A systematic approach is proposed to identify changes in a sequence of functional data, where the total number and the positions of the change-points are unknown and are to be estimated. The algorithm comprises the dynamic segmentation approach coupled with the backward elimination procedure. It recursively searches for the change-point candidates by dynamic segmentation, and these candidates are further assured in backward elimination using hypotheses testing for statistical significance. The selected change-point candidates are consistent with the actual ones if they exist. We illustrate the method through an application to highway traffic flow analysis and examine the practical performance of the algorithm via a simulation study.

EI0770: **jmcm: An R package for joint mean-covariance modelling of longitudinal data**
Presenter: **Yi Pan**, The University of Manchester, United Kingdom
Co-authors: Jianxin Pan
Longitudinal studies commonly arise in various fields such as psychology, social science, economics and medical research, etc. It is of great importance to understand the dynamics in the mean function, covariance and/or correlation matrices of repeated measurements. However, high-dimensionality (HD) and positive-definiteness (PD) constraints are two major stumbling blocks in modelling of covariance and correlation matrices. It is evident that Cholesky-type decomposition based methods are effective in dealing with HD and PD problems, but those methods were not implemented in statistical software yet, causing a difficulty for practitioners to use. We first introduce recently developed Cholesky decomposition based methods for joint modelling of mean and covariance structures, namely modified Cholesky decomposition (MCD), alternative Cholesky decomposition (ACD) and hyperspherical parameterization of Cholesky factor (HPC). We then introduce our newly developed R package jmcm which is currently able to handle longitudinal data that follows a Gaussian distribution using the MCD, ACD and HPC methods. Demonstration is provided by running the package jmcm and comparison of those methods is made through analysing two real data sets.

EO060 Room LSK1001 TIME SERIES MODELING AND ITS APPLICATIONS Chair: Heung Wong

EO0156: **Quantile autoregressive conditional heteroscedasticity**
Presenter: **Guodong Li**, University of Hong Kong, Hong Kong
Co-authors: Yao Zheng
A novel conditional heteroscedastic time series model is proposed by applying of the idea of quantile regression models to the ARCH(∞) form of the GARCH model. This new model includes the commonly used GARCH model as a special case, and it is also able to provide different structures for conditional quantiles of a time series. The strict stationarity and ergodicity is discussed, and the asymptotic distributions of the corresponding quantile estimation are derived. Simulation experiments are carried out to access the finite sample performance of the proposed methodology. An empirical example is presented to illustrate the usefulness of the new model.

EO0274: **Asymptotic inference for the threshold autoregressive model with a structural change**
Presenter: **Shiqing Ling**, HKUST, China
The least squares estimation of the threshold autoregressive model with a structural change is investigated. It is shown that both the estimated threshold and change-point are n-consistent, and they converge weakly to the smallest minimizer of a compound Poisson process and the location of minima of a two-sided random walk, respectively. When the magnitude of change in the parameters of the state regimes or the time horizon is small, it is further shown that these limiting distributions can be approximated by a class of known distributions. Other estimated slope parameters are $\sqrt{n}$-consistent and asymptotically normal. A simulation study is carried out to assess the performance of our procedure. The result is further applied to study the growth of tree rings in China, from which we obtain evidence of climate change during the period of 1641-1663. This supports historians' view that the bad weather is because of climate change, which probably is one reason in deep for the downfall of Chinese Ming Dynasty.

EO0193: **Moving average model with an alternative GARCH-type error**
Presenter: **Xingfa Zhang**, Guangzhou University, China
Co-authors: Huafeng Zhu, Xin Liang, Yuan Li
Motivated by the double autoregressive model with order p (DAR(p) model), the moving average model with an alternative GARCH error is studied. The model is an extension from DAR(p) model by letting the order p goes to infinity. The quasi-maximum likelihood estimator of the parameters in the model is shown to be asymptotically normal, without any strong moment conditions. Simulation results confirm that our estimators perform well. We also apply our model to study a real data set and it has better fitting performance compared to DAR model for the considered data.

EO0286: **Robust closed-form estimators for the integer-valued GARCH(1,1) model**
Presenter: **Fukang Zhu**, Jilin University, China
A closed-form estimator and its several robust versions for the integer-valued GARCH(1,1) model are proposed. These estimators are easy to implement and do not require the use of any numerical optimization procedure. Consistency and asymptotic normality for the non-robust closed-form estimator is established. The robustification of the closed-form estimator is done by replacing the sample mean and autocorrelations by robust estimators of them, respectively. The performances of these closed-form estimators are investigated and compared via simulations. New estimators

are applied to 5 stock-market data sets with different periods and time intervals, and their prediction performances are assessed by in-sample prediction, out-of-sample prediction and scoring rules. Other possible proposals related to the closed-form estimators are also discussed.

EO134 Room LSKG007 FINANCIAL AND RISK MANAGEMENT APPLICATIONS Chair: Toshiaki Watanabe

EO0157: **Forecasting limit order book liquidity supply-demand curves with functional autoregressive dynamics**
Presenter: **Ying Chen**, National University of Singapore, Singapore
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 the limit order book and demand curves and we utilize the fitted model to predict the joint evolution of the liquidity demand and supply curves. In the VFAR framework, we derive a closed-form maximum likelihood estimator under sieves and provide the asymptotic consistency of the estimator. In an application to limit order book records of 12 stocks in NASDAQ traded from 2 Jan 2015 to 6 Mar 2015, the VAR model presents a strong predictability in liquidity curves, with R^2 values as high as 98.5% for in-sample estimation and 98.2% in out-of-sample forecast experiments. It produces accurate 5-, 25- and 50-minute forecasts, with root mean squared error as low as 0.09 to 0.58 and mean absolute percentage error as low as 0.3 to 4.5%.

EO0159: **Understanding the two-way relationship between the ASX and NZX indexes: A vector threshold autoregressive approach**
Presenter: **Wai-Sum Chan**, The Chinese University of Hong Kong, Hong Kong
Co-authors: Johnny S-H Li, Andrew C-Y Ng
The non-linear dynamic relationship between the stock markets in Australia and New Zealand is studied. Specifically, we model the returns on the ASX 200 and NZX 50 indexes with a vector threshold autoregressive (VETAR) model. The model contains more than two regimes, each of which is characterized by a different parameter set and possibly a different autoregressive order. Therefore, in different regimes, the two stock indexes have different cross-correlations and are expected to increase or decrease at different rates. In our set-up, the switch between these regimes is governed by the average difference between the returns on the two stock indexes over a certain lookback period. The estimated model can hence be used to anticipate the behaviors of the two stock markets when one of them significantly out- or under-performs the other. An explanation of how investment decisions can be made on the basis the estimated VETAR model is given. This may help hedge funds that are interested in the oceanic stock markets to formulate investment strategies. It can also assist current and future retirees in Australia and New Zealand with the development of their retirement savings portfolios. By means of simulations, we demonstrate that using the VETAR model to make investment decisions may reduce the risk of an investment portfolio, in terms of risk measures such as Value-at-Risk.

EO0183: **Nonparametric tolerance limits for pair trading**
Presenter: **Cathy W-S Chen**, Feng Chia University, Taiwan
Co-authors: Tsai-Yu Lin
Tolerance intervals are an important statistical tool for determining the threshold of a certain reference. We propose to utilize nonparametric one-sided tolerance limits with three look-back window sizes for return spreads in order to find trading entry and exit signals. We illustrate how the proposed method helps to uncover arbitrage opportunities via the daily return spreads of 12 stock pairs in the U.S. markets and then report the performance of pair trading for two out-of-sample periods. The empirical results suggest that combining the minimum squared distance method and nonparametric one-sided tolerance limits generates positive excess returns, relative to the underlying stocks.

EO0663: **Bayesian shrinkage estimation of multivariate financial time series**
Presenter: **Mike So**, The Hong Kong University of Science and Technology, Hong Kong
Analyzing multiple financial time series is getting more challenging as there is growing demand in combining information from a large number of securities in forming portfolios or for risk management. In statistical terms, it is interesting to incorporate potential sparsity in modeling high-dimensional time series. We propose a Bayesian approach to formulate dynamic shrinkage in correlations while accounting for conditional heteroskedasticity in financial time series. Numerical studies using simulated and real data demonstrate possible financial applications of the methodology.

EO220 Room LSK1005 NEW METHODS AND APPLICATIONS IN QUANTILE REGRESSION AND BEYOND Chair: Weining Wang

EO0177: **Bayesian spectral analysis quantile regression models with shape restrictions**
Presenter: **Genya Kobayashi**, Chiba University, Japan
Co-authors: Taeryon Choi, Taeyong Roh
Some econometric models are presented based on the Bayesian quantile regression with shape-restricted functions using Gaussian process priors. By assuming the derivatives of the functions to be squares of Gaussian processes, the resulting functions are monotonic, monotonic convex or concave, U-shaped, and S-shaped. Introducing the shape restriction avoids overfitting and helps to smooth the estimates of the conditional quantiles especially for the extreme quantiles. The new shape-restricted quantile regression model is extended to deal with censored data and panel data. Specifically, for the panel data, the model with positive random effects are considered in the stochastic frontier framework. The usefulness of the proposed models are demonstrated using the simulated and real datasets.

EO0184: **Weighted composite quantile regression for single-index models**
Presenter: **Rong Jiang**, Donghua university, China
A weighted composite quantile regression (WCQR) estimation is proposed for single-index models. For the parametric part, the WCQR is augmented by using a data-driven weighting scheme. With the error distribution unspecified, the proposed estimators share robustness with the quantile regression and achieve nearly the same efficiency as the semiparametric maximum likelihood estimator for a variety of error distributions including the Normal, Students t, Cauchy distributions, etc. Furthermore, based on the proposed WCQR, we use the adaptive-LASSO to study variable selection for parametric part in the single-index models. For the nonparametric part, the WCQR is augmented by combining the equal weighted estimators with possibly different weights. Because of the use of weights, the estimation bias is eliminated asymptotically. By comparing the asymptotic relative efficiency theoretically and numerically, the WCQR estimation all outperforms the CQR estimation and some other estimation methods. Under regularity conditions, the asymptotic properties of the proposed estimators are established. Simulation studies and two real data applications are conducted to illustrate the finite sample performance of the proposed methods.

EO0537: **Nonlinear spatial hedonic quantile regression: Housing prices, relevant characteristics, and their shadow prices**
Presenter: **Joachim Schnurbus**, University of Passau, Germany
Co-authors: Markus Fritsch, Harry Haupt
In many applications of statistical real estate appraisal methods the following challenges arise simultaneously: [1] relevant characteristics of a property need to be identified, [2] shadow prices (marginal market valuation) of characteristics and [3] prices of bundles (of characteristics) not observed need to be estimated. State of the art hedonic housing price analysis comprises [i] modeling price functions nonlinearly, [ii] accounting for complex spatial association structures (horizontal market segmentation), and [iii] allowing for varying functional relationships across the conditional

price distribution (vertical market segmentation). We discuss two general classes of nonlinear quantile regression models which meet these criteria but pursue different avenues to simultaneously address the challenges [1]-[3]. Due to the underlying assumptions, the inference obtained from both model classes differs analytically and – more importantly – leads to different economic interpretations. The methods are illustrated by applying them to data generating processes with various degrees of functional and spatial complexity in a Monte Carlo study and to geo-referenced urban house price data.

EO0393: **A new approach to censored quantile regression estimation**
Presenter: **Xiaorong Yang**, Zhejiang Gongshang University, China
Co-authors: Naveen Narisetty, Xuming He
Quantile regression provides an attractive tool to the analysis of censored responses, because the conditional quantile functions are often of direct interest in regression analysis,and moreover, the quantiles are often identifiable while the conditional mean functions are not. Existing methods of estimation for censored quantiles are mostly limited to singly left or right-censored data, with some attempts made to extend the methods to doubly-censored data. We propose a new and unified approach, based on a variation of the data argumentation algorithm, to censored quantile regression estimation. The proposed method adapts easily to different forms of censoring including doubly censored and interval censored data, and somewhat surprisingly, the resulting estimates improve on the performance of the best known estimators with singly censored data.

EO254 Room LSKG003 NEW METHODS IN HIGH DIMENSIONAL DATA ANALYSIS Chair: Eric Chi

EO0208: **Tensor regression and sufficient dimension reduction with applications to neuroimaging analysis**
Presenter: **Xin Zhang**, Florida State University, United States
Regression models with tensors (multi-dimensional arrays) will be considered, and applications to some neuroimaging data sets will be discussed. Tensor versions of envelope models, partial least squares (PLS) algorithms, and discriminant analysis will be introduced.

EO0269: **Covariate information matrix for dimension reduction**
Presenter: **Weixin Yao**, UC Riverside, United States
Co-authors: Debmalya Nandy, Bruce Lindsay, Francesca Chiaromonte
Starting from the Density Information Matrix (DIM) approach, we develop a tool for Sufficient Dimension Reduction (SDR) in regression problems called Covariate Information Matrix (CIM). CIM exhaustively identifies the Central Subspace (CS) and provides a rank ordering of the reduced covariates in terms of their regression information. Compared to other popular SDR methods, CIM does not require distributional assumptions on the covariates, or estimation of the mean regression function. CIM is implemented via eigen-decomposition of a matrix estimated with the f2 method of computation, an efficient nonparametric density estimation technique. Expanding previous work, we also propose a bootstrap-based diagnostic tool for estimating the dimension of the CS. We use simulations and real data to demonstrate the effectiveness of our approach compared to other SDR methods.

EO0270: **On the estimation of ultra-high dimensional semiparametric Gaussian copula models**
Presenter: **Qing Mai**, Florida State University, United States
The semiparametric Gaussian copula model has wide applications in econometrics, finance and statistics. Recently, many have considered applications of semiparametric Gaussian copula model in several high-dimensional learning problems. We propose a slightly modified normal score estimator and a new Winsorized estimator for estimating both nonparametric transformation functions and the correlation matrix of the semiparametric Gaussian copula model. Two new concentration inequalities are derived, based on which we show that the normal score estimator and the new Winsorized estimator are consistent when the dimension grows at an exponential rate of the sample size. As demonstration, we apply our theory to two high-dimensional learning problems: semiparametric Gaussian graphical model and semiparametric discriminant analysis.

EO0837: **Bayesian regression trees for high-dimensional prediction and variable selection**
Presenter: **Antonio Linero**, Florida State University, United States
Decision tree ensembles are an extremely popular tool for obtaining high quality predictions in nonparametric regression problems. Unmodified, however, many commonly used decision tree ensemble methods do not adapt to sparsity in the regime in which the number of predictors is larger than the number of observations. A recent stream of research concerns the construction of decision tree ensembles which are motivated by a generative probabilistic model, the most influential method being the Bayesian additive regression trees (BART) framework. We take a Bayesian point of view on this problem and show how to construct priors on decision tree ensembles which are capable of adapting to sparsity in the predictors by placing a sparsity-inducing Dirichlet hyperprior on the splitting proportions of the regression tree prior. We characterize the asymptotic distribution of the number of predictors included in the model and show how this prior can be easily incorporated into existing Markov chain Monte Carlo schemes. We demonstrate that our approach yields useful posterior inclusion probabilities for each predictor and illustrate the usefulness of our approach relative to other decision tree ensemble approaches on both simulated and real datasets.

EO086 Room LSK1010 REGRESSION AND CLASSIFICATION IN HIGH-DIMENSIONAL SPACES Chair: Binyan Jiang

EO0243: **Sparse and multi-collinear regression via empirical Bayesian elastic net**
Presenter: **Lilun Du**, HKUST, China
Co-authors: Anhui Huang
High dimensional statistical inference is typically based on the assumption that the true regression function lies in a low dimensional manifold. Consequently, a sparse regression model is widely used in many problems of this type such as proteomics and genomics. As technology advancement has made high dimensional variables abundantly available, the aforementioned studies also require identifying causal effects from groups of highly correlated variables, pointing to the demands of both sparseness and multicollinearity. Existing algorithms, such as Lasso, empirical Bayesian Lasso (EBlasso), and elastic net (EN) trade statistical capabilities including hypothesis testing and/or collinear variable screening off for sparseness so as to infer such high dimensional models. We develop an empirical Bayesian elastic net (EBEN) algorithm for generalized linear models that enables both sparseness and multicollinearity, yet is able to infer the model accurately and efficiently. Simulation results demonstrated that EBEN detected more true effects with a similar number of false effects comparing to EN and EBlasso. Particularly, when there were groups of highly correlated variables, EBEN was able to select the whole group of variables while other methods typically selected one variable out of a group. When analyzing a real dataset, EBEN yielded more statistical meaningful results than those previously reported.

EO0636: **Rank based screening for high-dimensional survival data**
Presenter: **Jinfeng Xu**, University of Hong Kong, Hong Kong
Variable screening for censored survival data is most challenging when both survival and censoring times are correlated with a ultra high-dimensional vector of covariates. Existing approaches to handling censoring often make use of inverse probability weighting by assuming independent censoring with both survival time and covariates. This is a convenient but rather restrictive assumption which may be unmet in real applications, especially when the censoring mechanism is complex and the number of covariates is large. To accommodate covariate-dependent

censoring that is often present in high-dimensional survival data, we propose a rank based screening method to select features that are relevant to the survival time. The method is invariant to monotone transformations of the response and of the predictors, and works robustly for a general class of survival models. We establish the sure screening property of the proposed methodology. Simulation studies and a lymphoma data analysis demonstrate its favorable performance and practical utility.

EO0766: **Dynamic linear discriminant analysis in high dimensional space**
Presenter: **Binyan Jiang**, The Hong Kong Polytechnic University, Hong Kong
Co-authors: Ziqi Chen, Chenlei Leng
High-dimensional data that evolve dynamically feature predominantly in the modern data era. As a partial response to this, recent years have seen increasing emphasis to address the dimensionality challenge. However, the non-static nature of these datasets is largely ignored. Both challenges are addressed by proposing a novel yet simple dynamic linear programming discriminant (DLPD) rule for binary classification. Different from the usual static linear discriminant analysis, the new method is able to capture the changing distributions of the underlying populations by modeling their means and covariances as smooth functions of covariates of interest. Under an approximate sparse condition, we show that the conditional misclassification rate of the DLPD rule converges to the Bayes risk in probability uniformly over the range of the variables used for modeling the dynamics, even when the dimensionality is allowed to grow exponentially with the sample size. Minimax lower bound for the estimation of the Bayes risk is also established, which implies that the misclassification rate of our proposed rule is minimax-rate optimal. The promising performance of the DLPD rule is illustrated via extensive simulation studies and the analysis of a breast cancer dataset.

EC0324: **Tensor regression in economics and finance**
Presenter: **Giuseppe Brandi**, LUISS University, Italy
Multidimensional data (tensor data) is a relevant topic in statistical and machine learning research. Examples of multidimensional data are panel data (individuals $\times$ variables $\times$ time $\times$ locations) or higher order (> 2) multivariate portfolio moments (the coKurtosis) or 3D images (cube of pixels). Given their complexity, such data objects are usually reshaped into matrices and then analyzed. However, doing so poses other issues. First of all, it destroys the intrinsic interconnections among the datapoints in the multidimensional space and, secondly, the number of parameters to be estimated in a model increase exponentially. To alleviate these issues the data is treated as it is and a model able to deal with the multidimensionality of the dataset is developed. In particular, a parsimonious tensor regression (in which both the regressor and the response are tensors) is build such that it retains the intrinsic multidimensional structure of the dataset. Tucker decomposion is employed to achieve parsimony and an ALS algorithm is developed to estimate the model parameters. A simulation exercise is produced to validate the model and an empirical application to macroeconomic time series is carried over to compare its prediction ability with existing ones.

EO064 Room LSK1034 HETEROSKEDASTICITY AND AUTOCORRELATION ROBUST INFERENCE Chair: Cheng Liu

EO0278: **Self-normalized subsampling of time series**
Presenter: **Shuyang Bai**, University of Georgia, United States
Co-authors: Murad Taqqu, Ting Zhang
The inference procedure for the mean of a stationary time series is usually different under different model assumptions, because the sample sum behaves differently depending on whether the time series has short or long memory, and whether it has light or heavy tails. Typical inference procedures involve the estimation of some nuisance parameters. It is often challenging for the practitioners to decide which procedure to use given the data, and to know whether the estimation of the nuisance parameters is reliable. We propose a inference method which combines two ingredients: self-normalization and resampling. This method provides a unified inference procedure for the various aforementioned model assumptions. Furthermore, it avoids the estimation of nuisance parameters, and requires only the choice of one bandwidth.

EO0334: **An exactly (almost) unbiased long run variance estimator addressing the finite sample bias**
Presenter: **Jingjing Yang**, University of Nevada, Reno, United States
An (almost) exactly unbiased long run variance (LRV) estimator is proposed for a univariate time series with unknown mean that addresses the finite sample bias problem. This LRV estimator builds on the concept of an exactly unbiased autocovariance estimator recently proposed. The standard kernel LRV estimator with fixed-bandwidth (fixed-b) is known to be inconsistent and to be biased under finite samples. The exactly unbiased LRV estimator is shown to be composed of three parts: the standard kernel LRV estimator, the asymptotic bias correction, and the finite sample bias correction. The appealing feature of this estimator is to deal with the bias correction problem especially in finite samples. The fixed-b limiting distributions of the LRV estimator are derived. A different method was also previously used which has the asymptotic bias correction. However, the finite sample bias correction to the estimator is added, resulting in the exact unbiasedness of the estimator in finite samples. Simulations show that this LRV estimator has less bias and still maintains variance comparable with other estimators. Using this LRV estimator, the hypothesis test of a simple location model generally has a less distorted size control, while maintaining a power as high as the standard t test.

EO0607: **Simple and trustworthy asymptotic t tests in difference-in-differences regressions**
Presenter: **Cheng Liu**, Wuhan University, China
Co-authors: Yixiao Sun
Two asymptotically valid t tests in a difference-in-differences (DD) regression are proposed when the number of time periods is large while the number of individuals can be small or large. Each of the two t tests is based on a special heteroscedasticity and autocorrelation robust (HAR) variance estimator that is tailored towards the inference problem in the DD setting. The difference between the two t tests is that one is based on the sandwich variance estimator of a general form while the other is based on the sandwich variance estimator of a special form. By capturing the estimation uncertainty of the HAR variance estimators, both t tests have more accurate size than the corresponding normal tests. They are also as powerful as the latter tests. Compared to the nonstandard tests that are designed to reduce the size distortion of the normal tests, the proposed t tests are as accurate but are much more convenient to use, as critical values are from the standard t table.

EO0605: **Inference in time series models using smoothed clustered standard errors**
Presenter: **Seunghwa Rho**, Louisiana State University, United States
Co-authors: Timothy Vogelsang
A long run variance estimator is proposed for conducting inference in time series regression models that combines the traditional nonparametric kernel approach with a cluster approach. The basic idea is to divide the time periods into non-overlapping clusters. The long run variance estimator is constructed by first aggregating within clusters and then kernel smoothing across clusters. We develop an asymptotic theory for test statistics based on this smoothed clustered long run variance estimator. We derive asymptotic results holding the number of clusters fixed and also treating the clusters as increasing with the sample size. We find that the fixed number of clusters asymptotic approximation works well whether the number of clusters is small or large. Finite sample simulations suggest that clustering before kernel smoothing can reduce over-rejections caused by strong serial correlation without a great cost in terms of power. The simulations also suggest that the naive iid bootstrap mimics the fixed number of clusters critical values.

EO070 Room LSK1009 RECENT ADVANCES IN SPATIAL STATISTICS Chair: Huiyan Sang

EO0283: **Large dimensional penalized maximum likelihood estimation and variable selection in geostatistics**
Presenter: **Tingjin Chu**, Rennim University of China, China

In high dimensional spatial data analysis, the problem of selecting covariates and estimating parameters in spatial linear models with Gaussian process errors is considered. We propose two penalized methods for the spatial data with diverging number of covariates, based on penalized least square estimation and penalized maximum likelihood estimation, respectively. The optimization is carried out through a coordinate descent algorithm. The convergence rate for parameters' estimation error is investigated, and sparsistency results on model selection are obtained. Monte Carlo results show the proposed methods' better performance than other competitors. The proposed methods are applied on a real air quality index data and some useful insight is obtained.

EO0455: **Regionalization of multiple air pollutants based on functional principal component analysis**
Presenter: **Hui Huang**, Peking University, China

Emission patterns of air pollutants may vary across species and regions. They are closely related to local industry and energy consumption structures, but usually not directly observable. To make more effective environment policies, we need to better understand regional emission features. Conventional methods to regionalize air pollutants, such as Empirical Orthogonal Functions (EOF) or Self Organizing Maps (SOM), have limitations in combing multiple air pollutants with their spatial correlation structures. We propose a new method based on Functional Principal Component Analysis (FPCA), in which the associations among different pollutants and monitor sites are treated as correlations between functional PC scores. The regionalization is realized by a likelihood-based clustering. We apply our method to analyze 6 main pollutants over the North China Plain (NCP) from 2014 to 2016. Our findings indicate that different parts of NCP need different policies to control the air pollution.

EO0548: **Variogram models on all spheres**
Presenter: **Juan Du**, Kansas State University, United States

Variogram or variogram matrix functions play an important role in modeling dependence structure among multiple processes at different locations in spatial statistics. With more and more data collected on a global scale in environmental science, climatology, geophysics, and related fields, we focus on the characterizations of the variogram models on spheres of all dimensions for both stationary and intrinsic stationary, univariate and multivariate processes. Some efficient approaches are proposed to construct a variety of variogram functions including simple polynomial structures, or to show whether an existing function on Euclidean space is valid on all spheres. In particular, the series representation and spherical behavior of intrinsic stationary random fields are elaborated in both theory and simulation study. We also use simulation and real data analysis to demonstrate the application of the proposed models and theoretical results involved in terms of estimation and kriging.

EO0547: **Approaches for massive spatial data and applications in remote sensing**
Presenter: **Emily Kang**, University of Cincinnati, United States

With the development of new remote sensing technology, large or even massive spatial datasets from Earth observation become available. Statistical analysis of such data is challenging. We proposes a semiparametric approach to modeling and inference for massive spatial datasets. In particular, a Gaussian process with additive components is considered, with its covariance structure coming from two components: one part is flexible without assuming a specific parametric covariance function but is able to achieve dimension reduction; the second part is parametric and simultaneously induces sparsity. The inference algorithm for parameter estimation and spatial prediction is devised. The method is applied to simulated data and a massive dataset of sea surface temperature acquired from NASA's Terra satellite. The results demonstrate the computational and inferential benefits of the proposed method over competing methods and show that our method is more flexible and more robust against model misspecification. Other applications and extensions will also be discussed.

EO0845: **Spatially clustered coefficient models with large data sets**
Presenter: **Huiyan Sang**, Texas A&M University, United States
Co-authors: Furong Li

The availability of large spatial and spatial-temporal data geocoded at accurate locations has fueled increasing interest in spatial modeling and analysis. When data are collected over a large area, it is unlikely that the usual constant spatial regression models can adequately capture the spatially dynamic relationship between a response variable and explanatory variables. In many applications, the coefficients may exhibit cluster structure with homogeneity within clusters and abrupt changes across the boundary of clusters. We propose the spatially clustered coefficient (SCC) regression model to capture the spatial pattern, especially clustering pattern in the effect of explanatory variables. The method can be used to deal with irregularly spaced data and our empirical studies suggest that it works very effectively in estimation for data either with clustered coefficients or smoothly-varying coefficients. We establish theoretical properties of the estimators, which provides non-asymptotic error bounds on the estimation and prediction. An application of SCC method to the temperature and salinity data in the Atlantic basin is provided for illustration.

EO146 Room LSK1003 TOPICS IN FINANCIAL AND NONPARAMETRIC ECONOMETRICS Chair: Jeroen Rombouts

EO0284: **Large scale panel quantile regressions with unobservable effects**
Presenter: **Tomohiro Ando**, Melbourne Business School, Australia
Co-authors: Jushan Bai

A panel quantile regression model with interactive fixed effects, which allows multiple individual effects and can capture high-dimensional cross sectional/serial dependence, is introduced. The direct minimization of the quantile loss function is challenging because of the nonlinearity in both the functional form and the interactive effects. A new data-augmentation strategy for the estimation is introduced. The developed data-augmentation strategy provides simple inference methods and there is a rich opportunity to apply the proposed method. A large panel data set that consists of final asset returns is then analysed.

EO0354: **Two-sample estimation of varying coefficient models via nearest neighbor matching**
Presenter: **Masayuki Hirukawa**, Setsunan University, Japan
Co-authors: Artem Prokhorov

Economists often face the situation in which all necessary variables must be collected from more than one source when running a regression. We investigate the problem of estimating varying coefficient models using the combined sample that is constructed from two samples via the nearest neighbor matching. Our particular focus is on the kernel-smoothed local linear estimation of the varying coefficient. It is demonstrated that the local linear estimator using matched samples is inconsistent, as is the case with the ordinary least squares estimator of the linear regression model using matched samples. If only a few variables are used to impute the missing data, then it is possible to correct for the bias. We propose a bias-correction method for the local linear estimation and explore asymptotic properties of the bias-corrected estimator. Monte Carlo simulations confirm that the bias correction works very well in such cases.

EO0385: **Modelling variance risk premia via variance swap payoffs**
Presenter: **Jeroen Rombouts**, ESSEC Business School, France
Co-authors: Francesco Violante, Lars Stentoft
An approach using synthetic variance swap payoffs is proposed to estimate the VRP and information related to extreme payoff events. Variance swap payoffs are highly volatile series, with time varying variance levels and extreme payoffs during volatile market conditions. To cope with these features, we use signal extraction techniques based on a state-space representation of the model and the Kalman-Hamilton filter. Since we know from financial theory that the VRP is positive, we impose this economic constraint when estimating the model. This approach allows us to obtain measurement error free estimates of the smooth component of VRP, and construct variables indicating agents' expectations under extreme market conditions. This information is shown to be very useful in predictive return regressions on the S&P500, DJIA, NASDAQ and RUSSELL index markets.

EO0388: **Pricing individual stock options using both stock and market index information**
Presenter: **Lars Stentoft**, University of Western Ontario, Canada
Co-authors: Jeroen Rombouts, Francesco Violante
When it comes to individual stock option pricing, most, if not all, applications consider a univariate framework in which the dynamics of the underlying asset is considered without taking the evolution of the market or any other risk factors into consideration. From a theoretical point of view this is clearly unsatisfactory as we know, i.e. from the Capital Asset Pricing Model, that the expected return of any asset is closely related to the exposure to the market risk factors. On top of this theoretical inconsistency in empirical applications it is often difficult to precisely asses and appropriately measure risk premia from individual stock returns alone. We model the evolution of the individual stock returns together with market index in a bivariate model that allows us to estimate risk premia in line with the theory. We assess the performance of the model to price individual stock options on the constituent stocks in the Dow Jones Industrial Average.

EO038 Room LSK1027 ADVANCES IN STATISTICAL AND ECONOMETRIC MODELLING OF RISK PROCESSES Chair: Zudi Lu

EO0296: **Additive nonparametric models with time variable and both stationary and nonstationary regressors**
Presenter: **Chaohua Dong**, Southwestern University of Finance and Economics, China
Co-authors: Oliver Linton
Nonparametric additive models are considered that have deterministic time trend and both stationary and integrated variables as components. The diverse nature of the regressors caters for applications in a variety of settings. In addition, we extend the analysis to allow the stationary regressor to be instead locally stationary, and we allow the models to include a linear form of the integrated variable. Heteroscedasticity is allowed for all models. We propose an estimation strategy based on orthogonal series expansion that takes account of the different type of stationarity or nonstationarity possessed by each covariate. We establish the large sample distribution theory for all the estimated functions we consider. In spite of the entanglement of different kinds of regressors, we can separate out the distribution theory for each estimator. We provide Monte Carlo simulations that establish the favourable properties of our procedures in moderate sized samples. Finally, we apply our techniques to the study of a pairs trading strategy.

EO0665: **Semiparametric estimation for optimal dividend barrier with insurance portfolio**
Presenter: **Hiroshi Shiraishi**, Keio University, Japan
Co-authors: Zudi Lu
How an insurance portfolio is used to provide dividend income for insurance company's shareholders is an important problem in application of risk theory, where the premium income as dividends is paid to the shareholders, whenever the surplus attains a level barrier, until the next claim occurs. Under the aggregate claims process taken as a compound Poisson model, we define optimal dividend barrier as the level of the barrier that maximizes the expectation of the discounted dividends until ruin. In the literature, the optimal dividend barrier was derived explicitly under some fixed models concerning claim distribution where parametric estimation is possible. In practice, it may often be hard to provide with the claim distribution parametric model either only from theoretical point of view or from a finite sample, and thus non-parametric estimation is preferred. We consider the semi-parametric estimation of the optimal dividend barrier. A contribution in practice to decision-making on dividend barrier is made in the case where a new product is launched or optimality of an existing dividend barrier is tested.

EO0708: **Safety-first rule based portfolio selection with background risk**
Presenter: **Yuanyao Ding**, Ningbo University, China
A a safety-first rule based model is set up for portfolio selection with background risk, and then conditions are investigated for the existence of finite optimal portfolios and provides the analytical expressions of the optimal portfolios. Further, the relationship between the efficient portfolio frontier in a safety-first framework and the background risk are discussed, and the sensitivity of portfolio selection to background risk is analyzed. In the safety-first framework, it is shown that independent background risk only influences an investor's choice between risk-free asset and risky assets as a group, correlated background risk has impact on an investor's choice not only between risk-free asset and risky asset but also on an investor's choice among risky assets, and the impact of background risk on an investor's portfolio selection is related to the safety consciousness of the investor.

EO0686: **A review on semiparametric model averaging for dynamic time series with application to economic risk forecasting**
Presenter: **Zudi Lu**, University of Southampton, United Kingdom
The purpose is to review some recent progress on semiparametric model averaging schemes for nonlinear dynamic time series regression models with a very large number of covariates including exogenous regressors and autoregressive lags. Our objective is to obtain more accurate estimates and forecasts of time series by using a large number of conditioning variables in a nonparametric way. We have proposed several semiparametric penalized methods of Model Averaging MArginal Regression (MAMAR) for the regressors and auto-regressors either through an initial screening procedure to screen out the regressors whose marginal contributions are not significant in estimating the joint multivariate regression function or by imposing an approximate factor modelling structure on the ultra-high dimensional exogenous regressors with principal component analysis used to estimate the latent common factors. In either case, we construct the optimal combination of the significant marginal regression and auto-regression functions to approximate the objective joint multivariate regression function. Asymptotic properties for these schemes are derived under some regularity conditions. Empirical applications of the proposed methodology to forecasting the economic risk, such as inflation risk in the UK, will be demonstrated.

EO130 Room LSK1011 RECURRENT EVENTS Chair: Eric Beutner

EO0406: **A penalized algorithm for event-specific rate models for recurrent events**
Presenter: **Olivier Bouaziz**, University Paris Descartes, France
Co-authors: Agathe Guilloux
Statistical inference is considered for the rate function of a recurrent event process. We study two semi-parametric models stratified with respect to each recurrent event. This allows us to have a different covariate effect on each recurrent event. The first studied model was introduced in the 80ies and has a multiplicative form. The second one is based on the Aalen model introduced in the context of survival data and has an additive form. For reasonable sizes of sample in event-specific models the number of estimated parameters can be very large compared to the number of covariates. In order to remedy to this over-parametrization, a total-variation penalty is used which constrain some of the parameters to be constant. The asymptotic behavior of the penalized estimator is derived. Through a simulation study and analysis of real data, the performance of our estimator is compared with the unconstrained estimator and the Andersen and Gill constant estimator.

EO0413: **Consistent semiparametric estimation with recurrent event data based on virtual age models**
Presenter: **Laurent Doyen**, Univ Grenoble Alpes, France
Co-authors: Eric Beutner, Laurent Bordes
Virtual age models are useful to analyze recurrent events arising in epidemiology (e.g. relapse times of a disease), industry (e.g. repair times of a system), etc. The model consists of a composition of a baseline hazard rate function, characterizing the first time to event distribution, and an effective age function, that allows us to take into account events effects. In nonparametric setting all the existing results assume that the virtual age function is known. This means that events effects are known, what is generally not true in practice especially when the aim is to be able to assess those events effects. One way to overcome this difficulty is to consider semiparametric virtual age models with parametrized virtual age functions. Unfortunately, it is especially difficult to fit these models. Indeed, we show that the usual profile likelihood method fails to lead to consistent estimators. But, the expected consistency property is restored by smoothing the pseudo-estimator of the infinite-dimensional parameter of the model, i.e. the baseline hazard rate. Similar ideas have already been proposed in the literature for the accelerated failure time model. We show that the resulting estimators are asymptotically consistent. Our approach also show that empirical processes tools may be an efficient alternative to martingale methods to study the asymptotic properties of these inference methods.

EO0380: **Regression analysis of mixed and incomplete recurrent event data**
Presenter: **Jianguo Sun**, Jilin University, China
Event history studies occur in many fields including economics, medical studies, and social science. In such studies concerning some recurrent events, two types of data have been extensively discussed in the literature. One is recurrent event data that arise if study subjects are monitored or observed continuously. In this case, the observed information provides the times of all occurrences of the recurrent events of interest. The other is panel count data, which occur if the subjects are monitored or observed only periodically. This can happen if the continuous observation is too expensive or not practical, and in this case, only the numbers of occurrences of the events between subsequent observation times are available. We discuss a third type of data, a mixture of recurrent event and panel count data and present an estimating equation-based approach for regression analysis of the data.

EO0635: **Semiparametric inference with spatially correlated recurrent event data**
Presenter: **Akim Adekpedjou**, Missouri University of Science and Technology, United States
Co-authors: Sophie Dabo
The main driver is the need to analyze recurrent event data in the presence of spatial correlation. Existing models are based on the assumption that the units are independent, and cannot be applied to the situation where the units are subject to spatial correlation. We develop semiparametric models that account for spatial correlation among regions and where, marginally, the gap-time follows a general Cox-type model for recurrent events. Due to the high-dimensional of the integral involved in likelihood construction, the techniques of composite likelihood are used to estimate the regressions coefficients thereby allowing identification of risk factors at onset and further recurrence of events. The estimated parameters turn out to have population interpretation. We show that the inclusion of weights in the composite likelihood improved estimators properties. Asymptotic properties are easily obtained by applying the results of Guyon and some mixing properties. A simulation study and an application to real spatially correlated recurrent esophageal cancer data are provided to illustrate the models.

EO278 Room LSK1007 LEARNING THEORY AND BIG DATA Chair: Yiming Ying

EO0428: **Overlapping sliced inverse regression for dimension reduction**
Presenter: **Qiang Wu**, Middle Tennessee State University, United States
Co-authors: Ning Zhang
Sliced inverse regression (SIR) is a statistical tool for dimension reduction. It identifies the effective dimension reduction space, the subspace of significant factors with intrinsic lower dimensionality. We propose refined implementations of SIR algorithm by allowing slice overlapping. The new algorithms, called overlapping sliced inverse regression (OSIR), can estimate the effective dimension reduction space and determine the number of effective factors more accurately. We show that the overlapping technique codes the information of the differences (or derivatives in the population version) of the inverse regression curve, which helps to explain the superiority of OSIR. We also proved OSIR algorithms are $\sqrt{n}$-consistent and verified the effectiveness of OSIR algorithms by simulations and applications.

EO0563: **Strategies to facilitate access to detailed geocoding information based on synthetic data**
Presenter: **Jingchen Hu**, Vassar College, United States
Results are presented on whether generating synthetic data can be a viable strategy to provide access to detailed geocoding information for external researchers, without compromising the confidentiality of the units included in the database. This research was motivated by a recent project at the Institute for Employment Research (IAB) in Germany that linked exact geocodes to the Integrated Employment Biographies, a large administrative database containing several million records. Based on these data, we evaluate the performance of several synthesizers in terms of addressing the trade-of between preserving analytical validity and limiting the risk of disclosure. We propose strategies for making the synthesizers scalable for such large files, introduce analytical validity measures for the generated data, and provide general recommendations for statistical agencies considering the synthetic data approach for disseminating detailed geographical information

EO0555: **Kaczmarz algorithm in learning theory**
Presenter: **Xuemei Chen**, University of San Francisco, United States
Kaczmarz algorithm is an iterative algorithm that solves linear system of equations. It has got more attention recently due to its randomized version. We will address how the randomized Kaczmarz algorithm compared to some statistical learning algorithms and how the connection can be exploited to improve either subjects.

EO0215: **Convergence of the randomized Kaczmarz algorithm in Hilbert Spaces**
Presenter: **Xin Guo**, The Hong Kong Polytechnic University, Hong Kong
Co-authors: Junhong Lin, Ding-Xuan Zhou
The randomized Kaczmarz algorithm recently draws much attention. Existing results on analysis suffer from condition numbers of the linear equation systems. Although the randomized Kaczmarz algorithm has a natural generalization to Hilbert spaces (which covers online learning algorithms for a particular instance), the existing analysis does not. Despite the large-scale linear equation system is an ideal scenario for the randomized Kaczmarz algorithm to outperform direct solvers, it is also a scenario the existing analysis is not satisfactory. We introduce the regularity assumption widely adopted in learning theory and obtain the polynomial convergence rate of the randomized Kaczmarz algorithm in Hilbert spaces under a noise-free setting. We find that, by nature, the randomized Kaczmarz algorithm converges weakly. Meanwhile, with noisy data, we study the relaxation method and obtain a strong convergence arbitrarily close to the minimax optimal rate. The result applies to online gradient descent learning algorithms and significantly improves the existing learning rate in literature.

EO266 Room LSK1033 ADVANCES IN OPTIMAL PORTFOLIO ALLOCATION AND OPTION PRICING Chair: Simon Kwok

EO0438: **Specification test of multivariate calibrated models with application in option pricing**
Presenter: **Simon Kwok**, University of Sydney, Australia
Co-authors: Robert Jarrow
Despite its wide applications in empirical finance and economics, the exact calibration approach is often misunderstood to be a non-rigorous econometrics technique. We study the problems of estimating and testing the specification of calibrated models in the multivariate framework. Compared with the conventional least squares method, exact calibration may deliver an estimator with smaller asymptotic mean squared errors under certain model misspecifications. Furthermore, individual parameter estimates from exact calibration are informative for model diagnostics as they provide hints in identifying misspecifications along parameter directions that are economically interpretable. This is illustrated in simulation studies and an empirical application in option pricing.

EO0484: **A generalized measure for optimal portfolio selection**
Presenter: **Udi Makov**, University of Haifa, Australia
Co-authors: Zinoviy Landsman, Tomer Shushi
A novel class of utility functions applied to optimal portfolio selection is offered. This class incorporates as special cases important measures such as the mean-variance, Sharpe ratio, mean-standard deviation and others. In addition, a condition is provided for the existence of a one-to-one correspondence between the parameter of this class of utility function and the trade off parameter λ in the mean-variance utility function. We illustrate our results by taking a portfolio of stocks from NASDAQ.

EO0423: **A flexible generalised hyperbolic option pricing model and special cases**
Presenter: **Boris Choy**, University of Sydney, Australia
Co-authors: Simon Kwok, Claudia Yeap
A flexible generalised hyperbolic (GH) option pricing model is formulated, which has all four of its parameters free to be estimated. We also present six three-parameter special cases: a variance gamma (VG), t, hyperbolic, normal inverse Gaussian, reciprocal hyperbolic and normal reciprocal inverse Gaussian option pricing model. Using S&P 500 Index options, we compare the flexible GH, VG, t and Black-Scholes models. The flexible GH model offers the best out-of-sample pricing results and between the two three-parameter models, the t model outperforms the VG model for both in-sample and out-of-sample pricing. All three models ameliorate the Black-Scholes model's implied volatility biases.

EO0518: **On the matrix copula with application to portfolio analysis**
Presenter: **Nuttanan Wichitaksorn**, Thailand Development Research Institute, Thailand
A new matrix-copula construction is presented, and an application in finance is shown. The matrix copula is an implicit copula such that its construction is based on the associated matrix-variate distribution. The matrix copula based on the Gaussian distribution is presented to illustrate the formulation and estimation. The major and important feature of this matrix copula construction is that it embodies two dependence structures, among-row and among-column, that can be conveniently estimated via Markov chain Monte Carlo methods. As an application, we propose a new approach to analyze monthly stock loss data, whereby the model construction is based primarily on the so-called insurance collective risk model. The two dependence structures, for loss sizes and numbers of days in loss for stocks in a portfolio, are then modeled using the matrix Gaussian copula. Simulation and empirical studies are implemented to illustrate the performance and applicability of the proposed model and method.

EO236 Room LSK1014 THEORY AND NUMERICS IN ESTIMATING STOCHASTIC PROCESSES Chair: Hiroki Masuda

EO0442: **Model selection for stochastic differential equations in YUIMA package**
Presenter: **Shoichi Eguchi**, Kyushu University, Japan
There are several studies of model selection for stochastic differential equations (SDEs), which includes the contrast-based information criterion for ergodic diffusion processes (CIC) and the Schwarz type information criterion for locally asymptotically quadratic models (BIC, quasi-BIC). Based on these studies, we create the model selection function for SDEs in R package yuima. In particular, this function can calculate CIC, BIC and quasi-BIC for each candidate model. We will first overview the model selection methods for SDEs and then explain the specification of the model selection function. Some model selection examples are given in order to show how to use the function.

EO0506: **Capturing heterogeneous lead-lag relationships from ultra high frequency data**
Presenter: **Yuta Koike**, Tokyo Metropolitan University, Japan
A new framework for modeling lead-lag relationships in high frequency financial markets is proposed. The model can be accommodated to non-synchronous trading and market microstructure noise as well as the intraday heterogeneity of the lead-lag relationships, which are essential for empirical applications. A simple statistical methodology for analyzing the proposed model is presented as well. In particular, we can conduct a statistical test to detect the presence of such a lead-lag relationship. The methodology is illustrated by an empirical study to detect lead-lag relationships in multi-market trading.

EO0305: **Parametric inference for ruin probability under Levy insurance risks**
Presenter: **Yasutaka Shimizu**, Waseda University, Japan
The classical insurance ruin theory and its related field can revive interest in recent Enterprise Risk Management (ERM) because the theory gives us many tools for the dynamic risk management. A central issue in this context is estimating ruin probability under certain spectrally negative jump processes. Under a parametric assumption for the claim process, it is not so hard to construct an asymptotically normal estimator of ruin probability via the delta method given an asymptotically normal estimator of unknowns. However, the asymptotic variance of the estimator includes the derivative of the ruin probability with respect to the parameters, which is not easy to compute. To construct a confidence interval, we will give an approximation for the derivatives under a large initial surplus, and gives an approximated confidence interval.

EO0320: **Hybrid type estimation for ergodic diffusion processes based on reduced data**
Presenter: **Masayuki Uchida**, Osaka University, Japan

Parametric estimation is treated for both drift and volatility parameters of ergodic diffusion processes from discrete observations. As an initial estimator, we first obtain the Bayes type estimator based on reduced data, which has a non-optimal rate of convergence. Next, we compute the adaptive maximum likelihood (ML) type estimator with the initial Bayes type estimator, which is called the hybrid type estimator. In order to show the asymptotic normality and convergence of moments of the hybrid type estimator, the Ibragimov-Hasminskii program and the polynomial type large deviation inequality for the statistical random field are applied to the Bayes type estimation and the adaptive ML type estimation. We give an example and simulation results on the finite sample behavior of the initial Bayes type estimator and the hybrid type estimator from the viewpoint of numerical analysis.

EO276 Room LSK1032 FINANCIAL ECONOMETRICS Chair: Tommaso Proietti

EO0479: **Seasonal adjustment of daily data with JDemetra+: First results**
Presenter: **Dominique Ladiray**, INSEE, France
Co-authors: Jean Palate, Gian Luigi Mazzi, Tommaso Proietti

The progresses in information technology and survey methods have fostered the availability of daily and weekly time series. The seasonal adjustment of high-frequency time series poses several challenges. First of all the seasonal period of the annual cycle is neither constant nor an integral. Secondly, in order to accommodate complex seasonal patterns many individual effects might be required. Thirdly, the need for robust methods is reinforced by the fact that the effects of outlying observations is not smoothed by temporal aggregation and that they are relatively more frequent. The robustness can be enforced either by an outlier detection procedure or by robust filtering methods.Moving average methods (X-13ARIMA-SEATS) and ARIMA-based methods (TRAMO-SEATS) are recommended by Eurostat and are implemented in JDemetra+, the European software for seasonal adjustment. These methods only deal with monthly and quarterly series; they estimate first the outliers and calendar effects using a Reg-ARIMA model and then decompose the residual of the model into trend-cycle, seasonality and irregular component The first implementation of a Tramo-Seats algorithm and a X-12 algorithm to seasonally adjust high frequency data will be presented. We will show that these two methods can be extended to correct a series for multiple seasonality.

EO0464: **Time series seasonal adjustment using regularized singular value decomposition**
Presenter: **Wei Lin**, Capital University of Economics and Business, China
Co-authors: Jianhua Huang, Tucker McElroy

A new seasonal adjustment method is proposed based on the regularized singular value decomposition (RSVD) of the matrix obtained by reshaping the seasonal time series data. The method is flexible enough to capture two kinds of seasonality: the fixed seasonality that does not change over time and the time-varying seasonality that varies from one season to another. RSVD represents the time-varying seasonality by a linear combination of several seasonal patterns. The right singular vectors capture multiple seasonal patterns, and the corresponding left singular vectors capture the magnitudes of those seasonal patterns and how they change over time. By assuming the time-varying seasonal patterns change smoothly over time, the RSVD uses penalized least squares with a roughness penalty to effectively extract the left singular vectors. The proposed method applies to seasonal time series data with a stationary or nonstationary non-seasonal component. The method also has a variant that can handle that case that an abrupt change (i.e., break) may occur in the magnitudes of seasonal patterns. Our proposed method compares favorably with the state-of-art X-13ARIMA-SEATS method on both simulated and real data examples.

EO0620: **Testing stress**
Presenter: **Sulkhan Chavleishvili**, European Central Bank, Germany
Co-authors: Manfred Kremer

Financial stress can impair the regular functioning of certain financial market segments or of the financial system as a whole, often associated with adverse macroeconomic consequences. Although the concept of financial stress is central for macroeconomic policy interventions, it still remains unclear how it can be rigorously measured. A novel statistical framework is introduced for defining and testing financial stress. The framework can be flexibly applied to an arbitrary number of time series measuring stress in different economic units (like corporations) or different financial market segments. Our semiparametric framework defines financial stress as the event that the relative rank of a realized time-series observation (or the vector of ranks in the general multivariate setup) exceeds a particular non-stress level. The estimation of ranks is cast in a simple linear regression framework, and the asymptotic properties of the estimator are used to develop a test statistic for the null hypothesis of no-stress. We interpret this test statistic as a meaningful and statistically well-founded financial stress index. We apply our statistical framework to data for the US financial system from 1973 to 2016 and estimate, among other things, the effects which shocks in our financial stress index may have on different measures of real economic activity.

EO0615: **CISS in a time-varying environment: On the frequency of systemic distress**
Presenter: **Manfred Kremer**, European Central Bank, Germany
Co-authors: Eddie Gerba

The CISS indicator is a composite indicator of systemic stress in the financial system, which is highly time-varying. We incorporate this index in multiple macro-financial time-varying parameter VAR models of the euro area to evaluate its performance and measure its predictability of systemic events. Only during times of financial distress, the indicator increases strongly. Moreover during systemic events such as the post-2007 crises both the indicator as well as the uncertainty regarding the future path of key macro-financial variables increases in our model. This increase in the indicator precedes the increase in volatility of other variables in the model by 1-2 quarters, making it a good forecaster of crises events. Furthermore, systemic shocks (or shocks to the CISS indicator) have a stronger impact on the rest of the economy than any of the other shocks examined. Lastly we compare our model to a Markov-switching framework (ceteris paribus) and find that while the number of regime switches are fewer, increases in the mean and volatility of coefficients associated with each crisis event is higher in our current model.

Saturday 17.06.2017 08:30 - 09:50 Parallel Session K – EcoSta2017

EO150 Room LSKG001 ADVANCED GRAPHICAL AND COMPUTATIONAL METHODS Chair: Garth Tarr

EO0306: **Computationally intensive methods for model selection using the mplot R package**
Presenter: **Kevin Murray**, University of Western Australia, Australia
Model selection techniques have existed for many years however, simple, clear and effective methods of visualizing the model building process are sparse. Computationally intensive graphical methods are described which assist in the selection of models and comparison of many different selection criteria. Specifically, we describe for generalized linear models, how to visualize measures of description loss and of model complexity to facilitate the model selection dilemma. The use of the weighted bootstrap with exponential weighting is advocated to assess the stability of selected models and to enhance our graphical tools. Variable inclusion plots are demonstrated to assist in the determination of important variables. These methods are described and implemented using the mplot R Package and recent developments in this package are also discussed. Results using case studies demonstrate how these proposed tools are useful to learn more about important variables in the data and how these tools can assist the understanding of the model building process.

EO0161: **Sparse estimation of multi-class vector autoRegressive models**
Presenter: **Ines Wilms**, KU Leuven, Belgium
Co-authors: Luca Barbaglia, Christophe Croux
Vector AutoRegressive (VAR) models form a special case of multivariate regression models in that the response variables are observed over time and modeled as a function of their own past values. Assume we have K VAR models for K distinct but related classes. We jointly estimate these K VAR models to borrow strength across classes and to estimate multiple models that share certain characteristics. Our methodology encourages corresponding effects to be estimated similar across classes, while still allowing for small differences between them. Moreover, we focus on multi-class estimation of high-dimensional VAR models, i.e. models with a large number of time series relative to the time series length. Therefore, our estimate is sparse: unimportant effects are estimated as exactly zero, which facilitates the interpretation of the results. We consider a marketing application of the proposed methodology.

EO0581: **Cluster feature selection in high dimensional linear models**
Presenter: **Zhen Pang**, The Hong Kong Polytechnic University, Hong Kong
The focus is on variable screening when highly correlated variables exist in high dimensional linear models. We propose a novel cluster feature selection procedure based on the elastic net and linear correlation variable screening to enjoy the benefits of the two methods. When calculating the correlation between the predictor and the response, we consider the highly correlated group of the predictors instead of the individual ones. This is in contrast to the usual linear correlation variable screening. Within each correlated group, we apply the elastic net to select and estimate the variables. This avoids the drawback of mistakenly eliminating true non-zero coefficients for highly correlated variables like LASSO does. After our procedure, maximum absolute sample correlation coefficient between clusters becomes smaller and any common model selection methods like SIS or LASSO can be applied to improve the results. Extensive numerical examples including pure simulation examples and semi-real examples are conducted to show the good performances of our procedure.

EO0489: **Species distribution modelling for combined data sources: New extensions**
Presenter: **Ian Renner**, University of Newcastle, Australia
Co-authors: Olivier Gimenez
Species distribution models, which related occurrence information about species to environmental variables believed to influence their distributions, are fundamental tools in statistical ecology. A variety of methods have arisen in response to the variety of types of occurrence data available. As species occurrence information becomes more widely available, there is a need to develop methods which combine such occurrence information from multiple types and sources of data. Results will be presented from a model for the Eurasian lynx in the Jura Mountains of France which combines a set of repeated survey data from camera traps with two sources of "presence-only data", whereby reported observations of lynx are not balanced by any information about lynx absences. The model extends the approach of combining the likelihood expressions of occupancy and point process models by incorporating a lasso penalty to boost predictive performance and source weights to acknowledge differences in data quality.

EC286 Room LSK1033 CONTRIBUTIONS IN BAYESIAN ECONOMETRICS Chair: Jiri Witzany

EC0614: **Bayesian hierarchical model for technical efficiency using a stochastic frontier production function**
Presenter: **Ying-Ying Lee**, University of Cincinnati, United States
Co-authors: Myoungjin Jung, Chansoo Kim, David Yi, Younshik Chung
A stochastic frontier production function has been considered in the case that there is the effect of non-negative technical efficiency in the data. In the past decades, many studies have been discussed to determine the explanatory variables which affect the technical efficiency in the stochastic frontier production function. We consider a hierarchical model of the stochastic frontier production function to investigate data with multiple hierarchical structures. It turns out that the proposed model naturally gives us a dependent covariance structure within a sub-group and the independence between sub-groups. We consider Bayesian analysis to estimate the model parameters in the model as well as the technical efficiency using Markov Chain Monte Carlo (MCMC) methods.

EC0745: **Robust Bayesian exponentially tilted empirical likelihood**
Presenter: **Zhichao Liu**, Monash University, Australia
Co-authors: Catherine Forbes
A new robust Bayesian exponentially tilted empirical likelihood (RBETEL) inferential methodology is proposed which is suitable for moment condition models for data that may be contaminated by outliers. The foundations are on the original Bayesian exponentially tilted empirical likelihood (BETEL) method, justified by the fact that an empirical likelihood can be interpreted as the nonparametric limit of a Bayesian procedure when the implied probabilities are obtained from maximizing entropy subject to some given moment constraints. After first demonstrating through a simulation exercise that BETEL posteriors are susceptible to interference from outliers, a Markov chain Monte Carlo framework is developed, incorporating outlier information derived from a robust frequentist estimator of multivariate location and scatter. Controlled simulation experiments are conducted to investigate the performance of the proposed RBETEL method, and finds that the new approach improves on BETEL when outliers are present.

EC0730: **Partially censored posterior for accurate left tail density prediction**
Presenter: **Agnieszka Borowska**, Vrije Universiteit Amsterdam, Netherlands
Co-authors: Lennart Hoogerheide, Siem Jan Koopman
A novel approach to inference for a specific region of the predictive distribution is introduced. An important domain of application is accurate prediction of financial risk measures, where the area of interest is the left tail of the predictive posterior density of (log)returns. It originates

from the Bayesian approach to parameter estimation and time series forecasting, however it provides a more accurate estimation of the density in the region of interest in case of misspecification. In the proposed concept of the partially censored posterior the set of parameters is partitioned into two subsets: the first, for which we consider the standard marginal posterior, and the second, for which we consider a censored conditional posterior. The latter means that observations outside the region of interest are censored: for those observations only the probability of being outside the region of interest matters. In the second subset we choose parameters that are expected to benefit from censoring. This approach yields more precise parameter estimation than a fully censored posterior for all parameters, and has more focus on the region of interest than a standard approach. An extensive simulation study shows the ability of the introduced method to outperform standard approaches.

EC0505: **A Bayesian approach to backtest overfitting**
Presenter: **Jiri Witzany**, University of Economics in Prague, Czech Republic

Quantitative investment strategies are often selected from a broad class of candidate models estimated and tested on historical data. Standard statistical technique to prevent model overfitting such as out-sample back-testing turns out to be unreliable in the situation when selection is based on results of too many models tested on the holdout sample. There is an ongoing discussion how to estimate the probability of back-test overfitting and adjust the expected performance indicators like Sharpe ratio in order to reflect properly the effect of multiple testing. We propose a consistent Bayesian approach that consistently yields the desired robust estimates based on an MCMC simulation. The approach is tested on a class of technical trading strategies where a seemingly profitable strategy can be selected in the naive approach.

EC292 Room LSKG007 CONTRIBUTIONS IN FORECASTING Chair: Danilo Leiva-Leon

EC0709: **Forecasting the ex-vessel price of Canadian snow crab using a VARMAX approach**
Presenter: **Liang Chen**, University of Calgary, Canada

The fluctuation of ex-vessel prices of Canadian snow crab not only affect the export revenues of it, which is important for the Canadian fishery economy, but also influence the income and well-being of Canadian fishermen. As a result, building a reliable forecasting model that is able to predict the price trend of Canadian snow crab is of great significance as it helps policymakers to better manage the snow crab fishery industry. Using monthly data from December 1991 to December 2013, the forecasting model is based on the vector ARMAX (VARMAX) model rather than the vector error correction (VEC) model since the endogenous variables are either level stationary or trend stationary, which precludes the cointegrating relationship among the endogenous variables. Overall, the forecasting results based on the VARMAX(1,1,1) models are quite promising and they are able to capture the general trend and most of turning points in the ex-vessel price series. In addition, the VARMAX-GARCH model is also estimated due to the presence of GARCH errors and to capture the time-varying volatility of the ex-vessel price series. Expectedly, VARMAX-GARCH model produces even better forecasting results and smaller (95%) confidence intervals of forecasts.

EC0811: **Automatic ARIMA modeling and its application**
Presenter: **Dedi Rosadi**, Universitas Gadjah Mada, Indonesia

In some application of time series modelling, it is necessary to obtain forecast of various types of data automatically and possibly, in real-time. For instances, to forecast large number of univariate series every day, or to do a real-time processing of the satellite data. Various automatic algorithms for modeling ARIMA models are available in the literature, where here we will discuss three methods in particular. One of the method is based on a combination between the best exponential smoothing model to obtain the forecast, together with state-space approach of the underlying model to obtain the prediction interval. The second method, which is more advanced method, is based on X-13-ARIMA-SEATS, the seasonal adjustment software by the US Census Bureau. Last method is a heuristic method based on genetic algorithm approach. These approaches are implemented in our R-GUI package RcmdrPlugin. Econometrics which now already integrated in our new and more comprehensive R-GUI package RcmdrPlugin.SPSS. We provide application of the methods and the tool using real data.

EC0213: **Markov-switching three-pass regression filter**
Presenter: **Danilo Leiva-Leon**, Banco de España, Spain

A new approach is introduced for the estimation of high-dimensional factor models with regime-switching factor loadings by extending the linear three-pass regression filter to settings where parameters can vary according to Markov processes. The new method, denoted as Markov-Switching three-pass regression filter (MS-3PRF), is suitable for datasets with large cross-sectional dimensions since estimation and inference are straightforward, as opposed to existing regime-switching factor models, where computational complexity limits applicability to few variables. In a Monte Carlo experiment, we study the finite sample properties of the MS-3PRF and find that it performs favorably compared with alternative modelling approaches whenever there is structural instability in factor loadings. As empirical applications, we consider forecasting economic activity and bilateral exchange rates, finding that the MS-3PRF approach is competitive in both cases.

EC0813: **GARCH model for quantile income time series data and forecasting income inequality**
Presenter: **Haruhisa Nishino**, Hiroshima University, Japan

A lot of attention has been paid to income inequality in the Japanese economy as well as in the world economy. The aim is to analyze dynamic properties of the income inequality in Japan by using GARCH model. A parametric model and a Lognormal distribution are assumed. The lognormal distribution is better fitted to Japanese income data and useful for extracting inequality from income data because its scale parameter only represents inequality. A GARCH model is proposed which includes income inequality for income time series data. The GARCH model is suitable for modelling the scale parameter of the lognormal distribution. A joint distribution from selected order statistics enables us to construct a quasi-likelihood of the GARCH model from quantile income data, which is used for estimation. The proposed model has an inequality structure and a time series structure. It is useful for analyzing persistent income inequality and forecasting income inequality. The models are estimated from Japanese quantile income data (Family Income and Expenditure Survey by Ministry of Internal Affairs and Communications, Statistics Bureau, Japan) and the various GARCH models are compared and examined from the point of view of forecasting.

EC294 Room LSK1005 CONTRIBUTIONS IN STATISTICAL MODELLING Chair: Yao Rao

EC0764: **Abnormal events detection by a real-time surveillance: The case of influenza outbreaks**
Presenter: **Yao Rao**, The University of Liverpool, United Kingdom

A surveillance method is introduced by using deviations from probabilistic forecasts by comparing realised observations with probabilistic forecasts. The deviation metric is based on low probability events. Specifically, the problem of syndromic surveillance for influenza (flu) is addressed with the intention of detecting outbreaks, due to new strains of viruses, over and above the normal seasonal pattern. The syndrome is hospital admissions for flu-like illness and hence the data are low counts. In accordance with the count properties of the observations, an integer valued autoregressive process is used to model flu occurrences. Monte Carlo evidence suggests the method works well in stylised but somewhat realistic situations. The model is applied to real flu data and the application indicates that the model estimated on a short run of training data, did not declare false alarms, when used with new observations deemed in control, ex post and the model easily detected the 2009 H1N1 outbreak. Finally, we attempt to predict the timing of the flu season using multi-step ahead probability distribution forecasts.

EC0772: **Estimation of the population size of big data sets**
Presenter: **Jeffrey Chu**, University of Manchester, United Kingdom

In many areas of science, problems often arise which require the analysis of very large or very small data sets. For example, in the analysis and modelling of hard-to-reach populations - i.e. social networks, computer network traffic data, satellite and wireless communications/infrastructure etc. Often one of the first tasks in this type of analysis is to calculate the population size of such groups, whether it be the number of individuals, users, or some other quantity. However, more often than not, it is likely that we do not know the exact size of this quantity, thus, the figure must be estimated. We propose a population estimation method for big data sets based on (but not restricted to) random walk sampling and collisions, and compare it to two estimators known in the big data literature. Existing population estimation methods fail to deal with some of the consequences of random walk sampling leading to estimators with major restrictions such as a minimum number of collisions or a minimum sample size. The comparison is based on a simulation study and two real data applications (Twitter and LiveJournal social network data). The proposed estimator is shown to outperform the known estimators at least for small sample sizes.

EC0800: **Statistical inference for misspecified ergodic Levy driven stochastic differential equation models**
Presenter: **Yuma Uehara**, Kyushu University, Japan

The parametric estimation problem of misspecified ergodic Levy stochastic differential equation models under high frequency samples is considered. Our estimation method is based on a widely applicable and tractable Gaussian quasi likelihood which focuses on mean and variance structure. To handle misspecification effect, we invoke the theory of extended Poisson equation and give the properties of the solutions. Our result is that the Gaussian quasi maximum likelihood based estimator of drift and scale parameters have asymptotic normality, and that its convergence rate is the same as the correctly specified case.

EC0802: **Tail estimation for the cross-spectral density of a bivariate stationary Gaussian random field**
Presenter: **Joonho Shin**, Department of Statistics, Seoul National University, Korea, South

Multivariate stationary Gaussian random fields are widely used to fit multivariate spatial data. The one to one correspondence between (cross-)covariance functions and (cross-)spectral densities allows us to model (cross-)spectral densities instead of (cross-)covariance functions. We consider bivariate stationary Gaussian random field model. Under some assumptions on high-frequency behavior of (cross-)spectral densities, we introduce approach to estimate parameters that control tail behaviors by minimizing a modified version of multivariate Whittle likelihood objective function. We show consistency and asymptotic normality of the estimators with simulation results.

EG129 Room LSK1007 CONTRIBUTIONS IN SURVIVAL ANALYSIS Chair: Yan Shen

EC0814: **Comparison of hazard rates for paired and non censored data**
Presenter: **Felix Belzunce**, Universidad de Murcia, Spain
Co-authors: Carolina Martinez-Riquelme

The comparison of failure or hazard rates is a well-known topic in reliability and survival analysis. Recently, a new criteria to compare hazard rates for dependent random variables has been introduced and it is known as the weak joint hazard rate order. This criteria is useful not only in reliability and survival analysis but also in portfolio selection problems. Non parametric methods are presented for testing that two (dependent) random variables can be compared in terms of the weak joint hard rate order in the case of paired and non censored data.

EC0817: **Simultaneous phase-II monitoring of both scale and shape parameters of the Weibull distribution**
Presenter: **Min Xie**, City Univ of Hong Kong, Hong Kong
Co-authors: Amitava Mukherjee, Min Gong

The Weibull distribution is one of the most important probability models used in Statistics and Engineering apart from various other applied fields in the present era. It is a well-known life time distribution, and also used in modelling time between events, particle-sizes among others. Therefore, efficiently monitoring the parameters of a Weibull distribution is an important applied problem. Recently, a number of schemes for jointly monitoring the mean and variance of a normally distributed process using a single plotting statistic have been developed. The joint monitoring of the location and scale parameters of a two-parameter (shifted) exponential distribution is also available in the literature. Despite significant advantages of simultaneous monitoring, in general, there is still a dearth of researches in joint monitoring of multi-parameter non-normal distributions. Noting that the Weibull distribution is widely studied from various angles in the context of reliability estimation and life testing, in this context, we develop several joint monitoring schemes for the scale and shape parameters of a two-parameter Weibull distribution. We discuss the implementation procedures and study performance properties of various proposed schemes. We also offer an illustrative example along with a summary and recommendations.

EC0344: **Variable selection in proportional hazards cure model with time-varying covariates, application to bank failures**
Presenter: **Alessandro Beretta**, HEC Liege, Belgium
Co-authors: Cedric Heuchenne

In the last three decades, as a consequence of failures and corporate actions, the number of commercial banks in the United States has shrunk by two thirds. Empirical evidence in the analysis of bank failures suggests the existence of banks which are not susceptible to default. For this reason, we use a semi-parametric proportional hazards cure model with time-varying covariates to study their effects either on the probability that a bank is susceptible to default and on the survival time of failed institutions. We propose a penalized maximum likelihood method for the selection of the most significant variables, using a Smoothly Clipped Absolute Deviation (SCAD) penalty. A simulation study shows that this procedure performs reasonably well. We apply this methodology to a quite large sample of United States commercial banks insured by the Federal Deposit Insurance Corporation (FDIC) and with more than 50 million dollars of total assets during the last quarter of 2002. More in detail, we use bank-specific covariates observed on a quarterly basis until the end of 2015, that we use as proxies for capital adequacy, asset quality, earnings, management efficiency, liquidity, cost structure and size.

EC0793: **Improved likelihood inferences for Weibull regression model**
Presenter: **Yan Shen**, Xiamen University, China

A general procedure is developed for bias-correcting the maximum likelihood estimators (MLEs) of the parameters of Weibull regression model with either complete or right-censored data. Following the bias correction, variance corrections and hence improved t-ratios for model parameters are presented. Potentially improved t-ratios for other reliability-related quantities are also discussed. Simulation results show that the proposed method is effective in correcting the bias of the MLEs, and the resulted t-ratios generally improve over the regular t-ratios.

EC289 Room LSK1009 CONTRIBUTIONS IN MULTIVARIATE METHODS Chair: Marc Hallin

EC0706: **Overlapped groupwise dimension reduction**
Presenter: **Jingke Zhou**, Ningbo University, China
Existing groupwise dimension reduction requires given group structure to be non-overlapped. This confines its application scope. We aim at groupwise dimension reduction with overlapped group structure or even unknown group structure. To this end, existing groupwise dimension reduction concept is extended to be compatible with overlapped group structure. Then, the envelope method is ameliorated to deal with overlapped groupwise dimension reduction. As an application, Gaussian graphic model is employed to estimatethe structure between predictors when the group structure is not given, and the amended envelope method is used for groupwise dimension reduction with graphic structure. Furthermore, the rationale of the proposed estimation procedure is explained at the population level and the estimation consistency is proved at the sample level. Finally, the finite sample performance of the proposed methods is examined via numerical simulationsand a body fat data analysis.

EC0752: **A joint model of firm failure and credit ratings**
Presenter: **Laura Vana**, WU Wirtschaftsuniversitaet Wien, Austria
Co-authors: Rainer Hirk, Kurt Hornik
Credit risk modeling including the measurement of credit quality has been intensively investigated by academics and practitioners over the past decades. The aim is to contribute to this field by developing a framework for jointly modeling firm failures (e.g., bankruptcies) and ordinal credit ratings as outcomes. This model, unlike prior work, simultaneously incorporates failures and credit ratings and allows inference about the quantitative relationships between these outcomes by simultaneously making use of both sources of information. In addition the model does not require a balanced data set in the outcome variables, i.e., missing values in the outcomes are possible. We hypothesize that 1) the model alleviates the 'low failure portfolio' problem and outperforms univariate models of failure or credit rating models in terms of prediction accuracy; 2) useful insights into rating heterogeneity are gained; 3) information about the systematic rating patterns of the credit rating agencies is uncovered. The joint model will be formulated such that it takes the ordinal nature of the credit ratings into account and can therefore be incorporated in the class of multivariate ordinal models. Failure, firm-level and stock price data for publicly traded US companies as well as issuer credit ratings from the big three rating agencies (S&P, Moody's and Fitch) are collected and analyzed to illustrate the proposed framework.

EC0338: **Testing structural changes and change point estimation in panel data without boundary issue**
Presenter: **Michal Pesta**, Charles University, Faculty of Mathematics and Physics, Czech Republic
Co-authors: Barbora Pestova
Panel data of our interest consist of a moderate number of panels, while the panels contain a small number of observations. An estimator of common breaks in panel means without a boundary issue for this kind of scenario is proposed. In particular, the novel estimator is able to detect a common break point even when the change happens immediately after the first time point or just before the last observation period. Another advantage of the elaborated change point estimator is that it results in the last observation in situations with no structural breaks. The consistency of the change point estimator in panel data is established. The results are illustrated through a simulation study. As a by-product of the developed estimation technique, a theoretical utilization for correlation structure estimation, hypothesis testing, and bootstrapping in panel data is demonstrated. A practical application to non-life insurance is presented as well.

EC0180: **Testing Goodwin with a stochastic differential approach: The United States (1948-2015)**
Presenter: **Florent Mc Isaac**, University Paris Pantheon Sorbonne, France
The aim is to follow previous research in testing both Goodwin's predator-prey model and a further extension. A guideline is provided for the estimation and the backtesting strategy that can be applied to such a class of continuous-time macroeconomic model. The goal is to propose and test stochastic differential equations for Goodwin's model and one of its extension by using an estimation technique based on simulated maximum likelihood. The data considered here is that of wage share and employment rate in the United States from 1948:Q1 to 2015:Q4. Results show that models with two structural breaks and endowed with a CES production technology more accurately explains the behavior captured by this data than the Goodwin Leontief production function. These results are partially confirmed by a backtesting strategy which highlights the forecasting property of the Goodwin model on the considered data. Both the estimation and backtesting strategies can be used to assess the empirical improvement on any extension of the Goodwin model.

EG061 Room LSK1003 CONTRIBUTIONS ON TIME SERIES MODELING AND ITS APPLICATIONS Chair: Thomas Beissinger

EC0718: **On core inflation measures: Evidence from the European Union countries**
Presenter: **Aleksandra Halka**, Narodowy Bank Polski, Poland
Co-authors: Grzegorz Szafranski
The deep and long lasting decreases in consumer inflation have been observed since 2012 throughout many advanced and emerging economies. The negative headline inflation data in several economies are accompanied with core inflation recording the historical lows, which brings the questions about the sources of inflation decline as well as about the changes in the long-run level of inflation. To answer these questions we decompose the consumer inflation in several European economies into permanent (trend inflation) and transitory component extending a previous unobserved component model. In this context we investigate the role of the core inflation measures. We find that the core inflation measures that excludes the energy and food components excludes a significant part of the permanent component and for that reason is not the best approximation of the underlying inflation in particular in the period of the long-lasting commodity shocks. We evidence relatively strong correlation between core inflation and output gap. This result supports the hypothesis that the core inflation may to some extent reflect the changes in demand pressure in the economy. We also document that, the proper identification of sources of inflation variability may be also useful in predicting future inflation.

EC0815: **The importance of price and non-price competitiveness for Euro countries' exports: An unobserved components model**
Presenter: **Thomas Beissinger**, University of Hohenheim, Germany
Co-authors: Nathalie Chusseau, Joel Hellier, Martyna Marczak
Many studies explain the volume of exports by foreign demand and an indicator for the real exchange rate using an error-correction model that allows to distinguish between short-run effects and the long-run relationship between variables. However, if integrated variables are omitted, the estimated long-run relationship may be seriously misspecified. Such an omitted integrated variable is non-price competitiveness that matters for exports but is difficult to measure and therefore mostly ignored in empirical studies. To overcome this problem, an error-correction model is combined with an unobserved component model that allows to model non-price competitiveness as a time-varying trend. To get reliable estimates for all 19 countries of the Euro area separately, stochastic cyclical movements, outliers and structural breaks are taken into account as well. It is also analysed whether the role of foreign demand, price and non-price competitiveness in explaining export demand has changed in the wake of the financial crisis.

EC0747: **A data-cleaning augmented Kalman filter for robust estimation of state space models**
Presenter: **Martyna Marczak**, University of Hohenheim, Germany
Co-authors: Tommaso Proietti, Stefano Grassi
A robust augmented Kalman filter is presented that extends a previous the data-cleaning filter to the general state space model featuring nonstationary and regression effects. The robust filter shrinks the observations towards their one-step-ahead prediction based on the past, by bounding the effect of the information carried by a new observation according to an influence function. When maximum likelihood estimation is carried out on the replacement data, an M-type estimator is obtained. The performance of the robust AKF is investigated in two applications using as a modeling framework the basic structural time series model, a popular unobserved components model in the analysis of seasonal time series. First, a Monte Carlo experiment is conducted in order to evaluate the comparative accuracy of the proposed method for estimating the variance parameters. Second, the method is applied in a forecasting context to a large set of European trade statistics series.

EC0819: **State space modelling of long memory seasonal Gegenbauer processes: A final analysis**
Presenter: **Gnanadarsha Dissanayake**, University of Sydney Australia, Australia
Co-authors: Gnanadarsha Dissanayake, Shelton Peiris, Tommaso Proietti
An approximation of a Seasonal Gegenbauer autoregressive moving average (GARMA) process with long memory using a finite order moving average (MA) representation is considered. The state space form of the MA approximation is developed and the corresponding estimates are obtained by pseudo maximum likelihood using the Kalman filter. For comparative purposes, the same exercise is executed with an autoregressive (AR) approximation. Using an extensive Monte Carlo experiment, the optimal order of the chosen MA approximation is established, and found it was not very large (around 45) and rather insensitive to the sample size. Further evidence suggests that the approximation is reliable for forecasting and signal extraction with periodic long memory components. A rolling forecasting experiment was performed to validate the choice of optimal order of both AR and MA approximations in terms of predictive accuracy. Finally, the proposed methodology was applied to two yearly sunspots time series and the El Nino time series and compared with corresponding results proposed in the literature.

EC284 Room LSK1001 CONTRIBUTIONS IN TIME SERIES Chair: Jean Marc Bardet

EC0762: **Spatio-temporal GARCH models**
Presenter: **Sondre Holleland**, University of Bergen, Norway
Co-authors: Hans Arnfinn Karlsen
In a world where access of data is a diminishing issue, data with a space and time stamp has become very common. Having methodology for analysing such data is fundamental and therefore spatio-temporal models have become a hot topic of research. The model is mainly designed to handle temporal volatility and spatial dependency. For the spatial dependency structure, a closest neighbourhood principle is adopted. The formulation of the model enables generalization of theory from univariate GARCH and is denoted as spatio-temporal GARCH. A circular spatial assumption is applied in order to make the spatial area finite and the process stationary. Circular space means that the spatial boundaries are tied together through the neighbourhood set-up. Compared to related multivariate GARCH models this one is substantially simpler in several aspects and in particular it has fewer parameters. The largest distinction is the specific spatial dependency structure with circular boundaries. The model will be introduced and simulations from the model and parameter estimation procedures will be presented, along with asymptotic results for the Gaussian quasi maximum likelihood estimators of this model.

EC0769: **Multivariate Poisson autoregression**
Presenter: **Baard Stoeve**, University of Bergen, Norway
Co-authors: Paul Doukhan, Konstantinos Fokianos, Dag Tjoestheim
Modelling and inference for multivariate count time series data is studied. There are several approaches reported in the literature, but we focus on appropriate extensions of univariate generalized linear models which have been found quite useful in applications. The focus is on linear and log-linear models. For studying the properties of such processes we develop a novel conceptual framework which is based on copulaes. However, our approach does not impose the copula on a vector of discrete random variables; instead the joint distribution of the responses is determined by imposing a copula function to a vector of continuous random variables. This specific construction avoids conceptual difficulties associated with the joint distribution of discrete random variables, yet it keeps the properties of the Poisson process marginally. We employ Markov chain theory and the notion of weak dependence to study ergodicity and stationarity of the models we consider. We obtain easily verifiable conditions for both linear and log-linear models under both theoretical frameworks. Suitable estimating equations are suggested and studied in detail for estimating unknown parameters. The large sample properties of the resulting estimators are studied in detail. Some simulations and a real data example are shown.

EC0190: **Asymptotic behavior of the Laplacian quasi-maximum likelihood estimator of affine causal processes**
Presenter: **Jean Marc Bardet**, University Paris Pantheon-Sorbonne, France
The consistency and asymptotic normality of the Laplacian Quasi-Maximum Likelihood Estimator (QMLE) is proved for a general class of causal time series including ARMA, AR(∞), GARCH, ARCH(∞), ARMA-GARCH, APARCH, ARMA-APARCH,..., processes. We notably exhibit the advantages (moment order and robustness) of the Laplacian compared to the classical Gaussian QMLE. Numerical simulations confirms the accuracy of the Laplacian estimator.

EC0202: **State-space models on Stiefel manifold: Specification and estimation**
Presenter: **Yukai Yang**, Uppsala University, Sweden
Co-authors: Luc Bauwens
Novel state-space models are proposed. The latent states evolve on the Stiefel manifold and follow the matrix Langevin distribution. It offers the possibility for modelling high-dimensional time-varying parameter change with reduced rank property. It corresponds to the dynamic factor models, and also opens up the possibility to model the time-varying long-run relationship in cointegration models. The corresponding filtering and smoothing estimators and the algorithms have been developed. The maximum likelihood estimation methods have been investigated. Two illustrations with forecasting are provided.

EG053 Room LSK1010 CONTRIBUTIONS IN BOOTSTRAP METHODS Chair: Gil Gonzalez-Rodriguez

EC0827: **The adequate bootstrap: A new method for measuring model uncertainty**
Presenter: **Toby Kenney**, Dalhousie University, Canada
Co-authors: Hong Gu
In model adequacy testing, there is a fundamental disconnect between what is tested and what we would like to test. The usual approach is to test the null hypothesis "Model M is the true model". However, Model M is never the true model. This means that if we collect enough data, we will certainly reject the model adequacy test. The model might still be useful if it is close to the truth. We propose a new method for assessing this. The idea is to determine the sample size at which the model adequacy test would not be rejected, then perform inference based on this sample size. If the model is good, we will perform inference based on a large sample size, while if the model is not a good fit, we will perform inference based on a small sample size to incorporate the model uncertainty. The intuition is that if we had only taken a smaller sample, then we would not have rejected the model, so the inference we make under the model should be valid.

EC0850: **Bootstrapping generalized empirical likelihood with many weak moment conditions**
Presenter: **Wenjie Wang**, Hiroshima University, Japan
The main contribution is to theoretically analyze the application of bootstrap methods to generalized empirical likelihood (GEL) estimation when the available moment conditions may be weak and the number of moment conditions goes to infinity with the sample size. We demonstrate that previous nonparametric i.i.d. and efficient bootstrap procedures cannot consistently estimate the distribution of GEL estimators under many weak moment conditions. The primary reason is that these bootstrap procedures fail to capture the overall moment strength in the sample adequately. Such a bootstrap failure implies that the widely used inference approaches such as the bootstrap standard error and the percentile type method are invalid under many weak moments. We also find that the efficient bootstrap is asymptotically less distorted than the nonparametric i.i.d. bootstrap under the current context. Finally, we discuss modified bootstrap procedures that provide a valid distributional approximation.

EC0774: **A new block bootstrap method for dependent data**
Presenter: **Hans Arnfinn Karlsen**, University of Bergen, Norway
Block bootstrapping is a method for handling stationary time series data not having a residual structure. It is done by blocking the data into consecutive blocks. For a time series with a Markovian structure it seems to be an advantage relating the blocking structure to its regenerative property. However, with continuous state space regeneration depends on an embedded not observable renewal process. The method [SSBB] presented is simply defined by the successive visit epochs to a fixed chosen small set. An advantage of this approach is that no estimation is needed. For fixed sample size the resulting blocking structure is an approximation of the underlying regenerative structure when the time series is Markovian. The same is true for higher order Markov models provided that the small set is appropriate chosen. Simulations experiments for Markov models indicate quite clearly that SSBB has advantages when doing pivotal bootstrapping compared to for instance the moving block bootstrap methods [MBB]. For non-pivotal bootstrapping the results are less clear and it is hard to beat MBB. First order correctness for SSBB is proved and different examples are analyses. In particular an AR(2) model is implemented and discussed.

EC0715: **Bootstrap confidence bands for Levy densities under high-frequency observations**
Presenter: **Daisuke Kurisu**, University of Tokyo, Japan
Co-authors: Kengo Kato
Bootstrap methods are developed to construct uniform confidence bands for nonparametric spectral estimation of Lévy densities under high-frequency observations. We assume that we observe n discrete observations at frequency $1/\Delta$ with $\Delta > 0$, and work with the high-frequency setup where $\Delta = \Delta_n \to 0$ and $n\Delta \to \infty$ as $n \to \infty$. We employ a spectral estimator of the Lévy density, and develop novel implementations of Gaussian multiplier (or wild) and empirical (or Efron's) bootstraps to construct confidence bands for the spectral estimator on a compact set that does not intersect the origin. We provide conditions under which the proposed confidence bands are asymptotically valid. Our confidence bands are shown to be asymptotically valid for a wide class of Lévy processes. We also develop a practical method for bandwidth selection, and conduct simulation studies to investigate the finite sample performance of the proposed confidence bands.

EC285 Room LSK1034 CONTRIBUTIONS IN FINANCIAL ECONOMETRICS I Chair: Roderick McCrorie

EC0757: **Mild explosivity in recent crude oil prices**
Presenter: **Roderick McCrorie**, University of St Andrews, United Kingdom
Co-authors: Isabel Figuerola-Ferretti, Ioannis Paraskevopoulos
A new, mildly explosive/multiple bubbles technology is used to assess whether crude oil prices exhibited departures from martingale trend behavior over the last decade and to explore whether any such departures can be explained by fundamentals or other proxy variables. The test dates two significant time periods in both Brent and WTI, nominal and real front-month futures prices: a mildly explosive episode during the 2007-08 spike, prior to the peak of the Global Financial Crisis; and a negative such episode during the recent price decline, whose commencement is dated around a key OPEC meeting in November 2014. Evidence using other commodity prices points to explanatory factors beyond commodity markets. A demand-side fundamental is decisive in explaining the episode in mid-2008, in a way that above-ground inventories and excess speculation are not. U.S. fracking was found not to be decisive in the rejection of random walk behavior in oil prices in late-2014. In spite of some recent work tying the CBOE Volatility Index (VIX) to oil futures prices, we find no evidence that the VIX affected price levels during the sample period. The results, shown to be robust to a changing dollar exchange rate and to non-stationary volatility in the shocks driving the oil prices, are compared and contrasted with those obtained by Baumeister and Kilian (2016) using a forecasting approach based on a structural vector autoregressive model.

EC0734: **Misspecification tests in conditional covariances for large cross-sectional dimensions**
Presenter: **Bilel Sanhaji**, Paris VIII, France
Co-authors: Thomas Chuffart
Lagrange multiplier tests for nonlinearity in conditional covariances in multivariate GARCH models are proposed. The null hypothesis is the full BEKK model with variance targeting in which covolatilities of time series are driven by a linear function of their own lags and lagged squared innovations. The alternative hypothesis is an extension of the model in which covolatilities are modeled by a nonlinear function of the lagged squared innovations, represented by an exponential or a logistic transition function. Partial tests are also introduced in order to determine whether the relationship of time series or group of time series is linear or nonlinear. We investigate the size and power of these tests through Monte Carlo experiments, and we provide empirical illustrations in many of which cases these tests encourage the use of nonlinearity in conditional covariances.

EC0754: **The fractionally cointegrated VAR model with threshold adjustment**
Presenter: **Chi Wan Cheang**, University of Southampton, United Kingdom
The fractional cointegrated vector autoregressive model is extended by allowing two regimes in the speed of adjustment parameter in the error correction term, treating the long run cointegrating vector and the fractional and cofractional orders constant across regimes. Since the threshold parameter is not identified under the null hypothesis of no threshold, a SupLM test for the presence of a threshold is also proposed for fractional

cointegration model. The simulations for the null asymptotic distribution and asymptotic p-value are discussed. The finite sample performance of the threshold test is examined through a Monte Carlo simulation.The purposed threshold model is applied on the volatility index (VIX) and its related futures. By the fact that investors treat the VIX related products as a security on tailed risk, the products are more attractive in uncertain than in quite time. We argue that the adjustment between volatility index and its futures towards the equilibrium could be regime switching. Empirical tests show that there exists a regime threshold in the adjustment dynamic. The adjustment by VIX futures is negligibly small and insignificant when the error correction term $(s_t - \beta f_t)$ is less than or equal to the threshold, but it is highly mean reverting when the error correction is above the threshold. This result may explain the trading behaviour of volatility traders.

EC0791: **Semiparametric estimations and identification of models with interval-valued time series**

Presenter: **Yuying Sun**, Academy of Mathematics and System Science, Chinese Academy of Sciences,, China

Co-authors: Shouyang Wang, Yongmiao Hong

Modelling and forecasting interval valued time series (ITS) has received considerable attention in statistics and econometrics. This happens because more useful information, e.g., range and level, is contained in interval-valued observations than point-valued observations. The core of available methods on ITS analysis is based on various extensions of conventional linear modelling analysis. However, few works have considered possible nonlinearities in ITS data. We firstly propose a semiparametric regression to capture the nonlinear dynamics within an ITS system, and develop the minimum distance nonparametric estimation method subject to monotonicity constraint. Asymptotic properties of our proposed estimators and forecasts are established. Monte Carlo simulation is conducted to show the finite sample performance. Empirical application to SP 500 price index document yields a better forecast performance than some popular models available in the literature.

Saturday 17.06.2017 10:15 - 11:30 Parallel Session L – EcoSta2017

EO296 Room LSK1011 STATISTICAL MODELS WITH APPLICATIONS Chair: Seng Huat Ong

EO0214: **Constructing distributions on a circle through distributions on integers**
Presenter: **Tomoaki Imoto**, University of Shizuoka, Japan
Co-authors: Kunio Shimizu
A method is provided for constructing circular distributions through discrete distributions on nonnegative integers. Many well-known circular distributions can be constructed by the method. From this construction, we propose the modified von Mises distribution through Bernoulli and Poisson distributions. The symmetric case can be constructed through a Charlier series distribution and its p-th cosine moment is expressed in terms of the probabilities that the difference of the two independent and identically distributed Charlier series random variables takes values 0 and p. We consider some properties about the proposed distribution such as trigonometric moments, graphs of density, condition for unimodal and briefly describes random number generations. Fitting example is also given for showing its flexibility.

EO0301: **Complex multiplication models for circular regression**
Presenter: **Xiaoping Zhan**, Southwestern University of Finance and Economics, China
Circular data are studied as unit complex numbers and the dependencies between them are modeled by multiplying them in the complex theoretical frame. A general complex multiplication model is proposed for circular regression. The proposed structure is very flexible in the sense that many different models can be included. The minimum norm estimation is defined and the properties of the estimator are discussed. Two commonly used models are considered and two examples of real data analysis are provided.

EO0564: **A family of distributions for modelling dispersion in count data**
Presenter: **Seng Huat Ong**, University of Malaya, Malaysia
Co-authors: Shin Zhu Sim
A family of distributions is considered which has the flexibility to cater for under- , equi- and over- dispersion in count data. The motivation to introduce this family is because it has a simple finite mixture structure and which facilitates its application and analysis. Some basic and probabilistic properties have been derived. Tests of hypothesis for equi-dispersion and simulation study of their power, parameter estimation by maximum likelihood and a probability generating function based methods have been considered. The utility of the distributions are illustrated by application to real data sets which exhibit under, equi- and over dispersion. It is shown that the distribution fits better than the well-known generalized Poisson and COM-Poisson distributions.

EO222 Room LSKG007 ADVANCES IN COMPLEX TIME SERIES ANALYSIS AND ITS APPLICATIONS Chair: Guannan Wang

EO0232: **Spline estimation and variable selection for single-index prediction models with diverging number of index parameters**
Presenter: **Guannan Wang**, College of William & Mary, United States
Co-authors: Lily Wang
Single-index models are useful and fundamental tools for handling "curse of dimensionality" problems in nonparametric regression. Along with that, variable selection also plays an important role in such model building process when the index vectors are high-dimensional. Several procedures have been developed for estimation and variable selection for single-index models when the number of index parameters is fixed. In many high-dimensional model selection problems, the number of parameters is increasing along with the sample size. We consider weakly dependent data and propose a class of variable selection procedures for single-index prediction models, which are robust against model misspecifications. We apply polynomial spline basis function expansion and smoothly clipped absolute deviation penalty to perform estimation and variable selection in the framework of a diverging number of index parameters. Under stationary and strong mixing conditions, the proposed variable selection method is shown to have the oracle property when the number of index parameters tends to infinity as the sample size increases. A fast and efficient iterative algorithm is developed to estimate parameters and select significant variables simultaneously. The finite sample behavior of the proposed method is evaluated with simulation studies and illustrated by the river flow data of Iceland.

EO0661: **Variance change point detection under a smoothly-changing mean trend**
Presenter: **Pang Du**, Virginia Tech, United States
Co-authors: Zhenguo Gao, Zuofeng Shang
The extensive literature on change point analysis can be roughly divided into the parametric and nonparametric categories. Both categories of models require a sudden change in the data distribution, either in a few parameters or in the distribution as a whole. We are concerned with the scenario that during the observation period the variance of data may make a significant jump while the mean of data changes in a smooth fashion. We propose a penalized weighted least squares approach that naturally integrates the variance change point detection and the smooth mean function estimation. An iterative procedure is introduced to obtain all the parameter estimates. Given the variance components, the mean function is estimated by smoothing splines as the minimizer of the penalized weighted least squares. Given the mean function, we propose a likelihood ratio test statistic for identifying the variance change point. The null distribution of the test statistic is derived. Simulations show excellent performance of the proposed method. Application to an liver procurement experiment study offers numerical support to the non-invasive way of assessing organ viability through surface temperature monitoring.

EO0566: **Discussion on advanced topics in complex time series analysis**
Presenter: **Jing Wang**, University of Illinois at Chicago, United States
Time series analysis is an active area of research in many academic disciplines such as economics, finance, probability and statistics among others. One problem facing statisticians today is the increasingly complex structures of time series data collected by financial agencies, federal agencies, health and other organizations. Such series often exhibit more unique features such as nonlinearity, nonstationarity and other complex functional structures that warrant development of new statistical techniques on a range of practical issues. We will briefly summarize each talk's main message and its contribution to time series analysis. We will also describe how the work presented in this session fits into the complex time series literature.

EO048 Room LSKG001 CHALLENGES IN FUNCTIONAL DATA ANALYSIS Chair: Jeng-Min Chiou

EO0238: **Testing the equality of the covariance functions of several functional populations**
Presenter: **Jin-Ting Zhang**, National University of Singapore, Singapore
With modern recording equipments, functional data are collected frequently in many scientific fields. To analyze such functional data, more and more techniques have been proposed and studied. In this talk, we discuss an L2-norm based global test for comparing the covariance functions of several functional populations. We show that the proposed statistic has an asymptotic distribution of chi-square-type mixtures and enjoys good asymptotic powers. Two methods are proposed to approximate the underlying null distributions. The methodologies are illustrated via a real data application.

EO0250: **Semiparametric incomplete functional response models with scalar predictors**
Presenter: **Catherine Liu**, The Hong Kong polytechnic University, Hong Kong
Co-authors: Jin Yang, Tao Zhang
In functional data analysis, functional data are often assumed to be fully observed on the domain. However, in dealing with real data (for example, environmental pollution data), we often face the scenario that some functional data are fully observed and others are incompletely observed. The purpose is to propose a model where the predictor is a vector or scalar and the response is a random trajectory. This model incorporates the influence of the predictor both through the mean response function and by means of covariance function for this type functional data. Some asymptotic results of the proposed methods are established. Numerical studies demonstrate the established theories.

EO0424: **Partially linear functional additive models for multivariate functional data**
Presenter: **Yehua Li**, Iowa State University, United States
A class of partially linear functional additive models is investigated that predicts a scalar response by both parametric effects of a multivariate predictor and nonparametric effects of a multivariate functional predictor. We jointly model multiple functional predictors that are cross-correlated using multivariate functional principal component analysis, and model the nonparametric effects of the principal component scores as additive components in the model. To address the high dimensional nature of functional data, we let the number of principal components diverge to infinity with the sample size, and adopt the component selection and smoothing operator penalty to select relevant components and regularize the fitting. A fundamental difference between our framework and the existing high dimensional additive models is that the principal component scores are estimated with error, and the magnitude of measurement error increases with the order of principal component. We establish the asymptotic convergence rate for our estimator, while allowing the number of components diverge. When the number of additive components is fixed, we also establish the asymptotic distribution for the partially linear coefficients. The practical performance of the proposed methods is illustrated via simulation studies and a crop yield prediction application.

EO307 Room LSK1010 RECENT ADVANCES IN CAUSAL INFERENCE METHODS Chair: Qi Long

EO0295: **Matching using sufficient dimension reduction for causal inference**
Presenter: **Yeying Zhu**, University of Waterloo, Canada
Co-authors: Wei Luo
To estimate casual treatment effects, a new matching approach is proposed based on the reduced covariates obtained from sufficient dimension reduction. Compared to the original covariates and the propensity score, which are commonly used for matching in the literature, the reduced covariates are estimable nonparametrically under a mild assumption on the original covariates, and are sufficient and effective in imputing the missing potential outcomes. Under the ignorability assumption, the consistency of the proposed approach requires a weaker common support condition. In addition, the researchers are allowed to use different reduced covariates to find matched subjects for different treatment groups. We develop relative asymptotic results, and conduct simulation studies as well as real data analysis to illustrate the usefulness of the proposed approach.

EO0302: **A fully Bayesian approach to structural nested failure time models**
Presenter: **Jason Roy**, University of Pennsylvania, United States
Structural nested failure time models (SNFTMs) are a promising alternative to marginal structural models for survival data with time-dependent confounding. These models do not require the positivity assumption and do not suffer from the null paradox. The causal parameters of SNFTMs involve the effect of a 'blip' of treatment (additional treatment in one short period of time) as a function of treatment and covariate history up to that time. These models can be fitted using g-estimation. However, for multi-parameter causal models, this type of estimation can be problematic due to the discontinuities created by artificial censoring. We propose a fully Bayesian approach as an alternative to g-estimation. We first develop a fully parametric Bayesian model and quantify its performance in simulation studies. We then develop Bayesian nonparametric approaches to weaken some of the modeling assumptions.

EO0317: **Semiparametric models and inference for the treatment effect on nonnegative outcome clumped at zero**
Presenter: **Jing Cheng**, University of California, San Francisco, United States
Analyses are considered for nonnegative outcomes with a clump of observations at zero and continuously distributed positive values. Examples include medical care costs when a lot of subjects have zero costs, burden of illness when a lot of people do not experience the disease, earnings after job training program when many subjects have not got a job. We develop a class of semiparametric models and inference procedures that extend the widely used Tobit model. We consider more flexible models for the treatment effect than the shift effect of the Tobit model; and we make semiparametric inferences that allow the underlying latent variable to have any distribution, unlike the Tobit model that assumes normally distributed latent variable. We apply our method to data from the RAND Health Insurance Experiment and evaluate the treatment effect on medical care costs.

EO076 Room LSK1009 CIRCULAR TIME SERIES AND STATISTICAL INFERENCE Chair: Toshihiro Abe

EO0401: **Asymptotically optimal inference for two modal concentration of antipodally symmetric circular distributions**
Presenter: **Toshihiro Abe**, Nanzan University, Japan
Co-authors: Hiroaki Ogata, Takayuki Shiohama, Hiroyuki Taniai
The optimal asymptotic inference for antipodal symmetry is considered when the center direction of symmetry is specified by using a family of bimodal distributions on the circle. To adapt the Le Cam theory, we consider the reparameterization of the existing cosine perturbed bimodal models on the circle. The properties of the uniform local asymptotic normality (ULAN) for the cosine perturbed models are described. Finally, as an illustrative example, our developed tests are applied to the movements of 76 female turtles.

EO0461: **Multi-order circular Markov processes with canonical vine representations**
Presenter: **Hiroaki Ogata**, Tokyo Metropolitan University, Japan
A way of construction of multi-order circular Markov processes is introduced. Bivariate circular distributions with specified marginal distributions were previously proposed, and its representation naturally induces a circular Markov process with two arbitrary circular densities. One of the densities is called a binding density and it can be regarded as a copula density. This circular Markov process is extended to a multi-order one by employing a pair-copula decomposition of a multivariate distribution. The way of decomposition corresponds to the graphical model denoted as the canonical vine. Fitting the multi-order circular Markov process to real circular data is also considered.

EO0471: **Bayesian estimation for the inverse Batschelet distributions on the circle**
Presenter: **Takayuki Shiohama**, Tokyo University of Science, Japan
Co-authors: Toshihiro Abe, Yoichi Miyata
The inverse Batschelet distributions on the circle provide an asymmetrical shape of distribution. The distributions have flat-topped and sharply peaked properties by adding a shape parameter in the inverse monotone functions on the base symmetric densities. Bayesian estimation and finite sample inference techniques are considered. Different types of MonteCarlo methods are investigated and compared to access the performance of the proposed estimation scheme. An application on data analysis is also presented for the illustration purpose.

EO260 Room LSK1014 NEW DEVELOPMENTS IN BIOMEDICAL RESEARCH II Chair: Jinfeng Xu

EO0624: **Estimation of time-varying intervention effects using recurrent event data with application**
Presenter: **Feng Chen**, UNSW Syd, Australia
Co-authors: Xu Han, Eddy Lam, Yin Bun Cheung, Paul Milligan
The purpose is to consider the estimation of the effect of interventions such as malaria chemoprevention and vaccine booster doses that are applied intermittently in infectious disease control. A nonparametric function to model the time-varying intervention effect in the framework of Andersen-Gill model for recurrent event time data, is considered. The potential exciting or correcting effects due to past episodes of the disease is incorporated in the model. Another nonparametric function is used to model the effect of past disease episodes. The partial likelihood estimation approach is adopted and a simulation study is carried out to evaluate the performance of the proposed method. A method of assessing the overall goodness-of-fit of the counting process model is also proposed. The methodology is illustrated with the analysis of data from a malaria chemoprevention trial.

EO0529: **Optimal high-dimensional multiclass linear discriminant analysis**
Presenter: **Shan Luo**, Shanghai Jiao Tong University, China
The Bayes rule is reconsidered for multiclass discriminant analysis under high-dimensional situation where number of variables p depends on the total sample size n. Particularly, we focus on the scenario in which all K classes are assumed to come from multivariate normal distributions. When $p < n$, traditional linear discriminant analysis (LDA) estimate the population mean vectors by within class mean vector and the common covariance matrix is estimated by the pooled sample covariance matrix if necessary. We will provide theoretical results on its misclassification rate for $K = 3$. When $p > n$, we propose a novel estimation method and thus a new discriminant rule can be constructed. We will investigate its misclassification rate theoretically as well and its effectiveness will be illustrated through extensive numerical results.

EO0387: **Subgroup analysis with nonparametric unimodal symmetric error distribution**
Presenter: **Ao Yuan**, Georgetown University, United States
In clinical trials, one important objective is to classify the patients into treatment-favorable and non-favorable subgroups. Existing parametric methods are non-robust, and the commonly used classification rules do not consider the priority of the treatment-favorable subgroup. To address these issues, we propose a semiparametric model, with the sub-densities specified nonparametric. For nonparametric mixture identifiability, the sub-density is assumed symmetric, and unimodal to find its nonparametric maximum likelihood estimate. The semiparametric likelihood ratio statistics is used to test the existence of subgroups, while the Neyman-Pearson rule to classify each subject. Asymptotic properties are derived, simulation studies conducted to evaluate the performance of the method, and then it is used to analyze a real data.

EO182 Room LSK1003 COMPUTATIONAL METHODS IN FINANCIAL STATISTICS Chair: Chu-Lan Kao

EO0224: **Robust test of stock return predictability under heavy-tailed innovations**
Presenter: **Hsinchieh Wong**, National Central University, Taiwan, Taiwan
Co-authors: Cheng-Der Fuh, Menghua Chung
Conventional tests of the predictability of stock returns are usually based on the normal innovation assumption, which does not fit the empirical data well. To remedy this ideal assumption, we consider a predictive regression model $Y_t = \beta_0 + \beta_1 X_{t-1} + u_t$, $X_t = \rho X_{t-1} + e_t$ with heavy tails. That is, $\{u_t, t \geq 1\}$ and $\{e_t, t \geq 1\}$ are two sequences of random variables, in which the distributions are in the domain of attraction of the normal law with zero means and possibly infinite variances. To construct a robust statistic based on this model, we study asymptotic behavior of the estimators of β_0, β_1 and $(\beta_0, beta_1)^T$ for stationary as well as local to unity cases. Our results show that when $|\rho| < 1$ or ρ tends to unity but slowly enough, the proposed robust statistic is indeed pivotal, and can be used directly to test the predictability of stock returns. Next, under the dependent structure of u_t and e_t, in the case of local to unity, we propose a modified test based on the celebrated Bonferroni Q-test and present the efficiency of this test. Finally, based on our theoretical results, we study the relationships among heavy-tails, unit root and predictability. Numerical simulations and empirical studies are given for illustration.

EO0308: **Optimal search for parameters in Monte Carlo simulation for derivative pricing**
Presenter: **Chuan-Ju Wang**, Academia Sinica, Taiwan
A novel and general framework is provided for the problem of searching parameter space in Monte Carlo simulations. We propose a deterministic online algorithm and a randomized online algorithm to search for suitable parameter values for derivative pricing which are needed to achieve desired precisions. We also give the competitive ratios of the two algorithms and prove the optimality of the algorithms. Experimental results on the performance of the algorithms are presented and analyzed as well.

EO0382: **Correlated defaults under multi-factor models**
Presenter: **Chu-Lan Kao**, National Chiao-Tung University, Taiwan
Co-authors: Cheng-Der Fuh
Previous studies have shown that defaults correlated with each other through contagion effects and others. Under the widely used factor model, we directly capture this through the classical default barrier technique, and further quantify how a firm's default impacts others through asymptotic theory, both for a specific firm, or the first firm to default among a group of firms. Its relationship with the Merton's distance-to-default and mixture indicator models is discussed, and the application in pricing first-to-default swap is derived. Simulation studies are further provided.

EO030 Room LSK1007 RECENT ADVANCES IN DYNAMIC PANEL DATA AND FACTOR MODELS Chair: Bin Peng

EO0475: **Semiparametric single-index panel data models with interactive fixed effects: Theory and practice**
Presenter: **Bin Peng**, University of Bath, United Kingdom

A single index panel data model is proposed with unobserved multiple interactive fixed effects. This model has the advantages of being flexible and of being able to allow for common shocks and their heterogeneous impacts on cross sections, thus making it suitable for the investigation of many economic issues. We derive asymptotic theories for both the case where the link function is integrable and the case where the link function is nonintegrable. Our Monte Carlo simulations show that our methodology works well for large N and T cases. In our empirical application, we illustrate our model by analyzing the returns to scale of large commercial banks in the U.S. Our empirical results suggest that the vast majority of U.S. large banks exhibit increasing returns to scale.

EO0516: **Estimation and inference of panel data models with constrained interactive fixed effects**
Presenter: **Hongjun Li**, Capital University of Economics and Business, China

Panel data models with interactive fixed effects (PD-IFE) have important applications in empirical economic analyses as they can describe elaborate economic structures. We incorporate some prior information into PD-IFE to construct panel data models with constrained interactive fixed effects (PD-CIFE) and develop the estimation and inference methods for these models. Following a previous approach, we solve the first order conditions of the least squares problem simultaneously to estimate the unknown coefficients of exploratory variables, factors, and factor loadings. We derive the asymptotic properties of these estimators and provide testing methods of the constraints. Monte-Carlo simulations results are in favor of our methods, showing that our estimators are more efficient given that the constraints are correct.

EC0294: **Sufficient and linear dimension reduction methods in forecasting**
Presenter: **Efstathia Bura**, Vienna University of Technology, Austria
Co-authors: Alessandro Barbarino

Factor models are widely used in summarizing large datasets with few underlying latent factors and in building time series forecasting models for economic variables. In these models, the reduction of the predictors and the modeling and forecasting of the response y are carried out in two separate and independent phases. We introduce a potentially more attractive alternative,Sufficient Dimension Reduction (SDR), that summarizes x as it relates to y so that all the information in the conditional distribution of $y|x$ is preserved. We study the relationship between SDR and popular estimation methods, such as ordinary least squares (OLS), dynamic factor models (DFM), partial least squares (PLS) and RIDGE regression, and establish the connection and fundamental differences between the DFM and SDR frameworks. We show that SDR significantly reduces the dimension of widely used macroeconomic series data with one or two sufficient reductions delivering similar forecasting performance to that of competing methods in macro-forecasting.

EG239 Room LSK1033 CONTRIBUTIONS IN FINANCIAL ECONOMETRICS II Chair: Bernd Schwaab

EC0336: **Tail risk in government bond markets and ECB unconventional policies**
Presenter: **Bernd Schwaab**, European Central Bank, Germany
Co-authors: Xin Zhang

A novel observation-driven model is derived to study the time variation in the tail shape for time series observations from a wide class of fat-tailed distributions. Monte Carlo experiments suggest that the model reliably captures tail shape variation in a variety of simulation settings. In an empirical study of sovereign bond yields at a high frequency, we demonstrate that unconventional monetary policies adopted by the European Central Bank between 2010 and 2012, specifically its Securities Markets Programme and Outright Monetary Transactions, lowered the tail risk associated with holding certain sovereign bonds during the euro area sovereign debt crisis.

EC0682: **Investors' favourite: A different look at valuing individual labour income**
Presenter: **Jan Voelzke**, University of Muenster, Germany
Co-authors: Jeanne Diesteldorf, Fabian Goessling, Till Weigt

Human capital is a key economic factor in both macro- and microeconomics,and, at least for most people, by far their largest asset. Surprisingly, relatively little effort has been undertaken in the extant literature to empirically determine the value of individual human capital. The aim is to close this gap. We use the Substantial-Gain-Loss-Ratio to calculate GoodDeal bounds for securitizations of individual labour income one year ahead. Our procedure is applied to US data. We evaluate the attractiveness of hypothetical human capital contracts and can thereby identify investors' favourites.

EC0818: **Multivariate dominance among financial sectors**
Presenter: **Tomas Tichy**, VSB-TU Ostrava, Czech Republic
Co-authors: Noureddine Kouaissah, Sergio Ortobelli

A multivariate stochastic dominance comparison among different sectors is proposed from the point of view of different non-satiable investors. In particular, we consider different distributional hypotheses for the multivariate distribution of financial sectors and we examine if there exist some dominance among them. In this framework we also discuss the asymptotic dominance between financial sectors. Moreover, we determine the stochastic dominance rule for stable distributions, where the stability parameter plays an important role. Consequently, the multivariate rule for ordering sectors is based on a comparison between i) stability indices, ii) location parameters, iii) dispersion parameters and iv) skewness parameters. Finally, we empirically examine the choices of some non-satiable investors taking into account the proposed studies.

EG267 Room LSK1034 CONTRIBUTIONS IN OPTIMAL PORTFOLIO ALLOCATION AND OPTION PRICING Chair: Winfried Pohlmeier

EC0726: **Testing out-of-sample portfolio performance**
Presenter: **Winfried Pohlmeier**, University of Konstanz, Germany
Co-authors: Ekaterina Kazak

The quality of portfolio performance tests is studied based on out-of-sample returns. By disentangling the components of out-of-sample performance we show that observed differences are to a large extend driven by the differences in estimation risk. Our Monte-Carlo study reveals that the puzzling empirical results of inferior performance of the theoretically superior strategies based on the out-of-sample comparison mainly result from the low power properties of these tests. Thus, our results provide an explanation why the null hypothesis that the simple equally weighted portfolio cannot be outperformed by theoretically superior portfolio strategies in many out-of-sample horse races regardless of the underlying testing strategy. For the applied researcher we provide some guidance to cope with the problem of low power. In particular, we show how despite their low power out-of-sample performance tests can be optimally used for a pretest-based portfolio strategy.

EC0743: **A sequential convex approximation method for optimal portfolio allocation**
Presenter: **Xijun Liang**, China University of Petroleum, China

CVaR has received extensive attentions as a convex risk metric in recent years. However, the convex approximation in CVaR has large deviation from the true indicate function. The aim is to overcome this drawback and to propose a DC (difference of convex functions) approximation. The DC approximation makes the approximation error arbitrarily small. A DC programming is then designed for optimal portfolio allocation, where the DC approximation is employed and all the constraints are satisfied in the probabilistic sense. Moreover, a sequential convex approximation (SCA) algorithm is designed for solving the DC programming. The SCA algorithm approaches the optimal solution by iteratively solving a sequence of convex subproblems with Monte-Carlo method. The SCA algorithm has simple iteration scheme. Preliminary experimental results have shown that the performance of the method are better than the CVaR approximation, especially when the income factors have fat tail distributions.

EC0753: **Pricing Parisian options with an adaptive Monte Carlo method**
Presenter: **Sercan Gur**, Vienna University of Economics and Business, Austria

Parisian option is a type of barrier option, which can only be exercised if the underlying value process not only reaches a barrier level but remains a certain prescribed time (so-called window period) below (or above) this level. Closed form solutions for the value of these contracts do not exist. In order to price Parisian options, we use Monte Carlo simulation instead of partial differential equations, inverse Laplace transform or lattices. We propose a new Monte Carlo method which can be used to price Parisian options not only with constant boundary but with more general boundary. The advantage of this approach is that it can easily be adapted to compute the price of an option with more complicated path-dependent pay-off. We use an adaptive control variable to improve the efficiency of the Monte Carlo estimator. At last, we provide a numerical example to illustrate our method and a comparison of previous Monte Carlo methods with our technique.

EC287 Room LSK1005 CONTRIBUTIONS IN BAYESIAN STATISTICS Chair: Siew Li Linda Tan

EC0826: **Prior distributions for ranking problems**
Presenter: **Hong Gu**, Dalhousie University, Canada
Co-authors: Toby Kenney

The ranking problem is to order a collection of units by some unobserved parameter, based on observations from the associated distribution. This problem arises naturally in a number of contexts, such as business, where we may want to rank potential projects by profitability; or science, where we may want to rank variables potentially associated with some trait by the strength of the association. Many approaches to this problem are empirical Bayesian, where we use the data to estimate the hyper-parameters of the prior distribution, and then we use that distribution to estimate the unobserved parameter values. There are a number of different approaches to this problem, based on different loss functions for mis-ranking units. However, little has been done on the choice of prior distribution. Typical approaches involve choosing a conjugate prior for convenience, and estimating the hyper-parameters by MLE from the whole data set. We look in more detail at the effect of choice of prior distribution on Bayesian ranking. We focus on the use of posterior mean for ranking, but many of our conclusions should apply to other ranking criteria, and it is not too difficult to adapt our methods to other choices of prior distributions.

EC0220: **Bayesian inference for multiple Gaussian graphical models with application to metabolic association networks**
Presenter: **Siew Li Linda Tan**, National University of Singapore, Singapore
Co-authors: Ajay Jasra, Maria De Iorio, Timothy Ebbels

The aim is to investigate the effect of cadmium (a toxic environmental pollutant) on the correlation structure of a number of urinary metabolites using Gaussian graphical models (GGMs). The inferred metabolic associations can provide important information on the physiological state of a metabolic system and insights on complex metabolic relationships. Using the fitted GGMs, we construct differential networks, which highlight significant changes in metabolite interactions under different experimental conditions. The analysis of such metabolic association networks can reveal differences in the underlying biological reactions caused by cadmium exposure. We consider Bayesian inference and propose using the multiplicative (or Chung-Lu random graph) model as a prior on the graphical space. In the multiplicative model, each edge is chosen independently with probability equal to the product of the connectivities of the end nodes. This class of prior is parsimonious yet highly flexible; it can be used to encourage sparsity or graphs with a pre-specified degree distribution when such prior knowledge is available. We extend the multiplicative model to multiple GGMs linking the probability of edge inclusion through logistic regression and demonstrate how this leads to joint inference for multiple GGMs. A sequential Monte Carlo (SMC) algorithm is developed for estimating the posterior distribution of the graphs.

EC290 Room LSK1001 CONTRIBUTIONS IN APPLIED ECONOMETRICS AND STATISTICS Chair: Richard Gerlach

EC0690: **Modeling multi-state health transitions in China: A generalized linear model with time trends**
Presenter: **Han Li**, University of New South Wales, Australia
Co-authors: Katja Hanewald, Adam Shao

Population ageing has reached a new dimension in China. With rapid economic growth and medical advances, the longevity and health in China has been continuously improving during recent decades. Therefore, there is an increasing need to understand and analyse the ill-health transitions among Chinese elderly. We propose a generalized linear model to estimate the transition rates of functional disability. We use individual-level panel data from the Chinese Longitudinal Healthy Longevity Survey (CLHLS) for the period from 1989 to 2012. We made a formal comparison between male and female resident, for both rural and urban areas. Based on the results from proposed model, we also predicted the demand for aged care services in China.

EC0707: **Impact assessment of media investments on automobile brand market-shares using compositional models**
Presenter: **Joanna Morais**, Toulouse School of Economics CNRS, France
Co-authors: Christine Thomas-Agnan, Michel Simioni

Because media investments in the automobile market represent a very large amount of money, the interest is in modeling their impact on automobile brand market-shares. In the marketing literature, market-share models are quite simple, but require a large number of coefficient in order to take into account brand specific effects and cross-effects for example. The aim is to show that other mathematical models called compositional models, usually used in geology, can be adapted for market-share regression. These models are based on the geometry of the simplex, and thus respect the constraints of share data. They give good fitting results thanks to their flexibility, despite their sometimes abstract presentation. We prove that compositional models can be written in an attraction form, and thus can be compared to other market-share models. The interpretation of compositional models is not straightforward due to the fact that a change in a share implies a change on other shares because of the summing up to one constraint. We develop several elasticity measures in order to interpret the impact of media investments of brand l on market-share of brand j, and we show how these impact measures can be useful for decision making for an automobile firm.

EC0422: **A study of statistical and machine learning methods for cancer classification using cross-species genomic data**

Presenter: **Lan Gao**, The University of Tennessee at Chattanooga, United States

Use of gene expression profiling of animal model of a certain disease gives pre-clinical insights for the potential efficacy of novel treatments and drugs. Selection of an animal model, accurately resembling the human disease, profoundly reduces the research cost in resources and time. We introduce and compare three different methods for classification of sub-types of cancer via cross-species genomic data. A statistical procedure based on analysis of variance (ANOVA) of similarity of gene expression between human and animal is used to select the animal model that most accurately mimics the human disease. Two other commonly used methods, logistic regression and artificial neural networks are also examined and analyzed for the same data sets. Implementing procedure of each of these algorithms is discussed. Computational cost, advantage and drawback of each algorithm is scrutinized for classification of simulated data and a real example of medulloblastoma (a type of brain cancer).

Saturday 17.06.2017 14:00 - 15:40 Parallel Session N – EcoSta2017

EI004 Room LSKG001 BAYESIAN NONPARAMETRICS Chair: Byeong Park

EI0699: **On the computational complexity of MCMC methods for some high-dimensional models**
Presenter: **Lizhen Lin**, The University of Notre Dame, United States
Co-authors: Minwoo Chae, David Dunson
Markov chain Monte Carlo (MCMC) methods for high-dimensional models are often computationally inefficient due to the slow speed of convergence to the stationary distribution. We consider MCMC methods of high-dimensional models when parameters are divided into two groups: parameters of interest and nuisance parameters. Although the global convergence rate of the Markov chain is often slow, the rate for the parameter of interest can be much faster than that of the global rate under certain minorization conditions. As an example, we consider a semiparametric regression model with a Dirichlet process mixture prior for the error distribution in which a natural conjugate algorithm may slow the chain down. We propose an alternative algorithm that can resolve this problem and provide theoretical justifications.

EI0736: **Dynamic topic model using the power prior approach**
Presenter: **Yongdai Kim**, Seoul National University, Korea, South
The topic model provides a flexible mixed-membership so called topics to documents. We consider dynamic topic models, in which the generative model changes in time, and develop a novel algorithm to update the posterior distribution dynamically by combining the conjugate approximation and the power prior approach. An important advantage of the proposed algorithm is that it updates the posterior distribution by reusing a given batch algorithm without specifying a complicated dynamic generative model. Thus, the proposed algorithm is conceptually and computationally simpler. By analyzing real datasets, we show that the proposed algorithm is a useful alternative to other dynamic topic models.

EO232 Room LSK1026 MODEL ESTIMATION IN MATHEMATICAL FINANCE Chair: Xiaoling Dou

EO0158: **SVM-Jacobi for distribution fitting with applications to quantitative finance and actuarial science**
Presenter: **Xixuan Han**, The University of Hong Kong, Hong Kong
A method called SVM-Jacobi is proposed to approximate probability distributions by linear combinations of exponential distributions, associated with a comprehensive asymptotic analysis. In multivariate cases, the method also effectively works to provide approximations by linear combinations of products of independent exponential distributions. The proposed method is particularly applicable and useful in quantitative finance and actuarial science. Many pricing and hedging formulas have closed forms under exponential distributions. By approximating the real distributions, we are capable to use the closed-form formulas and fitted coefficients of SVM-Jacobi to approximate the prices and Greeks. In addition to the methodology, we give examples of approximating the credit value adjustment of defaultable bonds, financial derivatives with single payments and credit default swaps, and the value of equity-linked death benefits. Some numerical results also are presented for illustration.

EO0323: **Robust interpolation problems in L^p**
Presenter: **Yujie Xue**, Waseda University, Japan
Co-authors: Yan Liu, Masanobu Taniguchi
For a weakly stationary process $\{X_t, t \in Z\}$ with mean 0 and spectral distribution function $F(\lambda)$, the linear extrapolation problem can be transferred into a problem to get the distance from 1 to a completion of the linear hull of the set $\{e^{ij\omega}; j \in S\}$ where S is a subset of integers. When $S = \{..., -2, -1, 1, 2, ...\}$, in the L^2 space, the optimal interpolator is well known. The case of L^p where $p \geq 1$ is mainly discussed. It is shown that the closed form of the optimal interpolator can be given when $p > 1$, and that when $p = 1$, only value of minimum interpolation error can be obtained. Furthermore, the minimax problem when F is partially known is discussed.

EO0437: **Semi-parametric estimates of long-term background trend, periodicity, and clustering effect for a Hawkes point process**
Presenter: **Jiancang Zhuang**, Institute of Statistical Mathematics, Japan
Co-authors: Jorge Mateu
Past studies have shown that crime behaviors are clustered. A spatiotemporal Hawkes-type point-process model is proposed, which includes a background component with daily and weekly periodization and a clustering component that is triggered by previous events, for describing the occurrences of violence or robbery related to crimes in the city of Castellon, Spain, during 2012 and 2013. A nonparametric method, called stochastic reconstruction, is used to estimate each component, including daily and weekly periodicity of background rate, spatial background rate, long-term background trend, and the spatial and temporal response function in the triggering component, of the conditional intensity of the model. The results show that about 3 percents of the crimes can be explained by clustering. Residual analysis is carried out to validate the model formulation.

EO0227: **An investigation of Ochi's estimator**
Presenter: **Xiaoling Dou**, Waseda University, Japan
Ochi's estimator is proposed to estimate the autoregressive coefficient of the first order of autoregression (AR(1)) model by using two constants for the end points of the process. Classical estimators, such as the least square estimator, Burg's estimator and Yule-Walker's estimator of the parameter in AR(1) model are special choices of the constants in Ochi's estimator. First, we provide a simulation for AR(1) model and examine the performance of Ochi's estimator. Writing the autoregressive conditional heteroskedasticity model of order 1, ARCH(1), into a similar form to AR(1), we extend Ochi's estimator to the ARCH(1) model, introduce the ideas of the least squares estimator, Burg's estimator and Yule-Walker's estimator and compare the relationships of them with Ochi's estimator for ARCH(1) model. With a simulation, we investigate Ochi's estimator for ARCH(1) with different values of parameters and different sample sizes.

EO088 Room LSK1001 QUANTILE REGRESSION IN HIGH DIMENSIONS Chair: Zhongyi Zhu

EO0217: **High-dimensional partially linear additive quantile regression**
Presenter: **Ben Sherwood**, University of Kansas, United States
Quantile regression is a semiparametric approach to estimating conditional quantiles that can provide a more complete picture of the conditional distribution than focusing only on the conditional mean. Modeling non-central quantiles can capture heteroscedastic relationships, while modeling the condition median is a robust alternative to least squares. These properties are particularly appealing when dealing with large and noisy data sets. We will present the partially linear additive quantile regression model and a penalized estimator that simultaneously performs estimation and variable selection. The method will be evaluated using simulations and modeling birth weight using gene expression data.

EO0390: **High dimensional censored quantile regression**
Presenter: **Qi Zheng**, University of Louisville, United States
Censored quantile regression (CQR) has emerged as a useful regression tool for survival analysis. Stochastic integral based estimating equations are commonly used in the estimation of CQR, and pose new challenges in the analysis of CQR for high dimensional survival data. We study the high dimensional CQR simultaneously over a continuum of quantile indices. We propose a two-step penalization procedure, which accommodates stochastic integral based estimating equations and properly addresses the associated complications. We establish the uniform convergence rates for the proposed estimators, and investigate the properties on weak convergence and variable selection. We conduct extensive numerical studies to confirm our theoretical findings and illustrate the practical utility of our proposals.

EO0588: **Change point tests for stochastic M-estimation**
Presenter: **Zhou Zhou**, University of Toronto, Canada
Structural change testing is considered for a wide class of time series M-estimation with non-stationary predictors and errors. Flexible predictor-error relationships, including exogenous, state-heteroscedastic and autoregressive regressions and their mixtures, are allowed. New uniform Bahadur representations are established with nearly optimal approximation rates. A CUSUM-type test statistic based on the gradient vectors of the regression is considered. A simple bootstrap method is proposed and is proved to be consistent for M-estimation structural change detection under both abrupt and smooth non-stationarity and temporal dependence. Our bootstrap procedure is shown to have certain asymptotically optimal proper-ties in terms of accuracy and power. A public health time series dataset is used to illustrate our methodology, and asymmetry of structural changes in high and low quantiles are found.

EO0397: **Non-crossing multiple-index quantile regression**
Presenter: **Liping Zhu**, Renmin University of China, China
Though the theoretical properties of quantile regression have been extensively studied in the past three decades, in practice it is not unusual to obtain crossing quantile surfaces with regular approaches to estimating quantile functions at different quantile levels. The crossing quantile surfaces are intrinsically uninterpretable. To address this issue, we consider a semiparametric multi-index quantile regression subject to monotonicity restriction at different quantile levels. We first connect the semiparametric multi-index quantile regression model with a dimension-reducible model. Such a connection allows us to estimate the index coefficients consistently. The B-splines are then used to approximate the nonparametric function under the monotonicity restriction, which numerically corresponds to a constrained linear programming problem. To further improve efficiency, we estimate the B-spline coefficients based on a dual of linear programming. We assess the finite-sample performance of our proposed method through comprehensive simulations, and compare the prediction performance of different methods through an application to a real dataset.

EO078 Room LSK1005 MODELING AND TESTING PROBLEMS WITH COMPLEX HIGH-DIMENSIONAL DATA Chair: Ming-Yen Cheng

EO0226: **High-dimensional linear models with common breaks**
Presenter: **Jialiang Li**, NUS, Duke-NUS, SERI, Singapore
A general class of high-dimensional linear models indexed by time is considered, which includes the balanced panel data dynamic linear model and the spatio-temporal linear model, where the number of predictors tends to infinity at each time point. A procedure for simultaneously detecting multiple common breaks is developed rigorously via the construction of adaptive group lasso penalty. Consistency of the common breaks estimation is established under mild conditions even when the true number of breaks diverges with the sample size or the number of time points. The simulation studies demonstrate that our method is computationally fast and numerically stable and accurate. Two empirical data examples are provided where interesting findings are obtained.

EO0364: **A novel approach to evaluating relationships in data containing mixed variable types**
Presenter: **Susan Wilson**, UNSW, Australia
There is increasing availability of big data that contain many variables of different types, such as continuous, categorical and so forth. Such data sets are motivating statisticians to find methods of data exploration and testing that make minimal assumptions while having low computational cost. Mutual information (MI) has been proposed as a useful technique to meet these goals. An approach to estimating MI based on k-nearest neighbour distances has been proposed. Unfortunately this approach was restricted to continuous variables, and also the choice of a value for k is problematic and can result in substantial bias. Recently the k-nearest neighbour method has been extended to deal with data of all types. Further a local, adaptive approach to determine k that gives a minimally biased estimate of MI has been developed. The approach can be parallelised, and has relatively low computational cost. An R package that implements these advances, migen, is under development.

EO0612: **Dynamic latent class modeling with applications**
Presenter: **Wendy Lou**, University of Toronto, Canada
Motivated by a Canadian birth cohort study, where data were collected through multiple platforms across multiple sites over time, a robust approach is developed to describe various trajectories of subject changes (e.g. lung function) with associations of interest (e.g. gene, environment, etc). A class of dynamic models involving flexible mixture distributions and latent classes will be introduced, and practical examples will be given for time-varying measurements, such as nutrition and air quality. To illustrate the proposed approach, numerical results as well as real applications will be presented.

EO0580: **Regularized classification with its application to biomedical spectroscopic data**
Presenter: **Ying Zhu**, Nanyang Technological University, Singapore
Co-authors: Augustine Tuck Lee Tan, Wai Kwong Cheang
High-dimensional spectroscopic data consist of many overlapping absorption bands sensitive to the physical and chemical states of compounds. Often, only a small subset of spectral features is found essential. Direct implementation of linear discriminant methods on high-dimensional spectroscopic data provides poor classification results due to singularity problem and highly correlated spectral features. A regularized classification model incorporating complex spectral correlation structure enabled an automatic selection of a small number of informative spectral absorption bands for the purpose of classification and interpretation. This model has been applied on biomedical spectroscopic data. The well-performed selection of informative spectral features leads to substantial reduction in model complexity, improvement of classification accuracy, and is particularly helpful for providing us insights in the interpretation of the complex spectroscopic data regarding its active ingredients.

EO170 Room LSK1032 RECENT ADVANCES IN TIME SERIES ANALYSIS Chair: Chun Yip Yau

EO0251: **Robust optimal estimation of asymptotic covariance matrices in non-stationary multi-dimensional time series**
Presenter: **Kin Wai Chan**, Harvard University, United States

Robust optimal estimation is considered for the asymptotic covariance matrix (ACM) of the sample mean in multi-dimensional time series that potentially have hidden trends and structural breaks. Robust estimation of the ACM is crucial to statistical inference because the ACM is naturally involved in many statistical procedures, e.g., change point detection, and construction of simultaneous confidence bands of trends. We propose an estimator of the ACM, which is robust to unknown forms of trends and possibly divergent number of change points. We also propose a robust estimator of the optimal bandwidth in a close form so that the estimator can be easily operated with an asymptotically correct optimal bandwidth. The estimator is robust, statistically efficient, computationally fast, and easy to implement. An empirical study on the S&P 500 Index is presented.

EO0285: **Test for the existence of finite moments via bootstrap**
Presenter: **Wai Leong Ng**, The Chinese University of Hong Kong, Hong Kong
Co-authors: Chun Yip Yau

A bootstrap hypothesis test is developed for the existence of finite moments of a random variable, which is non-parametric and applicable to both independent and dependent data. The test is based on a property in bootstrap asymptotic theory, in which the m out of n bootstrap sample mean is asymptotically normal when the variance of the observations is finite. Consistency of the test is established. Monte Carlo simulations are conducted to illustrate the finite sample performance and compare it with alternative methods available in the literature. Applications to financial data are performed for illustration.

EO0468: **Multivariate stochastic regression models with multi-step ahead predictors in macroeconomic time series**
Presenter: **Ka Wai Tsang**, The Chinese University of Hong Kong, Shenzhen, China

Macroeconomic models and forecasts typically involve a multitude of economic time series. In econometrics, dynamic factor models were introduced to incorporate time series aspects in factor models for studying dynamic covariation of economic variables. While previous empirical research focused on estimating indexes of covariation, The dynamic factors were previously used for forecasting purposes with the premise that the information in the large number of predictors can be replaced by a handful of estimated factors. A brief review of dynamic factor models is first given and their applications to multi-step ahead macroeconomic forecasts are discussed. A new approach is then described to address some long-standing difficulties in choosing the factors, modeling their dynamics, and evaluating the model-based forecasts. This approach is built upon the recent work on high-dimensional multivariate stochastic regression models, capitalizing on their coefficient and rank sparsity, and upon new model selection and forecast evaluation criteria for multivariate time series. These criteria involve pertinent features of the predicted joint distribution of its components in multi-step ahead forecasts.

EO0379: **On the Bartlett correction of empirical likelihood for time series**
Presenter: **Kun Chen**, Southwestern University of Finance and Economics, China

Bartlett's correction, which improves the coverage accuracies of confidence regions, is one of the desirable features of empirical likelihood. For empirical likelihood with dependent data, previous studies on Bartlett's correction are mainly restricted to Gaussian processes. By establishing the validity of Edgeworth expansion for the signed root empirical log-likelihood ratio statistics, we show that Bartlett's correction is applicable to empirical likelihood for short-memory time series with possibly non-Gaussian innovations. In particular, the order of the coverage errors of Bartlett corrected confidence regions can be reduced from $O(n^{-1})$ to $O(n^{-2})$.

EO102 Room LSK1009 NEW ADVANCES IN STATISTICAL MODELING, COMPUTATION AND APPLICATIONS Chair: Tsung-I Lin

EO0260: **Joint tests of financial market contagion with applications**
Presenter: **Cody Yu-Ling Hsiao**, Macau University of Science and Technology, China

A new class of multiple-channel tests of financial market contagion is introduced in which the transmission channels of financial market crises are identified jointly through the correlation, co-skewness and co-kurtosis of the distribution of returns. The proposed tests have the advantage over the existing single-channel tests of contagion in the literature in that they yield the correct size in small samples which is typical of crisis periods. Regarding the power of the tests, the multiple-channel tests display the second highest power following the single-channel tests if the data generating process for an experiment contains the transmission channel of contagion consistent with the single-channel test. The proposed tests are applied to test for financial market contagion in equity markets during the three financial crises of 2007-12. The results show that the joint tests identify various combinations of transmission channels during the three crises.

EO0293: **Multivariate-t linear mixed models for multiple longitudinal data with censorship and fat-tailed behavior**
Presenter: **Wan-Lun Wang**, Feng Chia University, Taiwan
Co-authors: Tsung-I Lin, Victor Hugo Lachos Davila

The analysis of complex longitudinal data is challenging due to several inherent features: (i) more than one series of responses are repeatedly collected on each subject at irregularly occasions over a period of time; (ii) censorship due to limits of quantification of responses arises left- and/or right- censoring effects; (iii) outliers or heavy-tailed noises are possibly embodied within multiple response variables. The aim is to formulate the multivariate-t linear mixed model with censored responses (MtLMMC), which allows the analysts to model such data in the presence of the above described features simultaneously. An efficient expectation conditional maximization either (ECME) algorithm is developed to carry out maximum likelihood estimation of model parameters. The implementation of the E-step relies on the mean and covariance matrix of truncated multivariate-t distributions. The proposed methodology is illustrated through a simulation study and a real application on HIV/AIDS data.

EO0264: **Novel bayesian information criterion for choosing the number of factors in factor analysis with incomplete data**
Presenter: **Jianhua Zhao**, Yunnan University of Finance and Economics, China

The Bayesian information criterion (BIC), defined as the observed data log likelihood minus a penalty term depending on the sample size, is a popular model selection criterion for factor analysis with complete data. This definition has also been used for incomplete data except for the use of the 'complete' sample size in the penalty term (bic_c). However, it can be seen that this simple substitution implausibly ignores the amounts of missing information inherent in incomplete data. We propose a novel BIC criterion for factor analysis with incomplete data that only uses the actual amounts of observed information in the penalty term (bic_{obs}). Theoretically, we show that bic_{obs} is a large sample approximation of variational Bayesian (VB) lower bound and bic_c is a further approximation of bic_{obs}. We conduct experiments on synthetic and real data sets to compare bic_{obs} and bic_c and the results show that bic_{obs} is more accurate than bic_c, especially when the missing rate is not small.

EO0735: **A censored time series model for responses on the unit interval**
Presenter: **Mauricio Castro**, Universidad de Concepcion, Chile

An autoregressive model for time series is proposed in which the variable of interest lie on the unit interval being subjected to certain threshold values below or above which the measurements are not quantifiable. The model includes the independent beta regression as a special case. We

provide a full Bayesian approach for the estimation of the model parameters using standard Markov Chain Monte Carlo (MCMC) methods, simulating samples for the joint posterior distribution. We discuss the construction of the proposed model and compare it with alternative models using simulated and real data sets.

EO056 Room LSK1003 PERFORMANCE ANALYSIS Chair: Valentin Zelenyuk

EO0315: **On the competitiveness of countries**
Presenter: **Sung Ko Li**, Hong Kong Baptist University, Hong Kong
Co-authors: Chun-kei Tsang
Researchers of competitiveness tend to believe that higher competitiveness will lead to higher growth rate in the long run. Such a relation is crucial to the usefulness of competitiveness indices, but it has not been studied seriously in the literature. Evidence of the two-way causality between competitiveness and economic growth of countries in the world is found. We further found evidence that the causality from economic growth to competitiveness breaks down in developing countries. The result suggests competitiveness studies in developed countries fit in the scope of endogenous growth models, while competitiveness studies in developing countries fit in the scope of exogenous growth models. This can be explained by the effectiveness of their resource allocation. A more effective resource allocation country is more likely to have a growth pattern of endogenous growth models.

EO0622: **Inference for non-parametric stochastic frontier model**
Presenter: **Valentin Zelenyuk**, University of Queensland, Australia
The focus is on statistical testing in the framework of non-parametric and semi-parametric stochastic frontier via local least squares approach. In particular, we are concerned with inference about existence of production inefficiency in general and at particular points of interest as well as inference about significance of factors potentially influencing the conditional mean of efficiency. To reach our goals, we consider and adapt a previous bootstrap-based approach proposed.

EO0695: **Predicting pull and bear markets: On pooling opinions and sharpness**
Presenter: **Shui Ki Wan**, Hong Kong Baptist University, Hong Kong
Using monthly S&P 500 data, we assess the predictability of 20 financial and macroeconomic indicators for the bulls and bears. The time-varying likelihood ratio statistics suggest that some of the indicators do have in-sample predictability but are highly volatile. Benchmarking on the historical average, the out-of-sample statistic quadratic probability scores (QPS) suggest that the static dividend-price ratio, stock variance, non-farm payroll, federal fund rates and recession probabilities models are highly predictive for all the trend definitions. The dynamic model further improves the forecasting performance significantly. Their improvement is mainly attributed to the substantial increment in sharpness. Since investors are likely to be presented with a pool of forecasts, we also consider linear combinations of the probability forecasts. Although they lead to uncalibrated forecasts that are lack of sharpness, we find that the pooled forecasts are still well-calibrated with the sharpness that are close to the best model. Our studies also show that the adaptive opinion pooling using out-of-sample performance is better than other weighting scheme based on in-sample statistics. By putting sharpness into perspective, we find that trading strategy utilizing sharpness can generate a higher portfolio returns than the traditional threshold approach.

EO0724: **Social-economic environment and economic development: Evidence from nonparametric analysis**
Presenter: **Kai Du**, The University of Queensland, Australia
Co-authors: Valentin Zelenyuk
Most endogenous growth theories emphasize the importance of technological progress and the improvement of the social-economic environment. The aim is to investigate the impact of social-economic environment on country level economic efficiency and specifically to focus on the schooling, health status, and institutional arrangement. Using a data panel from 128 countries over the period of 2009 – 2014, we implement a dynamic version of the two-stage approach, which is based on the data envelopment analysis and truncated regression with double bootstrap. Our results confirm that the improvement of the social-economic environment has a positive contribution to the economic efficiency of a country.

EO311 Room LSK1034 RECENT ADVANCES IN JOINT MODELING Chair: Ming Wang

EO0319: **Causality in the joint analysis of longitudinal and survival data**
Presenter: **Lei Liu**, Northwestern University, United States
In many biomedical studies, disease progress is monitored by a biomarker over time, e.g., repeated measures of CD4. The endpoint of interest, e.g., death or diagnosis of a specific disease, is correlated with the longitudinal biomarkers. The causal relation between the longitudinal and time to event data is of interest. We examine the causality in the joint analysis of longitudinal and survival data. We consider four questions: (1) whether the longitudinal biomarker is a mediator between treatment and survival outcome; (2) whether the biomarker is a surrogate marker; (3) whether the relation between biomarker and survival outcome is due to unknown confounder; (4) whether there is a mediator moderator for treatment. We illustrate our methods by two real data.

EO0318: **Generalized bi-clustering analysis for integrative analysis with incorporation of structural information**
Presenter: **Qi Long**, University of Pennsylvania, United States
Advance in technology has enabled generation of multiple types of -omics data in many biomedical and clinical studies, and it is desirable to pool such data in order to improve the power of identifying important molecular signatures. However, such integrative data analyses entail new analytical and computational challenges. To address these challenges, we propose a Bayesian sparse generalized bi-clustering analysis (GBC) which enables integrating multiple omics modalities with incorporation of biological knowledge through the use of adaptive structured shrinkage priors. The proposed methods can accommodate both continuous and discrete data. MCMC and EM algorithms are developed for estimation. Numerical studies are conducted to demonstrate that our methods achieve improved feature selection and prediction in identifying disease subtypes and latent drivers, compared to existing methods.

EO0481: **A Bayesian joint frailty-copula approach for modeling recurrent events and a terminal event**
Presenter: **Ming Wang**, Pennsylvania State University, United States
In practice, recurrent events are always encountered, but sometimes will be censored by a terminal event. The non-informative censoring assumption is violated under this situation, and as a result, we cannot model the recurrent event process alone. The joint frailty model is widely used to jointly model these two processes. However, there exist limitations due to the assumption of conditional independence given a subject-level frailty and indirect estimate of their association. Besides this, Copula is a popular approach to model bivariate time to event processes. In order to relax the conditional independence assumption and estimate the association directly, we propose a joint frailty-copula approach under a Bayesian framework to model the terminal time-to-event process and the recurrent time-to-event process. We show that the joint frailty-copula model is a more generalized model, extended from a nested frailty model. Metropolis-Hastings within Gibbs Sampler algorithm is used for estimating parameters. Extensive simulation studies are performed to evaluate the performance of our method in terms of bias, mean squared error and

robustness. Finally, we apply our method to analyze the Marketscan data, studying the association between recurrent stroke process and the death process.

EO0633: **Joint modeling of repeated measures and competing failure events in a study of chronic kidney disease**
Presenter: **Dawei Xie**, University of Pennsylvania, United States
Co-authors: Wei Yang, Qiang Pan, Harold Feldman, Wensheng Guo
The motivation arises from the Chronic Renal Insufficiency Cohort (CRIC) study to identify risk factors for renal progression in patients with chronic kidney diseases. The CRIC study collects two types of renal outcomes: glomerular filtration rate (GFR) estimated annually and end stage renal disease (ESRD). A related outcome of interest is death which is a competing event for ESRD. A joint modeling approach is proposed to model a longitudinal outcome and two competing survival outcomes. We assume multivariate normality on the joint distribution of the longitudinal and survival outcomes. Specifically, a mixed effects model is fit on the longitudinal outcome and a linear model is fit on each survival outcome. The three models are linked together by having the random terms of the mixed effects model as covariates in the survival models. EM algorithm is used to estimate the model parameters and the non-parametric bootstrap is used for variance estimation. A simulation study is designed to compare the proposed method with an approach that models the outcomes sequentially in two steps. We fit the proposed model to the CRIC data and show that the protein-to-creatinine ratio is strongly predictive of both estimated GFR and ESRD but not death.

EO0720: **Dynamic prediction through landmark survival models**
Presenter: **Liang Li**, University of Texas MD Anderson Cancer Center, United States
Dynamic prediction is the personalized, real-time prediction of the risk of clinical events using longitudinal data. For any subject in the population under study, the prediction can be made at any time during the follow-up, adaptive to the time-varying at-risk population, personalized etiological history, and predictor-outcome association. We will review two popular approaches for dynamic prediction, joint modeling of longitudinal and survival data and landmark survival models, and will present some recent methodological development of the latter. First, we present a semi-parametric kernel based analytical framework for the landmark Cox model, including innovations to model formulation, estimation and prediction accuracy assessment. We also prove that there exists a joint distribution of the longitudinal and survival data that satisfies the modeling assumptions of the landmark Cox model. This result suggests that the landmark Cox model should not be viewed only as a working model or a prediction algorithm; it has probability basis and rigorous asymptotic analysis and simulation work can be conducted. We further generalize the results to the more flexible landmark linear transformation models, of which the landmark Cox model is a special case. We illustrate the proposed methodology by simulations and a data set from the African American Study of Kidney Disease and Hypertension.

EO028 Room LSKG007 FORECASTING ECONOMIC AND FINANCIAL TIME SERIES Chair: Alain Hecq

EO0345: **Parameter uncertainty in confidence and prediction intervals**
Presenter: **Alexander Heinemann**, Maastricht University, Netherlands
Co-authors: Eric Beutner, Stephan Smeekes
Whereas point forecasts and point estimates describe only one possible outcome, intervals provide a full range of likely outcomes to represent uncertainty. Intervals arising from dynamic models typically sidestep the matter of parameter estimation. Attempts to incorporate parameter uncertainty are typically based on the unrealistic assumption of observing two independent processes, where one is used for parameter estimation, and the other conditioned upon to obtain forecasts. The aim is to show how to construct and validate intervals that incorporate the uncertainty due to parameter estimation avoiding this unrealistic assumption. Our proposed solution is based on a simple sample-splitting approach. The analysis is embedded in the context of Markov chains nesting several important models such as ARMA and GARCH. We show that our intervals based on sample splitting merge with the standard intervals resting on the two sample assumption, thereby also providing a proper theoretical justification of these intervals.

EO0358: **Representation, estimation and forecasting of the multivariate index-augmented autoregressive model**
Presenter: **Gianluca Cubadda**, University of Rome Tor Vergata, Italy
Co-authors: Barbara Guardabascio
The aim is to examine the conditions under which each individual series that is generated by a vector autoregressive model can be represented as an autoregressive model that is augmented with the lags of few linear combinations of all the variables in the system. We call this modelling Multivariate Index-Augmented Autoregression (MIAAR). We show that the parameters of the MIAAR can be estimated by a switching algorithm that increases the Gaussian likelihood at each iteration. Since maximum likelihood estimation is known to perform poorly when the number of parameters approaches the sample size, we propose a regularized version of our algorithm to handle a medium-large number of time series. We illustrate the usefulness of the MIAAR modelling both by empirical applications and simulations.

EO0356: **Mixed causal-noncausal autoregressions with strictly exogenous regressors**
Presenter: **Alain Hecq**, Maastricht University, Netherlands
Co-authors: Sean Telg, Joao Victor Issler
Some authors propose mixed autoregressive causal-noncausal (MAR) models to estimate economic relationships involving expectations variables. These structural equations usually imply explosive roots in their autoregressive part but have stationary forward solutions. In previous work, possible exogenous variables in economic relationships are substituted into the error term and are assumed to also follow an MAR process to ensure the MAR structure of the variable of interest. To allow for the impact of exogenous fundamental variables directly, we instead consider a MARX representation which allows for the inclusion of strictly exogenous regressors. We develop the asymptotic distribution of the MARX parameters. We assume a Student's t-likelihood to derive closed form solutions of the corresponding standard errors. By means of Monte Carlo simulations, we evaluate the accuracy of MARX model selection based on information criteria. We investigate the influence of the US exchange rate and the US industrial production index on several commodity prices.

EC0783: **Combining high-dimensional multivariate volatility forecasts**
Presenter: **Alessandra Amendola**, Department of Economics and Statistics - University of Salerno, Italy
Co-authors: Vincenzo Candila, Giuseppe Storti
In a multivariate volatility framework, several options are available to estimate the conditional covariance matrix of returns. The effect of model uncertainty on multivariate volatility prediction is investigated. In particular, we investigate the profitability of combining forecasts coming from different model structures that are estimated using information at various frequencies. The accuracy of different combination strategies are assessed by means of an extensive application to portfolio allocation for a panel of U.S. stocks.

EO194 Room LSK1011 NONPARAMETRIC METHODS FOR VARIABILITY ESTIMATION Chair: Bo Li

EO0352: **Extremal notion of depth and central regions for functional data**
Presenter: **Naveen Narisetty**, University of Illinois at Urbana-Champaign, United States
There has been extensive work on data depths and their applications in the context of multivariate data. However, depth notions for infinite-dimensional objects such as functional data have received less attention. We propose a new notion of depth called Extremal Depth (ED) for functional data, discuss its properties, and compare its performance with existing concepts such as integrated data depth and band depth. The proposed notion is based on a measure of extreme outlyingness, similar to that for projection depth in the multivariate case. ED has many desirable properties as a measure of depth and is well suited for obtaining central regions of functional data. For constructing central regions, ED satisfies two important properties that are not shared by other notions: a) the central region achieves the nominal (desired) simultaneous coverage probability; and b) the width of the simultaneous region is proportional to that of the pointwise central regions. The usefulness of the method is demonstrated for constructing functional boxplots and detecting outliers.

EO0617: **Comparison between spatio-temporal random processes and application to climate model data**
Presenter: **Bo Li**, University of Illinois at Urbana-Champaign, United States
Comparing two spatio-temporal processes are often a desirable exercise. For example, assessments of the difference between various climate models may involve the comparisons of the synthetic climate random fields generated as simulations from each model. We develop rigorous methods to compare two spatio-temporal random processes both in terms of moments and in terms of temporal trend, using the functional data analysis approach. A highlight of our method is that we can compare the trend surfaces between two random processes, which are motivated by evaluating the skill of synthetic climate from climate models in terms of capturing the pronounced upward trend of real observational data. We perform simulations to evaluate our methods and then apply the methods to compare different climate models as well as to evaluate the synthetic temperature fields from model simulations, with respect to observed temperature fields.

EO0476: **On the choice of difference sequence in a unified framework for variance estimation in nonparametric regression**
Presenter: **Tiejun Tong**, Hong Kong Baptist University, Hong Kong
Difference-based methods do not require estimating the mean function in nonparametric regression and are therefore popular in practice. We propose a unified framework for variance estimation that combines the linear regression method with the higher-order difference estimators systematically. The unified framework has greatly enriched the existing literature on variance estimation that includes most existing estimators as special cases. More importantly, the unified framework has also provided a smart way to solve the challenging difference sequence selection problem that remains a long-standing controversial issue in nonparametric regression for several decades. Using both theory and simulations, we recommend to use the ordinary difference sequence in the unified framework, no matter if the sample size is small or if the signal-to-noise ratio is large. Finally, to cater for the demands of the application, we have developed a unified R package, named VarED, that integrates the existing difference-based estimators and the unified estimators in nonparametric regression and have made it freely available in the R statistical program http://cran.r-project.org/web/packages/.

EO0619: **Approximate inferences for massive space-time data**
Presenter: **Hao Zhang**, Purdue University, United States
For a large space-time dataset, approximation methods are often applied to compute and maximize the objective function which could be, for example, the negative likelihood function. Sometimes this is done deliberately in order to achieve the computational efficiency. We will focus on the situation when the approximation is necessary due to the ill condition of the covariance matrix.

EO090 Room LSK1027 RECENT DEVELOPMENTS IN TIME SERIES ANALYSIS AND RELATED TOPICS Chair: Sangyeol Lee

EO0410: **On the dependence between frequency-severity and bonus-malus systems**
Presenter: **Jae Youn Ahn**, Ewha Womans University, Korea, South
Co-authors: Sojung Park, Joseph Hyun Tae Kim, Woojoo Lee
Auto rate models traditionally have assumed independence between claim frequency and severity. With the invention of compound risk models which can accommodate dependence between claim frequency and severity, series of recent insurance literature revealed the dependence between frequency and severity. We first study the property of various statistical models which can accommodate the dependence between frequency and severity. With the help of such models, we found that the sign of dependence varies; liability claim data showed positive dependence whereas collision data showed negative dependence. We hypothesize that the dependence can be driven by bonus hunger phenomenons in bonus-malus systems, and this can explain the different dependency results from liability and collision data. Through the simulation study and data analysis, we confirm the hypothesis.

EO0372: **On-line estimation of case fatality rate of South Korean MERS using run-off triangle data**
Presenter: **Sungim Lee**, Dankook University, Korea, South
Co-authors: Johan Lim
The motivation comes from the recent Korean MERS outbreak. We propose an easy on-line estimation procedure for the case fatality rate- the proportion of deaths among the cases during the course of an epidemic disease - which is an important indicator of the severity of a disease. The key step in our procedure is the run-off triangle, which simultaneously takes into account two time axes, the calendar and disease-duration times. We restructure the original data into the run-off triangle, each of whose cell contains the numbers of patients cured, recessed, and still in disease status at a given calendar and disease-duration time combination. Based on this restructured run-off triangle data, we propose an on-line estimator of the CFR. We numerically compare the performance of the proposed estimator to two existing estimators in the literature. Finally, we apply our procedure to the 2015 Korean MERS outbreak data.

EO0430: **Wild bootstrap Ljung box test for cross correlations of multivariate time series**
Presenter: **Taewook Lee**, Hankuk University of Foreign Studies, Korea, South
The conventional Ljung-Box test for financial time series with ARCH effect (also known as conditional heteroscedasticity) is well-known to suffer from severe size distortions. The objective is to develop a wild bootstrap-based Ljung-Box test for crosscorrelations in mean of multivariate time series. According to our simulation study, the wild bootstrap-based Ljung Box test succeeds to achieve correct sizes and comparable powers in the presence of ARCH effect.

EO0404: **Information criterion and change point detection for general causal time series**
Presenter: **Chi Tim Ng**, Chonnam National University, Korea, South
It is illustrated that a general class of information criteria can be used in the change point analysis of a broad class of causal time series. In particular, asymptotic theory is established guaranteeing that the number of change points, locations of the change points, and the parameters of each regime can be estimated consistently. Computational issues are also discussed. Both dynamic programming approach and penalized likelihood approach

based on non-concave penalty function can be used to search the optimal model. The pros and cons of these methods are compared.

EO172 Room LSK1007 STATISTICAL INFERENCE AND THEIR APPLICATIONS TO COMPLEX PROBLEMS Chair: Dungang Liu

EO0425: **Confidence distribution a new inference tool for bridging Bayesian, frequentist and fiducial (BFF) inferences**
Presenter: **Min-ge Xie**, Rutgers University, United States

A confidence distribution (CD) is a sample-dependent distribution function that can serve as a distribution estimate, contrasting with a point or interval estimate, of an unknown parameter. It can represent confidence intervals (regions) of all levels for the parameter. It is to provide "simple and interpretable summaries of what can reasonably be learned from data", and it can give meaningful answers for all questions in statistical inference. An emerging theme is "Any statistical approach, regardless of being frequentist, fiducial or Bayesian, can potentially be unified under the concept of confidence distributions, as long as it can be used to derive confidence intervals of all levels, exactly or asymptotically". We articulate the logic behind the developments, and show how CD can potentially bridge posterior probabilistic inferences in Bayesian, frequentist and fiducial (BFF) schools in all aspects, including estimation, testing and prediction. If times allows, we will also present examples to show that the developments in CD lead to useful inference tools for statistical problems where methods with desirable properties have not been available or could not be easily obtained.

EO0415: **A probabilistic measure of cost-effectiveness for health economic evaluation**
Presenter: **Thomas Mathew**, University of Maryland Baltimore County, United States

Identifying treatments or interventions that are cost-effective (more effective at lower cost) is clearly important in health policy decision making, especially for allocating health care resources. Various measures of cost-effectiveness that are informative, intuitive and simple to explain have been suggested in the literature, along with statistical inference concerning them. Popular and widely used measures include the incremental cost-effectiveness ratio (ICER), defined as the ratio between the difference of expected costs and the difference of expected effectiveness in two populations receiving two treatments, and the incremental net benefit (INB), which is the difference between the incremental cost and the incremental effectiveness after multiplying the latter with a "willingness-to-pay parameter". Both ICER and INB are functions of population means, and inference concerning them has been widely investigated under a bivariate normal distribution, or under a log-normal/normal distribution for the cost and effectiveness measures. These will be briefly reviewed and an alternative probability-based approach will be introduced, referred to as cost-effectiveness probability (CEP), which is the probability that the first treatment will be less costly and more effective compared to the second one. Inference on the CEP will be discussed. Numerical results and illustrative examples will be given.

EO0416: **A sequential split-conquer-combine approach for Gaussian process modeling in computer experiments**
Presenter: **Ying Hung**, Rutgers University, United States
Co-authors: Min-ge Xie

Gaussian process (GP) models are widely used in the analysis of computer experiments. However, two critical issues remain unresolved. One is the computational issue in GP estimation and prediction where intensive manipulations of an n-by-n correlation matrix are required and become infeasible for large sample size n. The other is how to improve the naive plug-in predictive distribution which is known to underestimate the uncertainty. We introduce an unified framework that can tackle both issues simultaneously. It consists of a sequential split-conquer procedure, an information combining technique using confidence distributions (CD), and a CD-based predictive distribution. This framework provides estimators and predictors that maintain the same asymptotic efficiency as the conventional method but reduce the computation dramatically. The CD-based predictive distribution contains comprehensive information for statistical inference and provides a better quantification of predictive uncertainty comparing with the plug-in approach. Simulations are conducted to evaluate the accuracy and computational gains. The proposed framework is demonstrated by a data center example based on tens of thousands of computer experiments generated from a computational fluid dynamic simulator.

EO0677: **Surrogate residuals and diagnostics for regression model with an ordinal response**
Presenter: **Heping Zhang**, Yale University, United States
Co-authors: Dungang Liu

Ordinal outcomes are common in scientific research and everyday practice, and we often rely on regression models to make inferences. A long-standing problem with such regression analyses is the lack of effective diagnostic tools for validating model assumptions. The difficulty arises from the fact that an ordinal variable has discrete values that are labeled with, but not, numerical values. The values merely represent ordered categories. We propose a surrogate approach to defining residuals for an ordinal outcome Y. The idea is to define a continuous variable S as a "surrogate" of Y and then obtain residuals based on S. For the general class of cumulative link regression models, we study the residual's theoretical and graphical properties. We show that the residual has null properties similar to those of the common residuals for continuous outcomes. Our numerical studies demonstrate that the residual has power to detect misspecification with respect to 1) mean structures; 2) link functions; 3) heteroscedasticity; 4) proportionality; and 5) mixed populations. The proposed residual also enables us to develop numeric measures for goodness-of-fit using classical distance notions. Our results suggest that compared to a previously defined residual, our residual can reveal deeper insights into model diagnostics.

EO242 Room LSK1033 BAYESIAN MODELING FOR SPATIOTEMPORAL PHENOMENA Chair: Fumiyasu Komaki

EO0559: **Neural decoding based on an infinite mixture model**
Presenter: **Ryohei Shibue**, NTT Communication Science Laboratories, Japan
Co-authors: Fumiyasu Komaki

Neural decoding is a framework for reconstructing external stimuli from spike trains recorded by various neural recordings. We propose a neural decoding method based on an infinite mixture model and Bayesian nonparametrics. The proposed method improves decoding performance in terms of accuracy and computational speed. We apply the proposed method to simulation and experimental data to verify its performance.

EO0583: **Spatial distribution of coefficients of variation and Bayesian forecasts for recurrence intervals of earthquakes**
Presenter: **Shunichi Nomura**, The Institute of Statistical Mathematics, Japan
Co-authors: Yosihiko Ogata

A Bayesian method is proposed for probability forecasting for recurrent earthquakes of inland active faults in Japan. Renewal processes with the Brownian Passage Time (BPT) are applied for over a half of active faults in Japan by the Headquarters for Earthquake Research Promotion (HERP) of Japan. Long-term forecast with the BPT distribution needs two parameters; the mean and coefficient of variation (COV) for recurrence intervals. The HERP applies a common COV parameter for all of these faults because most of them have only one or a few specified paleoseismic events, which is not enough to estimate reliable COV values for respective faults. However, errors in COV estimates can make critical bias in forecasts and so COVs should be carefully selected for individual faults. The COVs of recurrence intervals depend on stress perturbation from nearby seismicity and have spatial trends. Thus we introduce a spatial structure on its COV parameter by Bayesian modeling with a Gaussian process prior. The COVs on active faults are correlated and take similar values for closely located faults. It is found that the spatial trends in the estimated COV values

coincide with the density of active faults in Japan. We also show Bayesian forecasts by the proposed model using MCMC methods. Our forecasts are different from HERP's forecast especially on the active faults where HERP's forecasts are very high or low.

EO0755: **Bayesian estimation of dynamic relationship between GDP and economic indicators for analyzing business cycles**
Presenter: **Koki Kyo**, Obihiro University of Agriculture and Veterinary Medicine, Japan
Co-authors: Hideo Noda
A traditional approach for analyzing business cycles is that uses a diffusion index, which is constructed based on several selected economic indicators. A problem of this approach is lack of analysis of the lead-lag relation and the difference of importance between the economic indicators. We analyze the lead-lag relation and the difference of importance between the economic indicators by taking GDP as a basis. We extract a stationary component from time series for GDP and each economic indicator using a set of state space models. The extracted stationary components are regarded as a kind of signal for business cycle analysis, then we estimate dynamic relationship between the stationary components in each economic indicator and GDP using Bayesian modeling methods. Thus, the lead-lag relation and the difference of importance between the economic indicators can be analyzed based on the estimates. It can also be aimed at estimating monthly series of GDP by extending the newly proposed approach.

EO0794: **Application of data assimilation to particle simulation and point process**
Presenter: **Kazuyuki Nakamura**, Meiji University, Japan
Co-authors: Noriho Fujioka, Yutaka Kono
Data assimilation is a method which combines a physical simulation system with observation data to obtain more precise simulation and prediction results. It has been developed in meteorology and oceanography, and recently its application fields are widely spreading such as geotechnical engineering, epidemiology, ecology and molecular dynamics. Data assimilation can be interpreted as Bayesian calculation and sequential Bayesian filtering, and therefore appropriate modeling of the prior and the system noise is important for obtaining suitable estimation. We show the rule for prior design, especially focusing on the constraint conditions for the real systems. Based on the above, we introduce several approaches of prior design for data assimilation of particle simulation and point process such as traffic flow and event occurrence. One of the approaches is the use of bounded Gaussian and uniform mixture, which can manage both fast estimation in sudden parameter change and estimation stability. The validity of the approach is also shown by numerical experiments.

EO106 Room LSK1010 LARGE-SCALE REGRESSION METHODS AND ALGORITHMS Chair: Jian Kang

EO0613: **A parallel algorithm for large-scale penalized quantile regression**
Presenter: **Nan Lin**, Washington University in St. Louis, United States
Penalized quantile regression (PQR) provides a useful tool for analyzing high-dimensional data with heterogeneity. However, its computation is challenging due to the nonsmoothness and (sometimes) the nonconvexity of the objective function. An iterative coordinate descent algorithm (QICD) was recently proposed to solve PQR with nonconvex penalty. The QICD significantly improves the computational speed but requires a double-loop. We propose an alternative algorithm based on the alternating direction method of multiplier (ADMM). By writing the PQR into a special ADMM form, we can solve the iterations exactly without using coordinate descent. This results in a new single-loop algorithm, which we refer to as the QPADM algorithm. The QPADM demonstrates favorable performance in both computational speed and statistical accuracy, particularly when the sample size n and/or the number of features p are large.

EO0381: **Kriging over space and time based on a latent low-dimensional structure**
Presenter: **Rongmao Zhang**, Zhejiang University, China
A new approach is proposed to represent nonparametrically the linear dependence structure of a spatio-temporal process in terms of latent common factors. Though it is formally similar to the existing reduce drank approximation methods, the fundamental difference is that the low-dimensional structure is completely unknown in our setting, which is learned from the data collected irregularly over space but regularly over time. We do not impose any stationarity conditions over space either, as the learning is facilitated by the stationarity in time. Krigings over space and time are carried out based on the learned low-dimensional structure. Their performance is further improved by a newly proposed aggregation method via randomly partitioning the observations accordingly to their locations. A low-dimensional correlation structure also makes the kriging methods scalable to the cases when the data are taken over a large number of locations and/or over a long time period. Asymptotic properties of the proposed methods are established. Illustration with both simulated and real data sets is also reported.

EO0833: **Variable selection for mixture and promotion time cure rate models**
Presenter: **Zhangsheng Yu**, Shanghai Jiao Tong University, China
Co-authors: Abdullah Masud, Wanzhu Tu
Failure-time data with cured patients are common in clinical studies. Data from these studies are typically analyzed with cure rate models. Variable selection methods have not been well developed for cure rate models. We propose two Least Absolute Shrinkage and Selection Operators (LASSO) based methods, for variable selection in mixture and promotion time cure models with parametric or nonparametric baseline hazards. We conduct an extensive simulation study to assess the operating characteristics of the proposed methods. We illustrate the use of the methods using data from a study of childhood wheezing.

EO0841: **Finding latent signals of dynamic correlation in high-throughput expression data**
Presenter: **Tianwei Yu**, Emory University, United States
In high-throughput biological data, dynamic correlation, i.e. changing correlation patterns under different biological conditions, can reveal important regulatory mechanisms. Current methods seek underlying conditions of dynamic correlation by using certain genes as surrogate signals. We describe a new method that directly identifies strong latent signals that regulate the dynamic correlation of many pairs of genes, named LDCA: Latent Dynamic Correlation Analysis. We validate the performance of the method with extensive simulations. In real data analysis, the method reveals biologically plausible latent factors that were not found by existing methods.

EO198 Room LSK1014 RECENT DEVELOPMENT IN STATISTICAL ANALYSIS OF FUNCTIONAL AND IMAGE DATA Chair: Chunzheng Cao

EO0625: **Roust clustering and prediction on functional data based on a heavy-tailed process**
Presenter: **Chunzheng Cao**, Nanjing University of Information Science and Technology, China
A functional mixed-effects model is introduced by using heavy-tailed processes for clustering functional data. The mixing proportions are defined through a logistic regression. The model can give robust inferences against data contamination or distribution misspecification, and hence yield precise clustering results. We propose an efficient algorithm for parameter estimation, and generate predictions and bootstrap-based confidence intervals for the subject-specific inferences. Numerical results including simulation studies and a real data application are presented.

EO0632: **Brain MRI analysis based on a finite mixture model**
Presenter: **Yunjie Chen**, Nanjing University of Information Science and Technology, China

An improved anisotropic multivariate student t-distribution based hierarchical fuzzy c-means method is proposed. Firstly, improved anisotropic spatial information, defined in the neighborhoods of each pixel, is proposed to overcome the effect of the noise and preserve more detail information, especially for points with less repetitive patterns, such as corner and end points. Secondly, the improved anisotropic spatially information is utilized into a negative multivariate student t-distribution based log-posterior as the dissimilarity function to improve the robustness and accuracy. Thirdly, we use the hierarchical strategy to construct a more flexible objective function by considering the improved dissimilarity function itself as a sub-FCM, to make the method more robust and accurate to outliers and weak edges. Finally, the intensity inhomogeneities is modeled as a linear combination of a set of orthogonal basis functions, and parameterized by the coefficients of the orthogonal basis functions. Then the objective function is integrated with the bias field estimation and makes the proposed method can estimate the bias field meanwhile segmenting images. The segmentation and the bias field estimation can obtain benefit from each other. Our statistical results on both synthetic and clinical images show that the proposed method can overcome the difficulties caused by noise and bias fields and obtain more accurate results.

EO0631: **Nonlinear mixed-effects scalar-on-function models and variable selection for kinematic upper limb movement data**
Presenter: **Yafeng Cheng**, MRC Biostatistics Unit, United Kingdom
Co-authors: Jian Qing Shi, Janet Eyre

Motivated by a collaborative research about modelling clinical assessments of upper limb function after stroke using 3D kinematic data, we present a new nonlinear mixed-effects scalar-on-function regression model with a Gaussian process (GP) prior focusing on variable selection from large number of candidates including both scalar and function variables. A novel variable selection algorithm has been developed, namely functional least angle regression (fLARS). As they are essential for this algorithm, we studied the representation of functional variables with different methods and the correlation between a scalar and a group of mixed scalar and functional variables. When the algorithm was applied to the analysis of the 3D kinetic movement data the use of the nonlinear random-effect model and the function variables significantly improved the prediction accuracy for the clinical assessment.

EO0629: **Symmetry and asymmetry of the brain graph: Volume entropy view point**
Presenter: **Kanghyun Choi**, Seoul National University, Korea, South

Brain is a paired organ. It consists of the right and the left brain. It looks symmetric, but is it really so? We will explore this question using fMRI human brain data and explain in terms of the ‘Volume entropy". We will explain how brain graph can be symmetric and asymmetric depending on the connectivity level between the right brain and the left brain.

Saturday 17.06.2017 16:10 - 17:50 Parallel Session O – EcoSta2017

EO082 Room LSK1007 ADVANCES IN HIGH-DIMENSIONAL DATA ANALYSIS Chair: Hao Chen

EO0155: **Graphical models for discrete and continuous data**
Presenter: **Johannes Lederer**, University of Washington, United States
Dependencies in multivariate observations are a unique gateway to uncovering relationships among processes. An approach that has proved particularly successful in modeling and visualizing such dependence structures is the use of graphical models. However, whereas graphical models have been formulated for finite count data and Gaussian-type data, many other data types prevalent in the sciences have not been accounted for. For example, it is believed that insights into microbial interactions in human habitats, such as the gut or the oral cavity, can be deduced from analyzing the dependencies in microbial abundance data, a data type that is not amenable to standard classes of graphical models. We present a novel framework that unifies existing classes of graphical models and provides other classes that extend the concept of graphical models to a broad variety of discrete and continuous data, both in low- and high-dimensional settings. Moreover, we present a corresponding set of statistical methods, theoretical guarantees, and software that allows for efficient estimation and inference in the framework.

EO0187: **Community detection with nodal information**
Presenter: **Yang Feng**, Columbia University, United States
Community detection is one of the fundamental problems in the study of network data. Most existing community detection approaches only consider edge information as inputs, and the output could be suboptimal when nodal information is available. In such cases, it is desirable to leverage nodal information for the improvement of community detection accuracy. Towards this goal, we propose a flexible network model incorporating nodal information, and develop likelihood-based inference methods. For the proposed methods, we establish favorable asymptotic properties as well as efficient algorithms for computation. Numerical experiments show the effectiveness of our methods in utilizing nodal information across a variety of simulated and real network data sets.

EO0256: **Homogeneity test of covariance matrices with high-dimensional longitudinal data**
Presenter: **Ping-Shou Zhong**, Michigan State University, United States
Co-authors: Runze Li
High-dimensional longitudinal data appear when a large number of variables are measured repeatedly over time. An important challenge of such data is the existence of both spatial and temporal dependence. The focus is on testing the temporal homogeneity of covariance matrices in high-dimensional longitudinal data with temporospatial dependence. The data dimension (p) is allowed to be much larger than the number of individuals (n). A new test statistic is proposed and the asymptotic distribution is established. If the covariance matrices are not homogeneous, an estimator for the location of the change point is given whose rate of convergence is established and shown to depend on p, n and the signal-to-noise ratio. The proposed method is extended to estimate multiple change points by applying a binary segmentation approach, which is shown to be consistent under some mild conditions. Simulation studies and an application to a time-course microarray data set are presented to demonstrate the performance of the proposed method.

EO0570: **Selecting the number of principal components: Estimation of the true rank of a noisy matrix**
Presenter: **Yunjin Choi**, National University of Singapore, Singapore
Co-authors: Jonathan Taylor, Robert Tibshirani
Principal Component Analysis (PCA) is one of the most popular methods in multivariate data analysis. Despite the popularity of the method, there is no widely adopted standard approach to select the number of principal components to retain. To address this issue, we propose a novel method utilizing the hypothesis testing framework and test whether the currently selected principal components capture all the statistically significant signals in the given data set. While existing hypothesis testing approaches do not enjoy the exact type 1 error property and lose power under some scenarios, the proposed method provides an exact type 1 error control along with decent size of power in detecting signals. Central to our work is the post-selection inference framework which facilitates valid inference after data-driven model selection; the proposed hypothesis testing method provides exact type 1 error controls by conditioning on the selection event which leads to the inference. We also introduce a possible extension of the proposed method for high-dimensional data.

EO120 Room LSK1033 NONLINEAR TIME SERIES Chair: Philip Yu

EO0174: **Group orthogonal greedy algorithm for change-point estimation of multivariate time series**
Presenter: **Yuanbo Li**, The Chinese University of Hong Kong, Hong Kong
Co-authors: NH Chan, Chun Yip Yau
A three steps method is proposed for the detection of multiple structural breaks in piecewise stationary vector autoregressive processes. The number of the structural breaks can be large and unknown. Moreover, the number and the location of the breaks are not necessarily the same in different components. The proposed method is based on a connection between the structural break problem and the high dimensional regression problem. With such a connection, we develop a group orthogonal greedy algorithm for efficient estimation of structural breaks inspired by the context of high dimensional variable selection. A high-dimensional information criterion is proposed to detect different structural breaks in different components. We prove the consistency of the estimators and provide Monte Carlo experiments for the finite sample performance.

EO0244: **Buffered vector error-correction models**
Presenter: **Renjie Lu**, The University of Hong Kong, Hong Kong
Co-authors: Philip Yu
Buffered autoregressive models are extended to buffered vector error-correction model (VECM). We propose a SupWald test for the presence of buffer-type threshold effect. We derive the null asymptotic distribution, and show how to make use of a bootstrap method to obtain the p-value. Least squares estimation is discussed, and the consistency of the corresponding estimators is derived. Furthermore, we discuss three different methods to determine the number of cointegrating vector for each regime. We investigate the effectiveness of the test and estimation by simulation studies. We apply our model to two data sets: the monthly bond rates of United States and high frequency data of futures and spot of S&P 500. We find the evidences that there exists buffering regime, and buffered VECM is more reasonable than the traditional two- and three-regime threshold VECM.

EO0241: **A nonparametric composite likelihood approach to multiple change-point problems**
Presenter: **Philip Yu**, The University of Hong Kong, Hong Kong
The non-parametric multiple change-point problem is considered by first embedding it in a parametric framework. We construct a composite likelihood function and search for the change points using a binary segmentation algorithm. The methodology is applied to both simulated data and human genomic data. It is seen that the proposed method performs very well and is able to identify change points previously undetected.

EO0248: **A dynamic matrix factor model for multivariate realized volatility**
Presenter: **Xiaohang Wang**, LIM Advisors Limited, China
Co-authors: Philip Yu
Modeling multivariate volatility in moderately large dimension with a feasible dynamic structure remains a challenging problem. With the availability of intra-day data, daily realized covariance matrices constructed form high-frequency data have started to be studied. The existing matrix factor (MFA) model transforms the realized covariance matrices to a lower dimensional subspace and reduce the number of parameters to be forecast. However, the loading matrix of MFA is constant over the time. We propose a new method, the dynamic matrix factor (DMF) model, for which the loading matrix is time-varying. The realized covariance matrices are assumed to follow a Wishart distribution, and the scale matrix adopts a spectral decomposition. An loading-driving process with scalar BEKK model are used to capture the dynamics of loading matrix, and the diagonal matrix is modeled by GARCH(1,1). We will further illustrate DMF model with both simulation and real data. Empirical studies of small and medium-size of daily realized covariance matrices process from the New York Stock Exchange demonstrate the benefits of the dynamic structure of DMF, and show that DMF outperforms the existing methods.

EO040 Room LSK1003 ENDOGENEITY AND NONPARAMETRICS IN MODELS OF PRODUCTION Chair: Artem Prokhorov

EO0181: **Endogenous environmental variables in stochastic frontier models**
Presenter: **Artem Prokhorov**, University of Sydney, Australia
Previous papers have considered the case that some of the inputs in a stochastic frontier model may be endogenous because they are correlated with either or both parts of the stochastic frontier model error term. These papers assumed that the errors were i.i.d. across observations. We consider the case that that the inefficiency term depends on some environmental variables that do not influence the frontier output, but which do influence the level of inefficiency. Specifically we assume a scaling model for the inefficiency. We first consider the case that some components of inputs and environmental variables are endogenous because they are correlated with the symmetric component of the error, though not with the inefficiency component. This case is easily handled. We then turn to the more novel and difficult case that some inputs and environmental variables are endogenous because they are correlated with the inefficiency component as well as possibly with the symmetric error component. We show how to estimate the model by IV and also by MLE. Neither method is simple, because a specific copula must be assumed to model the correlation of error components with the endogenous variables, and because simulation methods are necessary to form the IV criterion function or the likelihood. A novel and potentially important point is made: although endogeneity complicates estimation of the model, it also enables more precise prediction of the inefficiencies.

EO0218: **A partially parametric model**
Presenter: **Daniel Henderson**, University of Alabama and Nankai University, United States
Co-authors: Christopher Parmeter
A model is proposed which includes both a known (potentially) nonlinear parametric component and an unknown nonparametric component. This approach is feasible given that we estimate the finite sample parameter vector and the bandwidths simultaneously. We show that our objective function is asymptotically equivalent to the individual objective criteria for the parametric parameter vector and the nonparametric function. In the special case where the parametric component is linear in parameters, our single-step method is asymptotically equivalent to the two-step partially linear model estimator. Monte Carlo simulations support the asymptotic developments and show impressive finite sample performance. We apply our method to the case of a partially constant elasticity of substitution production function for an unbalanced sample of 134 countries from 1955-2011 and find that the parametric parameters are relatively stable for different nonparametric control variables in the full sample. However, we find substantial parameter heterogeneity between developed and developing countries which results in important differences when estimating the elasticity of substitution.

EO0233: **On the estimation of latent structure stochastic frontier panel data models**
Presenter: **Kien C Tran**, University of Lethbridge, Canada
A fixed-effects stochastic frontier panel data model with group-specific pattern heterogeneity in the slope coefficients is considered. In particular, we assume the slope coefficients are heterogeneous across groups but homogeneous within the group, and group membership is unknown. A Lasso-type penalized profile likelihood (PPL) estimation is proposed when the regressors are exogenous. Under endogeneity, a penalized GMM (PGMM) and/or penalized profile limited information maximum likelihood (PPLIML) is considered. Asymptotic properties of the proposed estimators for both scenarios are discussed. Monte Carlo simulations are performed to assess the finite sample performances of the proposed estimators. An empirical application is used to demonstrated the usefulness of the proposed models and methods in practice.

EO0312: **A time-varying true individual effects model with endogenous regressors**
Presenter: **Levent Kutlu**, Georgia Institute of Technology, United States
Co-authors: Kien C Tran, Mike Tsionas
A fairly general individual effects stochastic frontier model is proposed, which allows both heterogeneity and inefficiency to change over time. Moreover, our model handles the endogeneity problems if either at least one of the regressors or one-sided error term is correlated with the two-sided error term. Our Monte Carlo experiments show that our estimator performs well. We employed our methodology to the US banking data and found a positive relationship between return on revenue and cost efficiency. Estimators ignoring time-varying heterogeneity or endogeneity did not perform well and gave very different estimates compared to our estimator.

EO202 Room LSK1011 NONPARAMETRIC AND SEMI PARAMETRIC STATISTICS AND THEIR APPLICATIONS Chair: Wenbo Wu

EO0210: **Distance metrics for measuring joint dependence with applications**
Presenter: **Xianyang Zhang**, Texas A&M University, United States
New metrics which generalize the notion of distance covariance are proposed to quantify interaction and joint dependence among $d \geq 2$ random variables of arbitrary dimensions. We introduce the concept of high order distance covariance to measure Lancaster interactions. It is shown that linear combination of distance covariance and its higher order counterparts quantifies joint independence. Empirical estimators are constructed based on certain Euclidean distances between sample elements. We study the large sample properties of our estimators and propose a bootstrap procedure to approximate their sampling distributions. The new metrics are employed to check the goodness of fit for directed acyclic graph and to test the existence of independent components. Numerical studies demonstrate the effectiveness of our method.

EO0599: **A note on inverse regressions when responses are missing at random**
Presenter: **Cheng Yong Tang**, Temple University, United States
Co-authors: Yuexiao Dong, Qi Xia
Inverse regressions constitute a class of sufficient dimension reduction methods targeting at estimating the central space by regression-type approaches implemented inversely on the predictors and the responses. The most representative approach in this family is the seminal Sliced Inverse Regression (SIR) approach. We first show that missing responses generally affect the validity of the inverse regressions under the scheme of the

so-called missing at random, in the sense that the resulting estimations for the central space can be biased if data with missing responses are simply ignored. We then propose two simple and effective adjustments for missing responses that guarantees the validity of the inverse regressions. The proposed methods share the essence and simplicity of the inverse regressions. We demonstrate the performance of the proposed inverse regressions for dealing with missing responses by numerical and theoretical analyses.

EO0658: **Combining eigenvalues and variation of eigenvectors for order determination**
Presenter: **Bing Li**, The Pennsylvania State University, United States
In applying many statistical methods, such as principal component analysis, canonical correlation analysis, and sufficient dimension reduction, we need to determine how many eigenvectors of a random matrix are important for estimation. Such problems are known as order determination, and amount to estimating the rank of a matrix. Previous order-determination procedures rely either on the decreasing pattern, or elbow, of the eigenvalues, or on the increasing pattern of the variability in the directions of the eigenvectors. We propose a new order-determination procedure by exploiting both patterns: when the eigenvalues of a random matrix are close together, their eigenvectors tend to vary greatly; when the eigenvalues are far apart, their variability tends to be small. The combination of both helps to pinpoint the rank of a matrix more precisely than the previous methods. We establish the consistency of the new order-determination procedure, and compare it with other such procedures by simulations and in an applied setting.

EO0712: **Whittle likelihood estimation of the semi-varying coefficient dynamic panel data model with serially correlated errors**
Presenter: **Yingcun Xia**, NUS, Singapore
Co-authors: Lei Huang
In time series modelling, serial correlation in the regression errors may cause serious problems in estimation and inference. The problem becomes even worse for panel data with spatial and dynamic structure. However, most of the existing literature does not take this problem into consideration. The difficulty comes from the fact that, unlike what most estimation methods require, the conditional expectation of the errors on the regressors is not zero. Based on an extension of Whittle likelihood, the aim is to study a semi-parametric dynamic model with ARMA errors for panel data. Asymptotic normality of the estimator is established. Both simulated data and real data analyses indicate that incorporating serial correlation in residuals can improve estimation efficiency and prediction ability of the models.

EO188 Room LSK1001 APPLICATIONS AND EMPIRICAL RESEARCH IN ECONOMICS AND FINANCE Chair: Tsung-Chi Cheng

EO0258: **Empirical analysis on validity of economic approaches to environmental issues**
Presenter: **Yuejun Zheng**, Doshisha University, Japan
Various environmental issues caused by the economic development in China, Japan and South Korea, as the major powers of the East Asia, have been aggravating in both local and global levels in the past decades. It is needless to say that the improvement of individual consciousness and behavior on environment is an indispensable way to resolve environmental issues. In particular, the economic instrument has always been considered as a legal restraint to regulate the organizational or individual behaviors connected with the environmental destruction. The focus is on analyzing the validity of economic approaches, including environmental tax, decline in material comfort and control of economic growth, based on a data set collected from an East Asian cross-national survey on consciousness toward culture, life, and environment in 2011. Results derived from multivariate analysis on question items concerning consciousness toward environmental issues, economic development and their mutual relationships, have shown that the roles and limits of economic approach in different socioeconomic situations.

EO0496: **The spillover effects of US unconventional monetary policy on the Taiwanese economy**
Presenter: **Hsing-Hua Chang**, National Chengchi University, Taiwan
Co-authors: KuanChieh Chen
Due to the severity of the 2007-2009 financial crisis, the Federal Reserves of the United States aggressively lowered policy rate to zero and adopted several unconventional monetary policies, including Quantitative Easing. It has long suggested that US monetary policies had strong spillover effects on foreign countries. And indeed the governor of Taiwan's central bank expressed strong concerns of such ripple effects of US policy on the Taiwanese economy. We follow previous research to use Fed unconventional monetary policy announcements to check if US monetary policy has significant impacts on the Taiwanese economy and financial markets. We also compare these spillover effects of traditional monetary policies before the crisis with the unconventional policies implemented during and after the crisis. The results show that indeed US monetary policies exert significant impacts on the Taiwanese economy. We also discuss the Taiwan's central banks responses to these spillover effects.

EO0531: **Improvements for the estimation for the probability of informed trading models**
Presenter: **Hung-Neng Lai**, National Central University, Taiwan
Co-authors: Tsung-Chi Cheng
Although the probability of informed trading models has been widely used in many fields in finance, its estimation is believed to be notoriously difficult. Two advances are made to improve the estimation performance. First, an improved method of initial value selection is proposed to increase the chance of finding a global maximum during maximum likelihood estimation. Second, Poisson distributions are replaced by normal distributions when the numbers of trades in the sample are large. The proposed methods are implemented on the simulated data as well as the U.S. stock market data and are shown to be able to obtain more robust estimations.

EO0473: **Country heterogeneity, happiness and income inequality**
Presenter: **Shinn-Juh Lin**, National Chengchi University, Taiwan
National happiness level is related to income, income inequality and other crucial socioeconomic variables, such as happiness inequality, health condition, and social connection. In addition, countries may differ in many other aspects such as culture and regulations. To obtain an in-depth analysis of relations between happiness level and socioeconomic variables, data of the 1990s and the 2000s drawn from the World Database of Happiness are examined with two estimation approaches - pooled OLS and panel data analysis. The empirical results demonstrate that the role and the statistical significance of income inequality in explaining happiness clearly depends on unobservable country heterogeneity. Based on the pooled OLS, happiness and income inequality are positively related. However, after dividing all the countries into two groups, the relation between happiness and income inequality becomes insignificant for highly-developed countries, while that relation remains significantly positive for underdeveloped countries. Most interestingly, happiness turns out to be inversely related to income inequality with the panel data approach. These results favor the hypothesis that income inequality, instead of encouraging possibility of social mobility, entails more social comparison and hence reduces happiness level.

EO098 Room LSK1009 RECENT DEVELOPMENTS IN ECOLOGICAL STATISTICS Chair: Wen-Han Hwang

EO0280: **Variance estimation of a jackknife estimator of species richness**
Presenter: **Tsung-Jen Shen**, National Chung Hsing University, Taiwan
Species richness in an assemblage is an essential ingredient when monitoring or managing an ecosystem, though a large number of estimators about species richness have been proposing in the literature, this topic is still full of challenges. An important characteristic of the jackknife procedure is used to correct the statistical bias caused by a biased estimator of interest. Apparently, the number of species in a sample is only treated as a lower bound of species richness, so applying the jackknife method to the lower limit can give a series of jackknife estimators for adjusting the resulting bias by the lower limit. Nevertheless, determining the best one among a series of estimators is usually not straightforward due to a tradeoff between bias and variance for different jackknife orders; A sequence of testing hypotheses was suggested to be used to determine a relevant jackknife order for practical uses. Since the chosen jackknife order based on the criterion is no longer fixed, estimation of the variance for the chosen jackknife estimate of species richness is invalid by the standard asymptotic approach. Instead, we propose estimating the variance by using the distribution of the chosen jackknife order associated with an assumption on the testing statistics asymptotically following a multivariate normal distribution.

EO0209: **A generalized estimating equation approach to multivariate adaptive regression splines**
Presenter: **Jakub Stoklosa**, Dr, Australia
Co-authors: David Warton
Multivariate adaptive regression splines (MARS) is a popular nonparametric regression tool often used for prediction and uncovering important data patterns between the response and predictor variables. The standard MARS algorithm assumes normality and independence between continuous response variables. We extend MARS to generalized estimating equations, and we refer to this MARS-for-GEEs algorithm as MARGE. Through simulation we show that the proposed algorithm has improved predictive performance compared with the original MARS algorithm when using correlated and/or non-normal response data and is competitive with alternatives in the literature, especially for problems with multiple interacting predictors. The proposed algorithm is applied to various ecological data types.

EO0275: **Spatial capture-recapture models in continuous-time**
Presenter: **Matthew Schofield**, University of Otago, New Zealand
A key assumption in standard mark-recapture models is that of instantaneous sampling. This may not be appropriate, e.g. when sampling is conducted over several days, weeks or longer. It can be difficult to determine how many discrete sampling occasions to use and they are often determined arbitrarily. If animals are caught multiple times within a sampling occasion they are typically collapsed to an indicator that denotes that it was caught at least once. An alternative is to consider continuous-time models. We explore how spatial location can be incorporated into continuous-time capture-recapture models. We use notions of ancillarity to understand such models, exploring what information is used to inform various aspects of the model. We show how the models we develop can be used for straightforward model fitting.

EO0350: **Estimation of abundance from presence-absence maps using cluster models**
Presenter: **Wen-Han Hwang**, National Chung Hsing University, Taiwan
A presence-absence map consists of an indicator of the occurrence or nonoccurrence of a given species in each cell over a grid without counting the number of individuals in a grid once it is known it is occupied. They are often used to estimate the distribution of a species, but our interest is in using these data to estimate the abundance of the species. In practice, certain types of species (in particular flora types) may be spatially clustered across the presence-absence map, for example, some plant communities will naturally group together according to similar environmental characteristics such as soil type or temperature within a given area. To estimate abundance, several approaches have been proposed but often require prior knowledge of clusters of homogeneous cells, specific survey designs or additional information in the form of covariates. We relax some of these assumptions by developing an approach based on clustered negative binomial models of unknown cluster size. Our approach uses working clusters of cells to construct an estimator which we show is consistent. We also introduce a new concept called super-clustering used to estimate components of the standard errors. A simulation study is conducted to examine the performance of the estimators and they are applied to some real data. Whilst our methods are aimed towards quantitative ecologists and applied statisticians, our work also focuses on the model's theoretical properties.

EO142 Room LSK1010 RECENT ADVANCES IN BAYESIAN COMPUTATION Chair: Minh-Ngoc Tran

EO0365: **Gaussian variational approximation with structured covariance matrices**
Presenter: **David Nott**, National University of Singapore, Singapore
Co-authors: Michael Smith, Siew Li Linda Tan, Victor Ong
Variational approximation methods are an attractive scalable approach to approximate Bayesian inference suitable for large datasets or highly parametrized models. When a Gaussian approximating family is used, one difficulty that arises is that the number of variational parameters to be optimized grows like the square of the number of model parameters. The aim is to discuss how to implement Gaussian variational approximation methods efficiently using stochastic gradient ascent methods in the setting where the number of free parameters in the covariance matrix for the approximation is reduced, by assuming either a sparse precision matrix or a factor covariance structure.

EO0368: **Leverage, asymmetry and heavy tails in high-dimensional factor stochastic volatility models**
Presenter: **Marcel Scharth**, The University of Sydney Business School, Australia
Co-authors: Mengheng Li
It is known that time-varying volatility and leverage effects are often observed in financial time series which is believed to be asymmetrically distributed with heavy tails. The rich literature studying various forms of univariate stochastic volatility model tends to confirm such empirical findings. Yet the literature focusing on high dimensional stochastic volatility models lacks a corresponding general modelling framework and efficient estimation method due to curse of dimensionality. The aim is to propose a flexible factor stochastic volatility model with leverage based on generalized hyperbolic skew Student's t-error to model asymmetry and heavy tails. With shrinkage, the model leads to different parsimonious forms, and thus is able to disengage leverage effects and skewness in idiosyncratic noise from those in factors. A highly efficient Markov chain Monte Carlo estimation procedure which uses efficient importance sampling to exploit the Gaussian mixture representation of the error distribution is proposed to analyze the univariate version of the model. Multivariate extension is achieved with marginalization of factors and boils down to many univariate series which can be estimated in parallel. We assess the performance of our proposed method via a Monte Carlo study with both univariate and multivariate simulated data. Finally we apply our model to a equally weighted portfolio consisting of stocks from S&P/ASX50.

EO0562: **Advances in sequential Monte Carlo methods for static Bayesian models**
Presenter: **Christopher Drovandi**, Queensland University of Technology, Australia
Co-authors: South Leah, Anthony Pettitt, Nial Friel
Sequential Monte Carlo (SMC) methods are a powerful alternative to Markov chain Monte Carlo (MCMC) for sampling from Bayesian posterior distributions of static parameters when the posterior has some irregularities. In particular, SMC facilitates easy adaptation, is embarrassingly

parallelisable and produces an estimate of the model evidence that is useful in Bayesian model comparisons. SMC samples from a sequence of distributions that smoothly connects a distribution that is easy to sample from to the target posterior. A set of weighted samples (particles) are traversed through the sequence of distributions using re-weighting, re-sampling and mutation steps. The mutation step is often based on using MCMC kernels and is a computationally intensive step. We will present some of our recent work on improving SMC methods in terms of obtaining more accurate estimates of the evidence and more efficiently recycling all the information generated in the SMC process. In particular, we will discuss using an independent proposal within the MCMC step, methods for adapting tuning parameters in the MCMC proposal and various variance reduction tricks for obtaining more accurate estimates of the evidence.

EO0569: **The block pseudo-marginal approach**
Presenter: **Robert Kohn**, University of New South Wales, Australia
Co-authors: Minh-Ngoc Tran, Matias Quiroz, Mattias Villani
The pseudo-marginal (PM) approach is increasingly used for Bayesian inference in statistical models where the likelihood is intractable but can be estimated unbiasedly. Examples include random effect models, state-space models and data subsampling in big-data settings. Recently, it was shown how the PM approach can be made much more efficient by correlating the underlying Monte Carlo (MC) random numbers used to form the estimate of the likelihood at the current and proposed values of the unknown parameters. Their approach greatly speeds up the standard PM algorithm, as it requires a much smaller number of particles to form the optimal likelihood estimate. We present an alternative PM approach that divides the underlying random numbers into blocks so that the likelihood estimates for the proposed and current values of the parameters only differ by the random numbers in one block. Our approach has two advantages. First, it provides a direct way to control the correlation between the logarithms of the estimates of the likelihood at the current and proposed values of the parameters. Second, blocking is shown to be a natural way to carry out PM inmany problems, especially when the likelihood is estimated using randomised quasi-Monte Carlo instead of MC.

EO210 Room LSKG001 NEW DEVELOPMENTS IN FUSION LEARNING AND STATISTICAL INFERENCES Chair: Min-ge Xie

EO0417: **Joint estimation of multiple Gaussian graphical models via multiple hypothesis tests**
Presenter: **Faming Liang**, University of Florida, United States
Gaussian Graphical Models have been widely used to explore conditional independence relationships for a large set of random variables. We propose a new method for jointly estimating multiple Gaussian graphical models with observations belonging to distinct classes, which works under the framework of multiple hypothesis testing and includes an meta-analysis procedure to explicitly integrate the data information across distinct classes. The proposed method has some significant advantages over the existing ones. First, it provides a more explicit and sufficient way to integrate the data information across multiple classes. However, the existing methods often integrate the data information through prior distributions or penalty function, and this is often less sufficient. Second, it can provide an uncertainty measure for the edges detected in the multiple graphical models and the difference of edges detected in the graphical models under any two distinct conditions, while the existing methods only produce a point estimate or are feasible for very small size problems. We illustrate the performance of the proposed method using simulated and real data examples. The numerical results indicate the superiority of the proposed method over the existing ones.

EO0427: **Feature subset sampling for prediction in high-dimensional linear models**
Presenter: **Hua Liang**, George Washington University, United States
The high-dimensional prediction problem without the sparsity assumption is considered and a feature subset sampling (FSS) method is introduced, which chooses features by random sampling to get a feature subset, then does the prediction based on the sampled feature subset. This strategy greatly reduces the dimension and saves computational time. We explore this strategy under the ridge regression framework and derive the risk bound as a statistical guarantee of the method. We suggest two sampling strategies: uniform sampling and sampling proportional to the feature information score. Finally, we present detailed empirical studies for illustrating the numerical performance of the method. We demonstrate that the prediction risk of the FSS method can close to that of the full features under the mild assumptions.

EO0486: **Nonparametric fusion learning using confidence distribution, data depth and bootstrap**
Presenter: **Dungang Liu**, University of Cincinnati, United States
Co-authors: Regina Liu, Min-ge Xie
Fusion learning refers to synthesize inferences from diverse sources to provide more effective inference than any individual source. Commonly used methods rely on parametric model assumptions, such as the normality, which may not hold in practice. We propose a general nonparametric framework for fusion learning. The framework enables us to synthesize inferences of a set of target population parameters of interest in a nonparametric manner, i.e., without requiring any distribution specification. The key notion used in our development is the depth confidence distribution (CD), a summary of inferential information for the target parameters. We demonstrate that a depth CD is a useful inferential tool in the sense that it is an omnibus form of confidence regions and p-values, and its contours shrink to the true parameter values. To achieve nonparametric learning, we propose a fusion of depth CDs derived by nonparametric bootstrap in each study. This method can achieve high-order accuracy and Bahadur efficiency by specifying certain combining elements. It can also adapt to complex and irregular settings where the studies are heterogeneous. The advantages of our fusion method are also illustrated in a study of aircraft landings.

EO0834: **Approximate confidence distribution computing: An effective likelihood-free method with statistical guarantees**
Presenter: **Suzanne Thornton**, Rutgers, The State University of New Jersey, United States
Co-authors: Min-ge Xie
Approximate Bayesian computing (ABC) is a likelihood-free method that has grown increasingly popular since early applications in population genetics. However, the theoretical justification for inference based on this method has yet to be fully developed, especially pertaining to the use of non-sufficient summary statistics. We introduce a more general computational technique, approximate confidence distribution computing (ACC) to overcome a few issues associated with the ABC method, for instance, the lack of theory supporting the use of non-sufficient summary statistics, the lack of guardian for the selection of prior, and the long computing time. Specifically, we establish frequentist coverage properties for the outcome of the ACC method by using the theory of confidence distributions, and thus inference based on ACC is justified, even if reliant upon a non-sufficient summary statistic. Furthermore, the ACC method is very broadly applicable; in fact, the ABC algorithm can be viewed as a special case of an ACC method without damaging the integrity of ACC based inference. We supplement the theory with simulation studies and an epidemiological application to illustrate the benefits of the ACC method. It is demonstrated that a well-tended ACC algorithm can greatly increase its computing efficiency over a typical ABC algorithm.

EO186 Room LSK1005 SUFFICIENT DIMENSION REDUCTION IN SURVIVAL ANALYSIS Chair: Ming-Yueh Huang

EO0421: **A novel method of sufficient dimension reduction for censored survival data**
Presenter: **Ming-Yueh Huang**, University of Washington, Taiwan
When there is not enough scientific knowledge to assume a particular regression model, sufficient dimension reduction is a flexible yet parsimonious nonparametric framework to study how covariates are associated with an outcome. A novel estimator is introduced to flexibly summarize the contributions of covariates to a right-censored survival outcome into possibly low-dimensional composite scores. The proposed estimator determines the degree of dimension reduction adaptively from data; it estimates the structural dimension, the central subspace and a rate-optimal smoothing bandwidth parameter simultaneously from a single criterion. It is formulated in a counting process framework and is free of inverse probability weighting which is employed in existing methods but often leads to instability in small samples. The large sample properties for the estimated central subspace with data-adaptive structural dimension and bandwidth are also well-established. Some numerical simulations and real examples are also shown to illustrate the proposed method.

EO0445: **The survival function estimation of current status data with dependent censoring**
Presenter: **Jin-Jian Hsieh**, Department of Mathematics, NATIONAL Chung Cheng UNIVERSITY, Taiwan
The focus is on the survival function estimation of the failure time under the current status data. Because the failure time may be correlated with the observation time in practice, we investigate this topic under the current status data with dependent censoring. We use the Archimedean copula model to specify the dependency between the failure time and the observation time. Under the Archimedean copula assumption, we adopt a redistribution algorithm to estimate the survival function of the failure time. We examine the finite-sample performance of the proposed approach by simulation studies. We also apply our proposed method to analyze a practical tumorigenicity data.

EO0405: **Estimation and model checking in a latent semiparametric model for recurrent event data with informative censoring**
Presenter: **Hung-Chi Ho**, National Taiwan University, Taiwan
Te aim is to investigate the recurrent event process with informative censoring through more general semiparametric latent models. Instead of assuming the distributions of a subject- specific latent variable and censoring times, our approaches mainly rely on the features of recurrent event times and the number of recurrent events. We further note that the distinct distribution behaviors, which are related to the shape parameter, of recurrent event times should be carefully considered in estimation. In light of these findings, two types of estimation methods are thus proposed to deal with shape-dependent and shape-independent intensity regressions. Meanwhile, some test rules are further developed for variant competing intensity regression models without specifying a significance level. Moreover, the large sample properties of the proposed estimators of the index coefficients and numerical measures of the model selection criteria are established under some suitable conditions. In a variety of simulations, the estimators and numerical measures are also found to have satisfactory finite sample performance. As for the applicability of our methodology, it is illustrated by a recurrent event sample of intravenous drug users for inpatient care.

EO0467: **Estimation of the sufficient dimension reduction score with censored survival data**
Presenter: **Shao-Hsuan Wang**, Oncology Department, Johns Hopkins University, United States
Sufficient dimension reduction (SDR) is widely used for the analysis of high-dimensional data of complex structure. The concept of sufficient dimension reduction (SDR) score is first introduced, highlighting the importance and challenges for statistical analysis. To characterize the dependence of a response on covariates of interest, a monotonic structure is linked to a multivariate polynomial transformation of the central subspace (CS) directions with unknown structural degree and dimension. Under a very general semiparametric model formulation, such a sufficient dimension reduction (SDR) score is shown to enjoy the existence, optimality, and uniqueness up to scale and location in the defined concordance probability function. Two types of concordance-based generalized Bayesian information criteria are given to estimate the optimal SDR score and the maximum concordance index, denoted by Cmax. The estimation criteria are further carried out by effective computational procedures. The performance and practicality of the proposals are also investigated through simulations and empirical illustrations.

EO068 Room LSKG007 ADVANCES IN VOLATILITY MODELLING AND FORECASTING Chair: Boris Choy

EO0452: **New analysis method for forecasting foreign exchange rates volatility during the global financial crisis**
Presenter: **Henry Leung**, Sydney University, Australia
Co-authors: Tsung-Han Ke
The accuracy of volatility forecasting in exchange rates plays an important role in financial risk management, particularly for policy-makers, foreign exchange market agents, individual investors, and international traders. Studies have suggested the difficulty in predicting the movements of exchange rates during the global financial crisis. The aim is to improve the prediction ability of forecasting exchange rate volatility by adopting a new efficient method to extract useful information and plug it into the GARCH(1,1) model. Analyses indicate that the new method performs better than traditional methods in forecasting the volatility of exchange rate markets, and improves the volatility forecasting of several GARCH-family models. This study lays the groundwork for the understanding of extracting useful information from big data in the financial market.

EO0510: **Bayesian semi-parametric realized-CARE models for tail risk forecasting incorporating realized measures**
Presenter: **Richard Gerlach**, University of Sydney, Australia
Co-authors: Chao Wang
A new model framework called Realized Conditional Autoregressive Expectile is proposed, incorporating a measurement equation into the conventional CARE model, analogous to the Realized-GARCH model. Competing realized measures are employed as the dependent measurement variable and to drive expectile dynamics. The measurement equation here models the contemporaneous dependence between the realized measure and the latent conditional expectile. We propose employing the quantile loss function as the target criterion, instead of the conventional violation rate, during the expectile level grid search. For the proposed model, the usual asymmetric least squares (ALS) optimization often fails to converge. We incorporate a fast random walk Metropolis stochastic search method, combined with a more targeted grid search procedure, to allow improved accuracy in estimation. Given the convergence issue, Bayesian adaptive Markov Chain Monte Carlo methods are proposed for estimation, whilst their properties are assessed and compared with ALS via a simulation study. In a real forecasting study applied to 7 market indices and 2 asset returns, compared to the original CARE, the parametric GARCH and Realized-GARCH models, one-day-ahead Value-at-Risk and Expected Shortfall forecasting results favour the proposed Realized-CARE model, especially when incorporating the Realized Range and the sub-sampled Realized Range.

EO0465: **Forecasting risk via realized GARCH, incorporating the realized range**
Presenter: **Chao Wang**, The University of Sydney, Australia
Co-authors: Richard Gerlach
The realized GARCH framework is extended to incorporate the realized range, and the intra-day range, as potentially more efficient series of information than realized variance or daily returns, for the purpose of volatility and tail risk forecasting in a financial time series. A Bayesian adaptive Markov Chain Monte Carlo method is developed and employed for estimation and forecasting. Furthermore, the methods of sub-sampling

and scaling are applied to the Realized Range, to help deal with the inherent micro-structure noise of the realized measures. Compared to a range of well known parametric GARCH and realized GARCH models, predictive log-likelihood results across six market index return series clearly favor the realized GARCH models incorporating the realized range and sub-sampled realized range, over a six year period that includes the global financial crisis. Further, these same models, when combined with Student-t errors, also compare favourably for tail risk forecasting, both during and after the crisis period.

EO0444: **Flexible modelling of heavy tails and skewness in multivariate GARCH models**
Presenter: **Charles Au**, University of Sydney, Australia
Co-authors: Boris Choy
The application of the modified multivariate skew-*t* distribution to generalized autoregressive conditional heteroskedastic (GARCH) models in volatility forecasting is proposed. In addition to capturing the correlation structure, the modified skew-*t* distribution can flexibly model the heavy-tailedness and skewness of the returns for different assets. The constant conditional correlation structure will be considered. Bayesian methods will be used for statistical inference, as the modified skew-*t* distribution can be expressed using the scale mixtures of skew-normal (SMSN) representation but not in closed form. An empirical application of multivariate GARCH models with modified skew-*t* innovations to volatility forecasting will be illustrated. The performance of different GARCH models will also be compared.

EO268 Room LSK1027 SPATIAL ECONOMETRICS Chair: Sophie Dabo

EO0454: **Non parametric estimation of space-varying distribution**
Presenter: **Aboubacar Amiri**, Charles de Gaulle University, France
A spatial process with spatially varying distribution is considered, where the restrictions of the distribution function are locally separable meaning that the sampled data do not come from the same distribution of interest but rather from a slowly changing distribution with respect to the space (or the space and time). The evolution of such a process is modeled as follow: the spatial domain of observation is segmented into a set of disjoint regions and it is assumed that inside each of these regions, a number of observations is made, such as cases for some disease or the magnitude of an earthquake on the Richter scale, color of pixels. These observations form a sample from a local distribution. The intention of the exercise is to provide a space-varying estimation of this local distribution.

EO0501: **Generalized partially linear spatial Probit models**
Presenter: **Sophie Dabo**, University-Lille, France
Co-authors: Mohamed Salem Ahmed
Generalized partially linear probit model for spatially dependent data is considered. A triangular array setting is used to cover various patterns of spatial data. Conditional heteroscedasticity, non-identically distributed observations and a linear process for disturbances are assumed allowing various spatial dependencies. The estimation procedure proposed is a combination of a weighted likelihood and a generalized method of moments. We first fix the parametric components of the model and estimate the nonparametric one using weighted likelihood. The obtained estimate is then used to construct a GMM parametric component estimate. Consistency of the parametric and non-parametric components estimators and asymptotic normality of the parameter one are established under sufficient conditions. We present some simulated experiments including real data to investigate the finite sample performance of the estimators.

EO0511: **A space-time-categorical local linear smoother for predicting land/house price**
Presenter: **Ghislain Geniaux**, INRA, France
Co-authors: Davide Martinetti
A new class of data generating processes called MGWR-SAR was recently introduced, in which the regression parameters and the spatial dependence coefficient can vary over the space in order to take into account both spatial heterogeneity and spatial dependence. The estimator corresponds to a local linear smoother with a special kernel based only on Euclidean distance and a linearization of SAR regression using IV/2SLS method. Using simulated data and real price data (land, house), we proved this estimator to be more accurate than estimators integrating exclusively spatial heterogeneity or spatial autocorrelation. However, if any non-linearity exists between covariates, it would require to transform these variables to ensure linear relationship. We propose to improve that estimator using a General Product Kernel adding further smoothed variables to go beyond spatial heterogeneity. Firstly, "time" variable to consider space/time heterogeneity. Secondly, continuous variables to account for their non-linearity, and finally a categorical variable (for example number of rooms, or presence of a garden) to test the relevance of "ad hoc" market segmentation. Using a huge land/house sales spatial database in Provence, we compare prediction accuracy of the BLUP version of several models (OLS,NLS, SAR, SEM, SARAR, MGWR-SAR, REML estimator) against our proposition.

EO0503: **Flexible dependence modeling using convex combinations of different types of weight structures**
Presenter: **Nicolas Debarsy**, CNRS, France
Co-authors: James LeSage
There is a great deal of literature regarding use of non-spatial weight matrices or combinations of spatial and non-spatial weight structures. We explore alternative approaches for constructing convex combinations of different types of dependence between regions. Convex combinations of different weight matrices were previously used to form a single weight matrix that can be used in conventional spatial regression estimation and inference. We explore estimation issues that arise in producing estimates and inferences from these more general cross-sectional regression relationship in a Bayesian setting

EO204 Room LSK1014 FUNCTIONAL DATA ANALYSIS AND ITS APPLICATIONS Chair: Ci-Ren Jiang

EO0460: **Frechet regression for random objects with Euclidean predictors**
Presenter: **Alexander Petersen**, University of California Santa Barbara, United States
Co-authors: Hans-Georg Mueller
Increasingly, statisticians are faced with the task of analyzing complex data that are non-Euclidean and specifically do not lie in a vector space. To address the need for statistical methods for such data, the concept of Frechet regression is introduced, extending the classical concept of a Frechet mean to the notion of a conditional Frechet mean. Generalized versions of both global least squares regression and locally weighted least squares smoothing are developed. The target quantities are appropriately defined population versions of global and local regression for response objects in a metric space. Asymptotic rates of convergence are derived for the corresponding fitted regressions using observed data to the population targets under suitable regularity conditions by applying empirical process methods. For the special case of random objects that reside in a Hilbert space, such as regression models with vector predictors and functional data as responses, a limit distribution is obtained. The proposed methods have broad applicability, including responses that consist of probability distributions and correlation matrices, and the proposed global Frechet regression is demonstrated in these contexts for demographic and brain imaging data. Local Frechet regression theory is also developed and illustrated via a simulation with response data which lie on the sphere.

EO0276: **An iterative approach to fitting varying-coefficient additive models**
Presenter: **Xiaoke Zhang**, University of Delaware, United States
Co-authors: Jane-Ling Wang
The varying-coefficient additive model, which has recently been introduced, is a novel and powerful tool for analyzing functional data and longitudinal data. The model generalizes both the varying-coefficient model and the additive model, and thus inherits the merits of simplicity, flexibility, interpretability and preventability of the curse of dimensionality. The first estimation method for this model was designed and only feasible for data with time-invariant covariates, which is inevitably a limitation. To broaden its applicability, we propose a new approach to fitting the varying-coefficient additive model for data where both longitudinal and time-invariant covariates may be present. A simple iterative algorithm is developed that overcomes the computational difficulty caused by the non-convexity of the objective function. The consistency and the L2 rate of convergence are both established for the estimate of each relevant unknown function. The proposed method is convincingly illustrated by a simulation study and a real data application.

EO0568: **Understanding the time-varying associations between two functional measurements**
Presenter: **Haochang Shou**, University of Pennsylvania, United States
Co-authors: Simon Vandekar, Lihong Cui, Kathleen Merikangas, Vadim Zipunnikov
The availability of multiple domains of measurements in cohort studies poses challenges for statistical analysis. One motivating example is from the NIMH family study of spectrum disorders where daily physical activity profiles are continuously observed for 2 weeks using accelerometers, while in the mean time, ecological momentary assessments (EMA) are also surveyed 4 times a day for the same time period. Researchers are interested in understanding the time-varying associations between the two types of measurements and link such associations with mood disorders to better understand the disease-related deep phenotypes. In another example using Philadelphia Neurodevelopment Cohort (PNC), one is interested in the time-updated relationship of brain perfusion and body growth through child development, and how they are associated with gender or social cognition. We propose methods that estimate the time-dependent associations of the two types of functional data and test whether such associations vary over time.

EO0170: **Sensible functional linear discriminant analysis**
Presenter: **Ci-Ren Jiang**, Academia Sinica, Taiwan
Co-authors: Lu-Hung Chen
The aim is to extend Fisher's linear discriminant analysis (LDA) to both densely recorded functional data and sparsely observed longitudinal data for general c-category classification problems. We propose an efficient approach to identify the optimal LDA projections in addition to managing the noninvertibility issue of the covariance operator emerging from this extension. A conditional expectation technique is employed to tackle the challenge of projecting sparse data to the LDA directions. We study the asymptotic properties of the proposed estimators and show that asymptotically perfect classification can be achieved in certain circumstances. The performance of this new approach is further demonstrated with numerical examples.

EO316 Room LSK1034 FINANCIAL INTEGRATION AND CRISIS TRANSMISSION Chair: Vance Martin

EO0577: **A multivariate integer-valued model of the transmission of financial crisis across global asset markets**
Presenter: **Vance Martin**, University of Melbourne, Australia
A new class of multivariate models is developed to capture the spread of global financial crises across asset markets. Extending upon previous work on univariate integer-valued time series count models of financial crises, multivariate models of counts are specified to capture the dynamic spillover effects connecting global asset markets during periods of financial stress. Both vector autoregressive and moving-average dynamic channels are allowed for. A simulation estimator is proposed based on efficient method of moments as maximum likelihood methods are infeasible in the presence of moving average dynamics. The framework is applied to modelling the causal interlinkages between global asset markets during financial crises with special emphasis on banking, currency and equity asset markets

EO0627: **The count of Monte Carlo: Analysing global banking crises, 1800-2010**
Presenter: **Andrew Tremayne**, University of Liverpool, United Kingdom
Co-authors: Mardi Dungey, Vance Martin, Chrismin Tang
A new class of models is proposed to capture the dynamic transmission mechanisms of banking crises across national borders. As the data are characterized as counts of the number of countries experiencing a banking crisis within a year, a threshold integer autoregressive moving average model (TINARMA) is specified, where the threshold allows for the decomposition into systemic and nonsystemic crisis periods. As maximum likelihood methods are infeasible to estimate the models parameters, a simulation based procedure using efficient method of moments is adopted with standard errors based on subsampling. Applying the modelling framework using annual data from 1800 to 2010, shows that the model is able to capture the dynamic transmission mechanisms of banking crises over time. The forecasting properties of the model are also explored with a focus on periods of financial stress.

EC0634: **Global and regional financial integration in East Asia and the ASEAN**
Presenter: **Renee Fry-McKibbin**, The Australian National University, Austria
Co-authors: Cody Yu-Ling Hsiao, Vance Martin
Financial integration is East Asia and the ASEAN nations is investigated using a new approach based on entropy theory. An important advantage of the proposed framework is that it takes into account changes in integration arising from higher order moments including coskewness and cokurtosis as well as more traditional measures based on second order moments. The analysis focuses on the role of the U.S. in understanding changes in global integration over time, as well as changes in regional integration arising from China. Using daily stock returns from 1997 to 2016, the empirical results show an overall trend to improvements in financial integration over time with deviations from the trend occurring during periods of financial crises in Asia in 1997-98, the U.S. in 2007-10 and the recent European debt crisis. The influence of the Chinese economy mainly through its trading linkages is found to be an important determinant of financial integration over time both regionally as well as globally.

EO0703: **Financial factors and monetary policy: Determinacy and learnability of equilibrium**
Presenter: **Paul Kitney**, Australian National University, Australia
The aim is to contribute to the debate whether central banks should respond to asset prices and other financial factors in setting monetary policy, by evaluating determinacy and expectational stability of equilibria under various monetary policy rules. With adaptive learning, beliefs constitute an additional set of state variables, which may require more than a response to inflation, which traditionally is argued in the literature a sufficient to achieve central bank objectives under rational expectations. Furthermore, financial frictions are introduced by extending the determinacy and adaptive learning methodology embodied in an early work, beyond the New Keynesian modelling framework by incorporating a Financial Accelerator. A key result is that monetary policy rules responding to lagged asset prices and credit volume have less desirable determinacy and learnability characteristics than responding to current asset prices and credit spreads. This conclusion dovetails with recent research which shows that signals derived from credit spreads contain information which help explain business cycle fluctuations and demonstrate that a credit spread augmented

monetary policy rule dampens cycle variability. We also conclude that the mentioned earlier results are also robust to a New Keynesian model with financial frictions.

EO160 Room LSK1032 NEW DEVELOPMENTS IN FINANCIAL TIME SERIES Chair: Guodong Li

EO0339: **Quantile double autoregression**
Presenter: **Qianqian Zhu**, The University of Hong Kong, Hong Kong
Co-authors: Yao Zheng, Guodong Li

A novel conditional heteroscedastic time series model is proposed which is driven by functionally dependent autoregressive coefficients. This new model includes the commonly used double autoregressive model as a special case, where both the location and scale structures are incorporated in double autoregressive forms. However, due to a simple but nontrivial transformation, it enjoys much more flexible parameter structures without positive restrictions on the scale coefficients. The double autoregressive coefficients are monotone functions of a single random variable which is related to quantile levels in view of conditional quantiles of the process, and hence the associated quantile double autoregressive model is able to capture different structures for conditional quantiles of a time series. The strict stationarity of the proposed quantile double autoregressive process is discussed under mild conditions, and a self-weighted conditional quantile estimator is considered for the quantile double autoregressive model. The strong consistency and asymptotic normality of the proposed estimator are obtained with only fractional moments of the process, which makes the model possible to handle heavy-tailed data. Simulation experiments are carried out to access the finite sample performance of the proposed methodology, and an empirical example is presented to illustrate the usefulness of the new model.

EO0342: **Factor models for asset returns based on transformed factors**
Presenter: **Efang Kong**, University of Electronic Science and Technology of China, China

The Fama-French three factor models are built upon the assumption that the return of an asset can be accounted for directly by three factors (i.e. market, size and value factor), through a linear function. A natural question is whether some kind of transformed Fama-French three factors works better than the three factors. If so, what kind of transformation should be imposed on each factor in order to make the transformed three factors better account for asset returns. We address these questions through nonparametric modelling. We propose a data driven approach to construct the transformation for each factor concerned. A generalised maximum likelihood ratio based hypothesis test is also proposed to test whether transformations on the Fama-French three factors are needed for a given data set. Asymptotic properties are established to justify the proposed methods. Intensive simulation studies are conducted to show how the proposed methods work when sample size is finite. Finally, an application of the proposed methods to a real data set produces some interesting findings.

EO0376: **On the quasi-maximum likelihood estimation of threshold GARCH Models**
Presenter: **Yaxing Yang**, Xiamen University, China
Co-authors: Dong Li, Shiqing Ling

The asymptotic theory is studied for the quasi-maximum likelihood estimation (QMLE) for a threshold GARCH model. Under some mild condition, it is shown that the estimated threshold is n-consistent and converges weakly to the smallest minimizer of a two-sided compound Poisson process. The remaining parameters are $\sqrt{n}$-consistent and asymptotically normal. Simulation study is carried out to access the performance of the QMLE in finite sample and a real example is given.

EO0775: **Estimation of semivarying coefficient time series models with ARMA errors**
Presenter: **Lei Huang**, Southwest Jiaotong University, China

Serial correlation in the residuals of time series models can cause bias in both model estimation and prediction. However, models with such serially correlated residuals are difficult to estimate, especially when the regression function is nonlinear. Existing estimation methods require strong assumption for the relation between the residuals and the regressors, which excludes the commonly used autoregressive models in time series analysis. By extending the Whittle likelihood estimation, a semi-parametric autoregressive model with ARMA sequence of residuals is investigated in details. Asymptotic normality of the estimators is established, and a model selection procedure is proposed. Numerical examples are employed to illustrate the performance of the proposed estimation method and the necessity of incorporating the serial correlation in the residuals.

Authors Index