PROGRAMME AND ABSTRACTS

9th International Conference on

Computational and Financial Econometrics (CFE 2015)

http://www.cfenetwork.org/CFE2015

and

8th International Conference of the
ERCIM (European Research Consortium for Informatics and Mathematics) Working Group on

Computational and Methodological Statistics (CMStatistics 2015)

http://www.cmstatistics.org/CMStatistics2015

Senate House & Birkbeck University of London, UK
12 – 14 December 2015

ISBN 978-9963-2227-0-4

Technical Editors: Angela Blanco-Fernandez and Gil Gonzalez-Rodriguez.

Dear Friends and Colleagues,

We welcome you warmly to London, for the Ninth International Conference on *Computational and Financial Econometrics* (CFE 2015) and the Eighth International Conference of the ERCIM Working Group on *Computational and Methodological Statistics* (CMStatistics 2015). As many of you know, this annual conference has become a leading joint international meeting for the interface of statistics, econometrics, empirical finance and computing.

The conference aims at bringing together researchers and practitioners to discuss recent developments in computational methods for economics, finance, and statistics. The CFE-CMStatistics 2015 programme consists of 367 sessions, 5 plenary talks and over 1500 presentations. There are over 1650 participants. This is the biggest meeting so far of the conference series in terms of number of participants and presentations.

The co-chairs have endeavoured to provide a balanced and stimulating programme that will appeal to the diverse interests of the participants. The international organizing committee hopes that the conference venue will provide an appropriate environment to enhance your contacts and to establish new ones.

The conference is a collective effort by many individuals and organizations. The Scientific Programme Committee, the Session Organizers, the local hosting universities and many volunteers have contributed substantially to the organization of the conference. We acknowledge their work and the support of our hosts and sponsors, and particularly Queen Mary University of London and Birkbeck University of London, UK.

The new Elsevier journal, Econometrics and Statistics (EcoSta) is being inaugurated at the conference. The EcoSta is the official journal of the networks of Computational and Financial Econometrics (CFEnetwork) and of Computational and Methodological Statistics (CMStatistics). It publishes research papers in all aspects of econometrics and statistics and it comprises two sections, namely, Part A: Econometrics and Part B: Statistics. The participants are encouraged to submit their papers to special or regular peer-reviewed issues of EcoSta and its supplement Annals of Computational and Financial Econometrics.

Looking forward, the CFE-ERCIM 2016 will be held at the University of Seville, Spain, from Friday 9 to Sunday 11 December 2016. Tutorials will take place on Thursday the 8th of December 2016. You are invited to participate actively in these events.

We wish you a productive, stimulating conference and a memorable stay in London.

Ana Colubi, Erricos J. Kontoghiorghes and Herman K. Van Dijk: coordinators of CMStatistics & CFEnetwork.

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 journals 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 1150 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, 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 700 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

2015-12-12

Opening, 08:25

A - Keynote
CFE
08:40 - 09:30

B
CFE
09:40 - 10:55

Opening, 09:50

C - Keynote
CMStatistics
10:05 - 10:55

Coffee Break
10:55 - 11:25

D
CFE - CMStatistics
11:25 - 13:05

Lunch Break
13:05 - 14:25

E
CFE - CMStatistics
14:25 - 16:05

Coffee Break
16:05 - 16:35

F
CFE - CMStatistics
16:35 - 18:40

Welcome Reception
20:15 - 21:45

2015-12-13

G
CFE - CMStatistics
08:45 - 10:25

Coffee Break
10:25 - 10:55

H
CFE - CMStatistics
10:55 - 13:00

Lunch Break
13:00 - 14:30

I
CFE - CMStatistics
14:30 - 16:10

Coffee Break
16:10 - 16:40

J
CFE - CMStatistics
16:40 - 18:20

K - Keynote
CFE
18:30 - 19:20

Christmas Conference Dinner
20:30 - 23:30

2015-12-14

L
CFE - CMStatistics
08:30 - 10:10

Coffee Break
10:10 - 10:40

M
CFE - CMStatistics
10:40 - 11:55

N - Keynote
CFE - CMStatistics
12:05 - 12:55

Lunch Break
12:55 - 14:30

O
CFE - CMStatistics
14:30 - 15:50

Coffee Break
15:50 - 16:20

P
CFE - CMStatistics
16:20 - 18:00

Q - Keynote
CMStatistics
18:10 - 19:00

Closing, 19:00 - 19:15

TUTORIALS, MEETINGS AND SOCIAL EVENTS

WINTER SCHOOL AND TUTORIALS

The COST Action CRoNoS Winter Course on Robust methods and multivariate extremes takes place on Wednesday 9th to Friday 11th December 2015. On Wednesday and Thursday the course is given at the Senate Room of the Senate House and on Friday is given at the CLO B01 at the Clore Management Building of Birkbeck University of London. The courses on Friday are also designated as tutorials of the conference. The first tutorial is given by Prof. Michael Falk (An Offspring of Multivariate Extreme Value Theory: D-Norms) at 9:00-13:30. The second tutorial is given by Prof. Marc Hallin (Validity-Robust Semiparametrically Efficient Inference for Nonlinear Time Series Models) at 15:00 - 19:30.

SPECIAL MEETINGS by invitation to group members

- The *CSDA Editorial Board* meeting will take place on Friday 11th of December 2015 from 19:00-20:10.
- The *CSDA & Econometrics and Statistics (EcoSta) Editorial Boards* dinner will take place on Friday 11th of December 2015 at 20:20.
- The *Econometrics and Statistics (EcoSta) Editorial Board* meeting will take place on Saturday 12th of December 2015, 18:45-20:15.
- The *COST Action CRONOS* Management Committee lunch-meeting will take place on Saturday 12th of December 2015, 13:00-15:00, Lecture Room MAL B30.

SOCIAL EVENTS

- *The coffee breaks* will take place at the Crush Hall and MacMillan Hall of the Senate House, at rooms MAL B02 and MAL B04 of Birkbeck University of London and at the Foyer of the Clore Management Centre. You must have your conference badge in order to attend the coffee breaks.
- *Welcome Reception, Saturday 12th of December, from 20:15-21:45*. The Welcome Reception is open to all registrants and accompanying persons who have purchased a reception ticket. It will take place at the Senate House (see map at page IX). Conference registrants must bring their conference badge and ticket and any accompanying persons should 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.
- *Conference Dinner, Sunday 13th of December, from 20:30 to 23:45*. The conference dinner is optional and registration is required. It will take place at the Hotel Russell, 1-8 Russell Square (see map at page IX). Conference registrants and accompanying persons should bring their conference dinner tickets in order to attend the conference dinner.
- *Conference Buffet and Sandwich Lunches*. The conference lunches are optional and registration is required. The Buffet Lunch will be arranged at the Hotel Russell (1-8 Russell Square) on 12th, 13th and 14th of December 2015. The Sandwich lunch will be arranged at rooms MAL B02 and MAL B04 of Birkbeck University of London and at the Foyer of the Clore Management Centre. Conference registrants and accompanying persons should bring their conference lunch tickets in order to attend the conference lunches.

GENERAL INFORMATION

Addresses of venues

- Birkbeck and Clore Management Centre, University of London, Malet Street, London WC1E 7HX.
- University of London, Senate House, Malet Street, London WC1E 7HU.

Registration

The registration will be open on Friday afternoon 11th December 2015 at the Foyer of the tutorials venue, Clore Management Centre. The remaining days, that is, 12-14 December 2015, the registration will take place at the McMillan Hall of the Senate House.

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 McMillan and Crush Halls of the Senate House. 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

There will be a wireless Internet connection in the Macmillan Hall. You will need to have your own laptop in order to connect to the Internet. The daily login and password will be displayed on the announcement board by the registration desk of the Macmillan Hall at the Senate House. Participants from any eduroam-enabled institution can use the eduroam service in order to obtain access to Internet at Birkbeck.

Lecture rooms

The paper presentations will take place at Birkbeck, Clore and Senate House (see map in the next page). The list of rooms and their capacity is listed below. Due to health and safety regulations the maximum capacity of the rooms should be respected. There will be no signs indicating the location of the lecture rooms, and therefore we advise that you visit the venue in advance. The opening, keynote and closing talks will take place at the Beveridge Hall of the Senate House. The poster sessions will take place at the Macmillan Hall and Crush Hall of the Senate House.

Room	Capacity	Floor	Location	Room	Capacity	Floor	Location
Gordon	40	Ground	Senate House	Bloomsbury	50	Ground	Senate House
Bedford	50	Ground	Senate House	Beveridge Hall	450	Ground	Senate House
Woburn	50	Ground	Senate House	Torrington	50	First	Senate House
Court	70	First	Senate House	Jessel	50	First	Senate House
Chancellor's Hall	140	First	Senate House	G21A	30	Ground	Senate House
Athlone	30	Ground	Senate House	Holden	30	First	Senate House
Montague	30	Ground	Senate House	SH 349	80	Third	Senate House
Senate	80	First	Senate House	Macmillan Hall	220	Ground	Senate House
MAL B20	99	Basement	Birkbeck Malet St	MAL B33	165	Basement	Birkbeck Malet St
MAL B34	222	Basement	Birkbeck Malet St	MAL B35	125	Basement	Birkbeck Malet St
MAL B36	123	Basement	Birkbeck Malet St	MAL G15	48	Ground	Birkbeck Malet St
MAL B30	40	Basement	Birkbeck Malet St	MAL B29	30	Basement	Birkbeck Malet St
MAL 402	35	Fourth	Birkbeck Malet St	MAL 414	70	Fourth	Birkbeck Malet St
MAL 415	55	Fourth	Birkbeck Malet St	MAL 421	130	Fourth	Birkbeck Malet St
MAL 532	76	Fifth	Birkbeck Malet St	MAL 539	35	Fifth	Birkbeck Malet St
MAL 540	35	Fifth	Birkbeck Malet St	MAL 541	48	Fifth	Birkbeck Malet St
MAL 632	25	Sixth	Birkbeck Malet St	MAL 633	25	Sixth	Birkbeck Malet St
CLO B01	250	Basement	Clore	CLO 101	50	First	Clore
CLO 102	33	First	Clore	CLO 203	33	Second	Clore
CLO 204	33	Second	Clore	CLO 306	33	Third	Clore

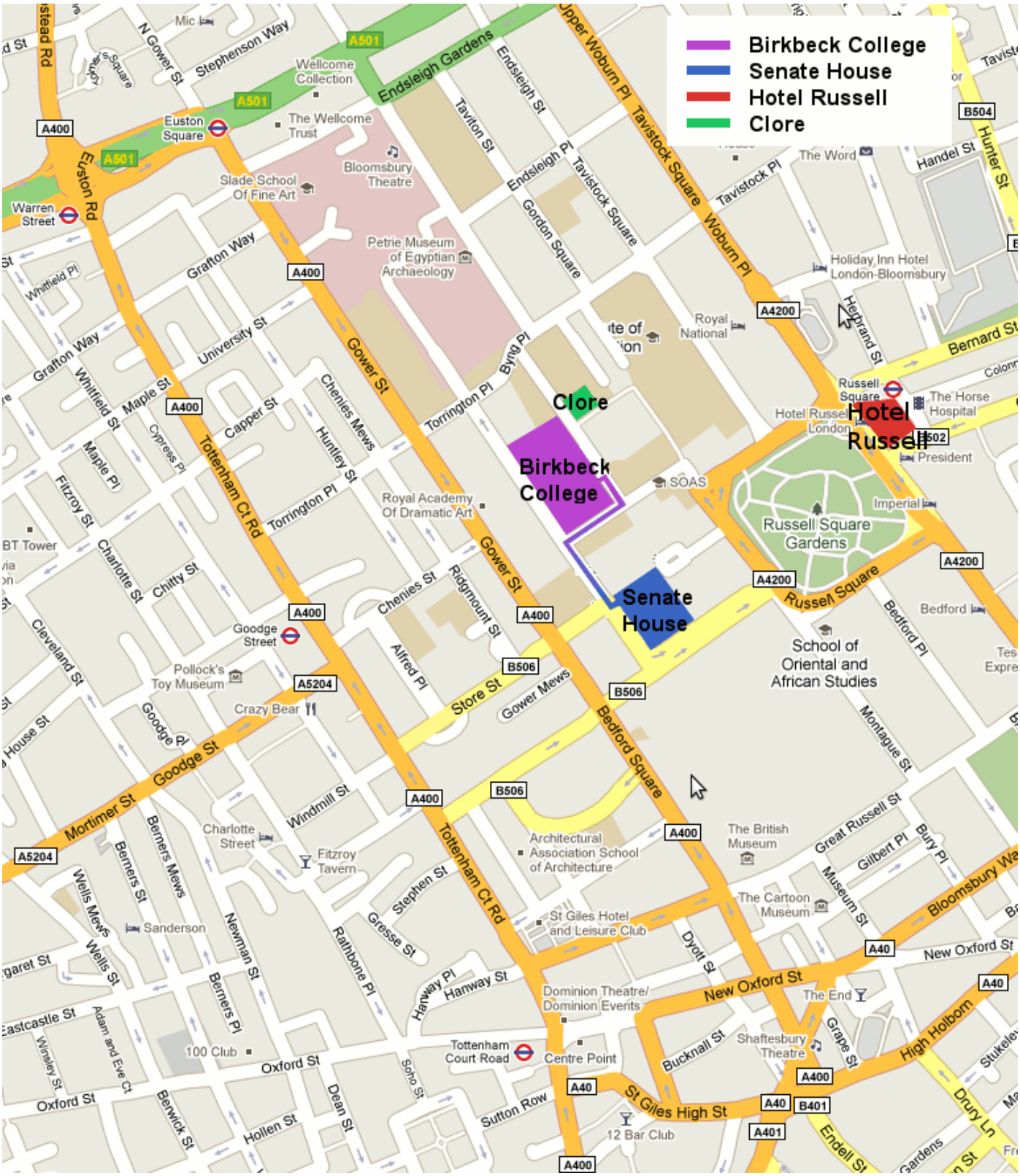
Information and messages

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, Numerical Algorithms Group (NAG), Springer, CRC Press (Taylor & Francis Group) and Atlantis Press.

Map of the venue and nearby area



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 and Computational and Methodological Statistics. 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. Interest is focused as well on well-founded applied econometric studies that demonstrate the practicality of new procedures and models. 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, and, 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, review 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 econometrics or statistics component 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 with deadline for submissions the 28th February 2016:

- (Part A: Econometrics) Annals of Computational and Financial Econometrics.
- Special Issue on Bayesian methods in statistics and econometrics.
- (Part A: Econometrics) Special Issue on Time series econometrics.
- (Part B: Statistics) Special Issue on Mixture models.
- (Part B: Statistics) Special Issue on Functional data analysis.

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 2016 the following special issues with deadline for paper submissions the 28th February 2016:

- 2nd Special Issue on Robust Analysis of Complex Data.
- Special Issue on Design of Experiments.
- Special Issue on Advances in Medical Statistics.

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

Saturday 12.12.2015 08:40 - 09:30 Room: Beveridge Hall Chair: Herman van Dijk Keynote talk 1

Realized volatility based forecasting models: Exploiting the errors

Speaker: **Tim Bollerslev, Duke University, United States** Andrew J Patton, Rogier Quaedvlieg

We propose a new family of easy-to-implement realized volatility based forecasting models. The models exploit the asymptotic theory for high-frequency realized volatility estimation to improve the accuracy of the forecasts. By allowing the parameters of the models to vary explicitly with the (estimated) degree of measurement error, the models exhibit stronger persistence, and in turn generate more responsive forecasts, when the measurement error is relatively low. We document significant improvements in the accuracy of the resulting volatility forecasts compared to the forecasts obtained from some of the most popular existing realized volatility based forecasting models. We also discuss multivariate extensions of the new class of models and various applications thereof, including portfolio allocation decisions and systematic risk measurement.

Saturday 12.12.2015 10:05 - 10:55 Room: Beveridge Hall Chair: Byeong Park Keynote talk 2

Random projection ensemble classification

Speaker: **Richard Samworth, University of Cambridge, United Kingdom**

We introduce a very general method for high-dimensional classification, based on careful combination of the results of applying an arbitrary base classifier to random projections of the feature vectors into a lower-dimensional space. In one special case that we study in detail, the random projections are divided into non-overlapping blocks, and within each block we select the projection yielding the smallest estimate of the test error. Our random projection ensemble classifier then aggregates the results of applying the base classifier on the selected projections, with a data-driven voting threshold to determine the final assignment. Our theoretical results elucidate the effect on performance of increasing the number of projections. Moreover, under a boundary condition implied by the sufficient dimension reduction assumption, we show that the test excess risk of the random projection ensemble classifier can be controlled by terms that do not depend on the original data dimension. The classifier is also compared empirically with several other popular high-dimensional classifiers via an extensive simulation study, which reveals its excellent finite-sample performance.

Sunday 13.12.2015 18:30 - 19:20 Room: Beveridge Hall Chair: Stephen Pollock Keynote talk 3

Estimation and Inference for random time varying coefficient models

Speaker: **George Kapetanios, Queen Mary University of London, United Kingdom**

Estimation and inference for random time varying coefficient models is considered when the random coefficient processes are persistent and bounded. More complex settings than previously studied, including large dimensional datasets and instrumental variable estimation, are investigated. The first setting focuses on the estimation of time varying covariance matrices for large datasets. A variety of popular such estimators are generalised to the time varying case.Theoretical results are presented and data dependent methods for selecting tuning parameters are discussed. The methods are applied to the construction of portfolios using a large number of assets. The second setting focuses on time varying instrumental variable estimation that allows for the endogeneity status of regressors to change over time. Theoretical and simulation results are presented and the methods are applied to a macroeconomic empirical application.

Monday 14.12.2015 12:05 - 12:55 Room: Beveridge Hall Chair: Hans-Georg Mueller Keynote talk 4

Distributed estimation and inference with statistical guarantees

Speaker: **Jianqing Fan, Princeton University, United States** Heather Battey, Han Liu, Junwei Lu, Ziwei Zhu

The focus is on hypothesis testing and parameter estimation in the context of the divide and conquer algorithm. In a unified likelihood based framework, we propose new test statistics and point estimators obtained by aggregating various statistics from k subsamples of size n/k. In both low dimensional and high dimensional settings, we address the important question of how to choose k as n grows large, providing a theoretical upper bound on the number of subsamples that guarantees the errors due to insufficient use of full sample by the divide and conquer algorithms are statistically negligible. In other words, the resulting estimators have the same inferential efficiencies and ℓ_2 estimation rates as a practically infeasible oracle with access to the full sample. For parameter estimation, we show that the error incurred through the divide and conquer estimator is negligible relative to the minimax estimation rate of the full sample procedure. Thorough numerical results are provided to back up the theory.

Monday 14.12.2015 18:10 - 19:00 Room: Beveridge Hall Chair: Yi Li Keynote talk 5

Model-based geostatistics for prevalence mapping in low-resource settings

Speaker: **Peter Diggle, Lancaster University and University of Liverpool, United Kingdom** Emanuele Giorgi

Statistical methods and software associated with the standard model are first reviewed, then several methodological extensions are considered whose development has been motivated by the requirements of specific applications. These include: low-rank approximations for use with large data-sets; methods for combining randomised survey data with data from non-randomised, and therefore potentially biased, surveys; spatio-temporal extensions; spatially structured zero-inflation. Finally, we will also describe disease mapping applications that have arisen through collaboration with a range of African public health programmes.

Saturday 12.12.2015 09:40 - 10:55 Parallel Session B – CFE

CO416 Room Chancellor's Hall ANALYSIS OF HIGH-DIMENSIONAL TIME SERIES I Chair: Marco Lippi

CO0378: **Generalized dynamic factor models and volatilities**
Presenter: **Matteo Barigozzi**, London School of Economics, United Kingdom
Co-authors: Marc Hallin

In large panels of time series with a dynamic factor structure on the levels or returns, the volatilities of the common and idiosyncratic components are often strongly correlated suggesting the presence of common volatility shocks, i.e. they also admit a dynamic factor decomposition. Based on this observation, we propose an entirely non-parametric and model-free two-step general dynamic factor approach which accounts for the factor structure both of returns and of volatilities. We propose a two-step estimation procedure based on one-sided representations of dynamic factor models, and we give some preliminary conditions for consistent estimation of common components of volatilities when both n and T are large. Finally, we apply our method to the panel of SP100 asset returns in order to build GARCH based volatility forecasts. Results show that our approach is superior to existing univariate and multivariate methods when predicting daily highlow range.

CO0399: **Optimal linear prediction of stochastic trends**
Presenter: **Tommaso Proietti**, University of Roma Tor Vergata, Italy
Co-authors: Alessandro Giovannelli

A recent strand of the time series literature has considered the problem of estimating high-dimensional autocovariance matrices, for the purpose of out of sample prediction. For an integrated time series, the Beveridge-Nelson trend is defined as the current value of the series plus the sum of all forecastable future changes. For the optimal linear projection of all future changes into the space spanned by the past of the series, we need to solve a high-dimensional Toeplitz system involving n autocovariances, where n is the sample size. A non parametric estimator of the trend is proposed that relies on banding, or tapering, the autocovariance sequence to achieve consistency. We derive the properties of the estimator and compare it with alternative parametric estimators based on the direct and indirect finite order autoregressive predictors. We then consider the estimation of trends within a multivariate system composed of a target series (e.g. gross domestic product) and a set of observable dynamic factors.

CO1238: **Estimation of generalized linear dynamic factor models: The single and the mixed frequency case**
Presenter: **Manfred Deistler**, Vienna University of Technology, Austria
Co-authors: Alexander Braumann, Elisabeth Felsenstein, Diego Fresoli, Lukas Koelbl

We consider generalized linear dynamic factor models in a stationary framework; the observations are represented as the sum of two uncorrelated component processes: the so-called latent process, which is obtained from a dynamic linear transformation of a low dimensional dynamic factor process and which shows strong dependence of its components, and the noise process, which shows weak dependence of the components. The latent process is assumed to have a singular rational spectral density. In high dimensional time series often the univariate component series are available at different sampling frequencies. This is called mixed frequency observations. We discuss estimation, first for the single frequency case, consisting of the following steps: 1) Denoising, i.e. obtaining estimates of the latent variables and of the static factors from the observations, e.g. via a PCA or a Kalman filtering procedure described; 2) Estimation of a VAR for the static factors, e.g. by AIC and Yule Walker estimators; 3) Estimation of the dimension of the dynamic factors via a PCA on the innovations of the VAR. This procedure is generalized to the mixed frequency case. Several overall estimation procedures, both for the single and for the mixed frequency case, are evaluated and compared by Monte Carlo simulations.

CO623 Room Bloomsbury STATE SPACE MODELS AND COINTEGRATION Chair: Martin Wagner

CO0469: **Polynomial cointegration**
Presenter: **Martin Wagner**, Technical University Dortmund, Germany
Co-authors: Dietmar Bauer

A complete definition of polynomial cointegration is provided, including notions of non-triviality and minimum-degree. The relations of our definition to existing definitions in the literature are clarified in detail. We use the state space framework to describe the polynomial cointegrating spaces for multivariate unit root processes. This is particularly convenient as it allows to transform dynamic polynomial cointegration into a static cointegration problem of a suitably defined process. We furthermore derive expressions that allow to construct tests for the occurrence of certain polynomial cointegrating relationships for a given system. Finally, it is shown that the state space representation allows for an easy understanding of the driving forces underlying polynomial cointegration. This is demonstrated for the $I(2)$ case.

CO0521: **Asymptotic properties of subspace methods for the estimation of seasonally cointegrated models**
Presenter: **Dietmar Bauer**, University Bielefeld, Germany
Co-authors: Rainer Buschmeier

The aim is to investigate the asymptotic properties of the so-called CCA (canonical correlation analysis) algorithm in the case when the data generating process is a minimal state space system containing unit roots at the seasonal frequencies. It is shown that in this situation under mild assumptions on the noise process and under the assumption of known system order, the algorithm provides strongly consistent estimates of the impulse response coefficients without any knowledge on the dynamic properties in terms of (co-)integration of the data generating process. If the number of common cycles for each unit root (i.e. the unit root structure) is known, consistent system matrix estimators can be obtained. Consistency is robust to the extraction of a deterministic trend, constant and deterministic seasonal cycles using seasonal dummies prior to the application of the algorithm. Furthermore, three different procedures for estimating and testing for the number of common trends are proposed and investigated. Two of the suggested procedures mimic tests that could be applied if the true state was known and operate on each unit root separately. The small sample size and power properties are compared for the three tests as well as to standard estimators and tests in the VAR setting.

CO1692: **Some results on the structure theory of cointegrated state space systems**
Presenter: **Massimo Franchi**, University of Rome La Sapienza, Italy

Minimality of the state space representation of a stochastic process places restrictions on the rank of certain matrices that shape its integration and cointegration properties. These restrictions characterize unit root processes in the state space framework. Building on these results, the cointegration and polynomial cointegration properties of the process are described in the $I(d)$ - d greater or equal to 1 - case. The $I(1)$ and $I(2)$ cases and their connection with factor models are discussed in detail.

CO596 Room Montague TEMPORAL AND SPATIAL ECONOMETRIC MODELLING AND TESTING Chair: Zudi Lu

CO0541: **Indirect inference in spatial autoregression**
Presenter: **Maria Kyriacou**, University of Southampton, United Kingdom
Ordinary least squares (OLS) is well known to produce an inconsistent estimator of the spatial parameter in pure spatial autoregression (SAR). The potential of indirect inference to correct the inconsistency of OLS is explored. Under broad conditions, it is shown that indirect inference (II) based on OLS produces consistent and asymptotically normal estimates in pure SAR regression. The II estimator is robust to departures from normal disturbances and is computationally straightforward compared with pseudo Gaussian maximum likelihood (PML). Monte Carlo experiments based on various specifications of the weighting matrix confirm that the indirect inference estimator displays little bias even in very small samples and gives overall performance that is comparable to the Gaussian PML.

CO0555: **Cross-validation bandwidth selection for kernel density estimation with spatial data**
Presenter: **Zudi Lu**, University of Southampton, United Kingdom
Co-authors: Zhenyu Jiang, Nenxiang Ling
Nonparametric kernel method has become increasingly useful in exploring non-Gaussian distribution and nonlinearity with spatial data, in which bandwidth selection is a fundamental component. Cross validation (CV) has been a popular approach for bandwidth selection in nonparametric analysis of independent or time series data. However, although CV has been used in selecting bandwidth for some empirical analyses of real spatial data in the literature, there has been neither theory of asymptotic optimality nor study of its finite sample performance that justifies the CV method for spatial data. In fact, because of multi-direction and lacking of natural ordering in space, it becomes more challenging to establish a theory for the bandwidth selection with spatial data than in time series case. We aim to make a first step to bridge the gap by developing large-sample theory for the CV bandwidth selection in kernel density estimation of spatial stationary lattice random fields. The convergence and asymptotic optimality of the CV selected bandwidth are developed under some mild conditions. The finite-sample performances of these CV properties are also examined by simulation.

CO0746: **Adjusted MLE for the spatial autoregressive parameter**
Presenter: **Federico Martellosio**, University of Surrey, United Kingdom
The quasi-maximum likelihood estimator (QMLE) of the autoregressive parameter in a spatial autoregression can suffer from substantial bias. Part of this bias is due to the presence of other parameters in the model, and a simple approach to reduce the impact of nuisance parameters is to recenter the profile score. We study properties of the resulting estimator, named the adjusted QMLE. Despite being first-order asymptotically equivalent, the QMLE and the adjusted QMLE estimators behave quite differently in many cases of practical interest. In particular, we show that the supports of their distributions may be different, which implies that care needs to be taken when comparing the two estimators.

CO524 Room Athlone MODELING COMMODITY PRICES AND VOLATILITY Chair: Helena Veiga

CO0623: **On the impact of macroeconomic uncertainty on the volatility of commodity prices**
Presenter: **Marc Joets**, Ipag Business School and University Paris Ouest, France
The aim is to analyze the impact of macroeconomic uncertainty on a large sample of 19 commodity markets. We rely on a robust measure of macroeconomic uncertainty based on a wide range of monthly macroeconomic and financial indicators, and we estimate a structural threshold VAR (TVAR) model to assess whether the effect of macroeconomic uncertainty on commodity price returns depends on the degree of uncertainty. Our findings show that whereas the safe-haven role of precious metals is confirmed, agricultural and industrial markets are highly sensitive to the variability and the level of macroeconomic uncertainty, respectively. In addition, we show that the recent 2007-09 recession has generated an unprecedented episode of high uncertainty in numerous commodity markets that is not necessarily accompanied by a subsequent volatility in the corresponding prices, highlighting the relevance of our uncertainty measure in linking uncertainty to predictability rather than to volatility.

CO0723: **Spillover effect of stock market panic on crude oil and natural gas markets**
Presenter: **Julien Chevallier**, IPAG Business School, France
Co-authors: Yue-Jun Zhang
The Volatility Threshold Dynamic Conditional Correlations (VT-DCC) approach is introduced to investigate the spillover effect of stock market panic on crude oil and natural gas markets conditional on volatility regimes. Methodologically, the DCC-MVGARCH model allows the dynamics of correlations to depend on asset variances through a threshold structure. The empirical study of our model to a sample of US stock market panic, represented by the volatility index (VIX), during 1996-2015 indicates that the periods of stock market investors' panic are significantly associated with an increase in cross-market co-movement. The modeling framework represents a useful tool for the study of market contagion.

CO1046: **Pricing the (European) option to switch between two energy sources: An application to crude oil and natural gas**
Presenter: **Hayette Gatfaoui**, IESEG School of Management, France
We consider the viewpoint of a firm, which can choose between crude oil and natural gas to run its business. The trade-off between those two energy sources is straightforward since the firm selects the energy source, which minimizes its energy or production costs at a given time horizon. Assuming the energy strategy to be established over a fixed time window, the energy choice decision will be made at a given future date T. In this light, the firms' energy cost can be considered as a long position in a risk-free bond by an amount of the terminal oil price, and a short position in a European put option to switch from oil to gas by an amount of the terminal oil price too. As a result, the option to switch from crude oil to natural gas allows for establishing a hedging strategy with respect to energy costs. Modelling stochastically the underlying asset of the European put, we propose a valuation formula of the option to switch and calibrate the pricing formula to empirical data on a daily basis. Hence, our innovative framework handles widely the hedge against the price increase of any given energy source versus the price of another competing energy source (i.e. minimizing energy costs). Moreover, we provide a price for the cost-reducing effect of the capability to switch from one energy source to another one (i.e. hedging energy price risk).

CO464 Room Woburn EARLY WARNING SYSTEM AND SYSTEMIC RISK INDICATORS I Chair: Gian Luigi Mazzi

CO0697: **Logit and multinomial logit models for early warning systems: On the duration of systemic banking crises**
Presenter: **Leone Leonida**, Queen Mary University of London, United Kingdom
Co-authors: George Kapetanios, Pietro Calice, Giovanni Caggiano
The performance of the binomial and multinomial logit models in forecasting systemic banking crises is compared in the context of building early warning systems. We argue that, because the average duration of the crises in the sample of countries is longer than one year, the predictive performance of binomial logit models is likely to be hampered by what we define the crisis duration bias. The bias arises from the decision to either treat crisis years after the onset of a crisis as noncrisis years or remove them altogether from the model. A simple Monte Carlo experiment shows that, if compared to the binomial logit model, the multinomial logit approach improves the predictive power of the EWS. We apply the latter to a sample of world economies. Results strongly support the use of the multinomial logit model.

CO0972: **Identifying indicators for stress in the banking system: A simulation based approach**
Presenter: **Makram El-Shagi**, Henan University, China
Co-authors: Jacky Mallett
Due to the experience in the most recent financial crises, that caught politicians and scientists alike by surprise, assessing the fragility of the financial system and detecting stress in the banking system has become a major concern. We propose an agent based simulation tool to identify potential measures of financial stress. Our simulation is based on a double entry book keeping reproduction of bank operations to obtain a highly detailed representation of bank balance sheets which are subject to both economic and regulatory constraints. We use this simulation within a simplified economic framework to recreate the pathology of a banking crisis in a controlled environment. Based on this simulation we identify crisis indicators both on the bank level and the system level. While the main objective is identifying potential measures of crisis risk where the necessary data is not yet being collected, we test the predictive performance of selected indicators that are available to validate our simulations.

CO1139: **Quantifying systemic risk**
Presenter: **Nina Boyarchenko**, Federal Reserve Bank of New York, United States
Co-authors: Domenico Giannone, Tobias Adrian
Using a large panel of macroeconomic, financial and balance sheet explanatory variables, we construct the conditional distribution of future GDP growth, inflation and unemployment from quantile regressions. We document a systemic risk-return trade-off: higher median projections of future GDP growth come at the cost of a larger interquartile range. Furthermore, periods of low realized volatility are associated with large left tails of future GDP growth. Finally, we show that the quantile regression methodology is robust in out-of-sample testing.

CO392 Room Senate GARCH INNOVATIONS Chair: Christian Francq

CO0856: **Testing for policy effects in ARMA-GARCH model**
Presenter: **Wojciech Charemza**, University of Leicester, United Kingdom
Co-authors: Christian Francq, Svetlana Makarova, Jean-Michel Zakoian
We propose tests for detecting outcomes of policy decisions in ARMA-GARCH model. It is assumed that neither timing nor the magnitude of such decisions is known, and the outcomes can be detected by testing the distribution of innovations of the model. It is proved that the Lagrange Multiplier-type tests have well defined asymptotic properties. It is also shown that the power is reasonable for a range of alternatives. Finite sample critical values are obtained by simulation. Empirical application leads to identifying countries with significant fiscal policy effects in series of daily and monthly 10-year government bonds for 37 countries.

CO0878: **Root$-n$ consistent estimation of the density of a GARCH(1,1) process**
Presenter: **Lionel Truquet**, ENSAI, France
It is well known that the density of a function of several independent random variables can be estimated at the usual parametric rate of convergence, using $U-$statistics arguments and kernel density estimation. For time series, such results are available for linear processes. We explain that this approach can be generalized to estimate the marginal density of non linear processes and we give some assumptions under which one can estimate the density of a GARCH(1,1) process with a root$-n$ consistent estimator.

CO1362: **Two step estimation of multivariate GARCH and stochastic correlation models**
Presenter: **Jean-Michel Zakoian**, CREST, France
Co-authors: Christian Francq
The estimation of a wide class of multivariate volatility models is investigated. Instead of estimating an $m-$multivariate volatility model, a much simpler and numerically efficient method consists in estimating m univariate GARCH-type models Equation by Equation (EbE) in the first step, and a correlation matrix in the second step. Strong consistency and asymptotic normality (CAN) of the EbE estimator are established in a general framework, including Dynamic Conditional Correlation models. The EbE estimator can be used to test the restrictions imposed by a particular MGARCH specification. For general Constant Conditional Correlation models, we obtain the CAN of the two-step estimator. Comparisons with the global method, in which the model parameters are estimated in one step, are provided. Monte Carlo experiments and applications to financial series illustrate the interest of the approach.

CO528 Room Holden FUNDS PERFORMANCE MEASUREMENT Chair: Spyridon Vrontos

CO0988: **Diversification benefits of commodities: A stochastic dominance efficiency approach**
Presenter: **Nikolas Topaloglou**, Athens University of Economics and Business, Greece
In light of the growing interest in investments in commodities, we revisit the open question whether commodities should be included in investors' portfolios. We deviate from the previous literature and employ for the first time a stochastic dominance efficiency (SDE) approach. SDE is a non-parametric concept that accommodates deviations from normality and circumvents the necessity to specify a utility function to describe investor's preferences. We find that commodities provide diversification benefits both in- and out-of-sample and regardless of the SDE criterion and performance evaluation measure. This evidence is stronger when commodity indices which mimic dynamic commodity trading strategies are used.

CO1340: **A comprehensive approach to survival analysis of hedge funds**
Presenter: **Spyridon Vrontos**, University of Essex, United Kingdom
Co-authors: Ekaterini Panopoulou
The aim is to estimate the probability of survivorship of individual hedge fund data reported in BarclayHedge database. We employ a plethora of alternative models and a comprehensive data set of indicator variables and hedge fund characteristics. Emphasis is given on the significant factors that affect the lifetime of hedge funds.

CC0982: **Divergence and performance: A new measure of hedge fund distinctiveness**
Presenter: **Nikolaos Voukelatos**, University of Kent, United Kingdom
Co-authors: Ekaterini Panopoulou
We explore whether the level of strategy distinctiveness is associated with future hedge fund performance. We propose a new measure of the distinctiveness of a hedge fund's strategy based on the fund's contribution to the cross-sectional dispersion of returns computed from the universe of all hedge funds. We argue that this Dispersion Contribution Index (DCI) can serve as a meaningful and easily computable measure of hedge fund distinctiveness. The intuition behind DCI is that, the more distinct the investment strategy pursued by a hedge fund, the more likely the fund's returns to diverge from the mean return in the market and, by extension, the larger its contribution to the cross-sectional dispersion of returns in the industry. We investigate the time-series and cross-sectional properties of DCI computed from US hedge fund returns. More importantly, we empirically examine if a higher level of strategy distinctiveness, as measured by DCI, is related to better subsequent performance.

CO496 Room SH349 TIME SERIES Chair: Marco Reale

CO1074: **Generalised linear models for the spectrum of a time series**
Presenter: **Alessandra Luati**, University of Bologna, Italy
Co-authors: Tommaso Proietti
The aim is to introduce the class of generalised linear models with Box-Cox link, which is based on the truncated Fourier series expansion of the Box-Cox transformation of the spectral density. The coefficients of the expansions (which we term the generalised cepstral coefficients) are related to the generalised autocovariances of the series. The link function depends on a power transformation parameter and encompasses the exponential model, which is based on the Fourier series expansion of the logarithm of the spectral density. Other important special cases are the inverse link, which leads to modelling the inverse spectrum), and the identity link. One of the merits of this model class is the possibility of nesting alternative spectral estimation methods (autoregressive, exponential, etc.) under the same likelihood-based framework.

CO1415: **Testing asset price bubbles with financial data**
Presenter: **Yang Zu**, City University London, United Kingdom
Existing literature studying the asset price bubbles does not consider the nonstationary volatility in the data. The bubble testing problem with nonstationary volatility in a continuous-time model is studied. First, the asymptotic distribution of the classical Dickey-Fuller t statistic under the nonstationary volatility is derived. Then a Dickey-Fuller $t-$type statistic motivated by a weighted least squares regression is proposed, where the weight is defined by a preliminary estimated spot error volatility, and its asymptotic distribution is also studied. These statistics are then used to construct tests for asset price bubbles. Monte Carlo simulation and empirical examples show that the bubble test based on the weighted least squares t statistic is reasonably well sized and has superior power property.

CO1780: **Edge deletion tests in graphical models for time series**
Presenter: **Marco Reale**, University of Canterbury, New Zealand
Co-authors: Rory Ellis, Anna Lin, Chris Price
Sparse structural VAR's can be effectively identified with graphical models. A fundamental step in this approach is the estimation of the conditional independence graph. We give a brief overview of the methodology and compare different tests for edge deletion.

CO420 Room Jessel ROBUST METHODS Chair: Stephane Guerrier

CO1242: **The robust lasso for high dimensional regression**
Presenter: **Christophe Croux**, Leuven, Belgium
Co-authors: Andreas Alfons
Sparse model estimation is a topic of high importance due to the increasing availability of data sets with a large number of variables. Another common problem in applied statistics is the presence of outliers in the data. We discuss several possibilities to combine robust regression and sparse model estimation. We focus on the sparse least trimmed squares estimator, a robustified version of the lasso. This estimator has a high breakdown point, and a fast algorithm for its computation is available. An important advantage of this estimator is that no initial sparse robust estimator is needed, neither does the scale of the error terms needs to be known. We discuss in more detail the estimation of the scale of the error terms, needed for constructing outlier detection rules in robust regression.

CO0173: **Semiparametrically efficient rank-based estimation for dynamic location and scale models**
Presenter: **Davide La Vecchia**, University of Geneva, Switzerland
Co-authors: Marc Hallin
New rank-based procedures are introduced to conduct semiparametric inference on time series models. In the considered setting, the conditional location and scale of the process depend on an Euclidean parameter, while the innovation density is an infinite dimensional nuisance parameter. Easy-to-implement rank-based estimators (R-estimators) are derived and their properties are discussed, with emphasis on semiparametric efficiency and root-n consistency even in the presence of misspecification. The developed methodology has a wide range of applications, including linear and nonlinear models, in either discrete- or continuous-time, with either homo- or heteroskedasticity. Numerical examples about the modeling of the two scale realized volatility illustrate the performances of the proposed R-estimators. Finally, some extensions related to constrained inference on conditional duration models for market microstructure analysis and multivariate time series are briefly discussed.

CO0400: **Robust inference for time series models: A wavelet-based framework**
Presenter: **Stephane Guerrier**, University of Illinois at Urbana-Champaign, United States
Co-authors: Roberto Molinari
The aim is to present a new framework for the robust estimation of time series models which is fairly general and, for example, covers models going from ARMA to state-space models. This approach provides estimators which are (i) consistent and asymptotically normally distributed, (ii) applicable to a broad spectrum of time series models, (iii) straightforward to implement and (iv) computationally efficient. The framework is based on the recently developed Generalized Method of Wavelet Moments and a new robust estimator of the wavelet variance. Compared to existing methods, the latter directly estimates the quantity of interest while performing better in finite samples and using milder conditions for its asymptotic properties to hold. Hence, not only we provide an alternative estimator which allows us to perform wavelet variance analysis when data are contaminated, but also a general approach to robustly estimate the parameters of a variety of time series models. The simulations confirm the better performance of the proposed estimators and the applications show the usefulness and broadness of the proposed methodology in domains such as hydrology and engineering with sample sizes up to 500,000.

CO538 Room Gordon MACROECONOMIC ANALYSIS Chair: Peter Zadrozny

CO1435: **An econometric method for decomposing total-input productivity into input-specific productivities**
Presenter: **Peter Zadrozny**, Bureau of Labor Statistics, United States
The aim is to develop and illustrate with aggregate U.S. data an econometric method for decomposing total-input productivity or the Solow residual into input-specific productivities or input-specific Solow residuals, that requires the same data as the Solow residual, in particular, requires no input-specific information. The method is based on an estimated Cobb-Douglas Marginal Production (CDMP) function. Commonly used production functions such as Cobb-Douglas and Constant Elasticity of Substitution (CES) production functions are special cases of the considerably more general CDMP function. Instead of the usual approach of explaining the Solow residual in terms of input-specific information, the proposed method suggests more directly explaining input-specific Solow residuals in terms of input-specific information, which should result in more accurate inferences on the sources of productivity change.

CO1446: **Information and inequality**
Presenter: **Xiaowen Lei**, Simon Fraser University, Canada
Co-authors: Kenneth Kasa
The relationship between information acquisition and capital income inequality is studied. Investors have a Shannon information-processing constraint, which can be relaxed through costly information acquisition. A competitive equilibrium is characterized by solving a Pareto problem with endogenous Pareto weights. Heterogeneous beliefs evolve endogenously as a function of the Pareto weights. Allocations are decentralized by introducing equity and bond markets, and a market for information. The equilibrium features a wealth inequality amplification mechanism previously highlighted in the literature, in which wealthy investors optimally choose to acquire more information. Evidence is provided suggesting that rising capital income inequality can be partially attributed to endogenous information acquisition.

CC1465: **Doubts, inequality, and bubbles**
Presenter: **Kenneth Kasa**, Simon Fraser University, Canada
Co-authors: In-Koo Cho
Two agents share a common benchmark model for dividends. Each is risk-neutral but uncertainty averse, i.e., preferences are linear in consumption, but each agent has doubts about the specification of the dividend process. These doubts manifest themselves has a preference for robustness. Robust preferences introduce pessimistic drift distortions into the benchmark dividend process. These distortions increase with the level of wealth, and give rise to endogenous heterogeneous beliefs. Belief heterogeneity allows asset price bubbles to emerge. A novel implication of our analysis is that bubbles are more likely to occur when wealth inequality increases. A key advantage of our analysis is that detection error probabilities can be used to assess whether empirically plausible doubts about dividends can explain observed bubble episodes.

CO418 Room Court CO-MOVEMENTS IN MACRO AND FINANCE TIME SERIES Chair: Alain Hecq

CO0523: **Index-augmented autoregressive models: Representation, estimation, and forecasting**
Presenter: **Gianluca Cubadda**, University of Rome Tor Vergata, Italy
Co-authors: Elisa Scambelloni
The purpose is to examine the condition under which each individual series that is generated by a $n-$dimensional Vector Auto-Regressive (VAR) model can be represented as an autoregressive model that is augmented with the lags of q linear combinations of all the variables in the system. We call this modelling Index-Augmented Autoregression (IAAR). We show that the parameters of the IAAR can be estimated by a switching algorithm that increases the Gaussian likelihood at each iteration. Provided that the number of factors q times the VAR order p is small compared to the sample size T, we propose a regularized version of our algorithm to handle a medium-large number of time series. We illustrate the usefulness of the IAAR modelling both by empirical applications and simulations.

CO1456: **Testing for news and noise in non-stationary time series subject to multiple revisions**
Presenter: **Michalis Stamatogiannis**, University of Bath, United Kingdom
Co-authors: Alain Hecq, Jan PAM Jacobs
Before being considered definitive, data currently produced by statistical agencies undergo a recurrent revision process resulting in different releases of the same phenomenon. The collection of all these vintages is referred to as a real-time data set. Economists and econometricians have realized the importance of this type of information for economic modeling and forecasting. We focus on testing non-stationary data for forecastability, i.e. whether revisions reduce noise or are news. To deal with historical revisions which affect the whole vintage of time series due to redefinitions, methodological innovations etc. we employ the recently developed Impulse Indicator Saturation approach, which involves adding an indicator dummy for each observation to the model. We illustrate our procedures with the Real Gross National Product series from ALFRED and find that revisions to this series neither reduce noise nor can be considered as noise.

CC1075: **Cross-border effects of fiscal policy in the Eurozone**
Presenter: **Andreea Bicu**, Bank of England, United Kingdom
Co-authors: Lenard Lieb
We empirically assess spillovers from fiscal policy in the Euro area. We propose a structural multi-country factor-augmented vector autoregression model identified with sign restrictions and analyse the domestic and international effects of fiscal policy measures. By extracting information from an extended set of country specific and cross-border variables, we are able to account for the different channels through which government expenditure shocks are transmitted within as well as across borders. We find significant negative effects of fiscal consolidations on domestic output, private consumption and investment. More importantly, spending cuts in Italy and Spain induce significant and persistent output spillovers on Germany and France.

CO424 Room Torrington TECHNICAL ANALYSIS AND ADAPTIVE MARKETS Chair: Robert Hudson

CO0526: **On technical trading rules**
Presenter: **Andrew Urquhart**, University of Southampton, United Kingdom
Co-authors: Robert Hudson, Bartosz Gebka
A previous study found that technical trading rules have strong predictive power. Since then, there has been an explosion of studies examining the potential profitability of technical trading rules. However the majority of these studies examine the performance of technical trading rules over some pre-determined period and through a number standard parameters of technical trading rules. The aim is to investigate whether investors could have exploited various technical trading rules in the DJIA and FT30 over a long horizon. Employing a parameter sweep, we report the most successful technical trading rules over various subperiods and show how the most successful set of parameters of technical trading rules change over time. This provides strong evidence of the Adaptive Market Hypothesis and that investors need to be weary of the parameters they employ when using technical trading rules.

CO1348: **On the determinants of stability in financial markets**
Presenter: **Mona Soufian**, Hull University Business School, United Kingdom
With an intention to investigate the determinants of stability in financial market, we attempt to build further on the work of fresh thinking on systematic risk and use Agent Based Modelling to examine systematic risk and stability in financial markets. We consider financial market as a complex system and examine complexity and stability, which are evidentially the common ground between financial systems and network science. We examine whether the stability of financial markets can be influenced by the systems complexity, agents population and the connectedness between agents. The results would be beneficial to build and further develop fresh thinking on systematic risk in financial markets.

CO1810: **Investigating the profitability of technical analysis using cross-sectional country data**
Presenter: **Robert Hudson**, University of Hull, United Kingdom

We initially present results showing the profitability of technical analysis for a comprehensive list of countries. We then test whether, as in the prior literature on US stocks, rule returns are positively related to risk as measured by standard deviation. To further understand our findings we consider the distinction between the economic returns generated by rules and the ability of rules to predict market direction. We do this both algebraically and by empirical investigation of the data. We subsequently consider the effect of time varying standard deviation and investigate other potential causes of predictability relating to different national characteristics.

CG577 Room Bedford CONTRIBUTIONS ON NONSTATIONARY TIME SERIES AND PANELS Chair: Michael Vogt

CC0260: **The long-run dynamics of the real sovereign bonds: What it can be learnt from Net Foreign Assets**
Presenter: **Melika Ben Salem**, University Paris-Est Marne-la Vallee, France
Co-authors: Barbara Castelletti Font

Many theoretical and empirical studies have been devoted to explain the long-run dynamics of real interest rates. Economic theory indicates GDP, the rate of return on investment, time preference and risk whereas the relationship with fiscal policy variables remains an open issue. Sovereign bond yields have been declining since the 1980s, which casts doubt to their reaction to the deterioration of budget deficits in OECD countries. It seems then useful to disentangle short-run and long run determinants. If short-run determinants allow us to proxy the shifts in expectations about the fundamentals, open economies with high degree of financial integration should share common transmission mechanisms for their real interest rates, at least in the long run. On the contrary expectations could differ in the short-run. These two features explain the choice of an heterogeneous panel error-correction model. Departing from the existing studies the net international investment position, regarded as an indicator of default risk is introduced. The idea that net foreign assets play a role on macroeconomic performance and stability is at the heart. Our analysis yields a key finding: net foreign assets combined nonlinearly with government debt do explain the long-run dynamics of sovereign bond yields, even with short-term interest rates.

CC1565: **New results on the power of some tests for a fixed unit root under a stochastic unit root alternative**
Presenter: **Julio Angel Afonso-Rodriguez**, University of la Laguna, Spain

The nonstationary behaviour of some macroeconomic and financial time series seems to be not well characterized by the standard unit root process. Within the class of nonlinear unit root processes, we concentrate on the so-called stochastic (or randomized) unit root (STUR) process which introduces an additional persistent component that could explain some of the conflicting empirical results arising when testing for the null of a fixed unit root against a stationary alternative, and nests two different nonstationary alternatives to the standard fixed unit root process such as the pure STUR and the bilinear unit root (BLUR) processes. By introducing general assumptions about serially and mutually correlated errors and a proper normalization of the parameter governing the effect of the additional component, we study the power properties of a great variety of parametric and semiparametric tests for the null of a fixed unit root against the alternatives of stationarity and nonlinear nonstationarity, as well as for some tests for the null of stationarity. We obtain some new results that could explain the apparent conflicts arising when combine the outcomes of some of these complementary testing procedures, and propose an alternative testing procedure to discriminate between a fixed unit root against any of these two sources of nonlinear nonstationarity, incorporating a semiparametric correction for serially correlated errors with good size and power properties.

CC1588: **Estimating structural parameters in regression models with adaptive learning**
Presenter: **Michael Massmann**, WHU - Otto Beisheim School of Management, Germany
Co-authors: Norbert Christopeit

The ordinary least squares (OLS) estimator of the structural parameters is examined in a class of stylised macroeconomic models in which agents are boundedly rational and use an adaptive learning rule to form expectations of the endogenous variable. The popularity of this type of model has recently increased amongst applied economists and policy makers who seek to estimate it empirically. Two prominent learning algorithms are considered, namely constant gain and decreasing gain learning. For each of the two learning rules, the analysis proceeds in two stages. First, we derive the asymptotic properties of agents' expectations. At the second stage, we derive the asymptotics of OLS in the structural model, taken the first stage learning dynamics as given. In the case of constant gain learning, the structural model effectively amounts to a static, cointegrating or co-explosiveness regression. With decreasing gain learning, the regressors are asymptotically collinear such that OLS does not satisfy, in general, the Grenander conditions for consistent estimability. Nevertheless, we show that the OLS estimator remains consistent in all models considered. We also show, however, that its asymptotic distribution, and hence any inference based upon it, may be non-standard.

Saturday 12.12.2015 11:25 - 13:05 Parallel Session D – CFE-CMStatistics

CI016 Room Beveridge Hall SPECIAL SESSION ON BAYESIAN METHODS IN ECONOMICS AND FINANCE Chair: Knut Are Aastveit

CI0215: **Bayesian estimation of sparse dynamic factor models with order-independent and ex-post mode identification**
Presenter: **Sylvia Kaufmann**, Study Center Gerzensee, Switzerland
Co-authors: Christian Schumacher
Common variation in N series of a large panel data set is extracted into few k factors, where $k \ll N$. The factors are dynamic and follow a vector autoregressive process. We estimate the factor loadings under a sparse prior which includes positive mass at zero. Non-zero factor loadings in columns provide an explicit interpretation of the factors. Rows of zero factor loadings indicate irrelevant variables which do notcontain information for factor estimation. We propose a way of estimating and identifying the factor model which is independent of variable ordering. The posterior mode, i.e. factor sign and factor order, is identified ex-post by post-processing the draws from the posterior distribution.

CI0485: **Mutual fund dynamic risk allocation and skill level with a Bayesian seminonparametric change-point model**
Presenter: **Mark Jensen**, Federal Reserve Bank of Atlanta, United States
Co-authors: Mark Fisher, Paula Tkac
The aim is to model the skill and risk-factor allocation of a mutual fund dynamically as a change-point process of the four factor Fama-French, momentum model of mutual fund returns. Each mutual funds change point probability is assumed to be a random draw from an unknown distribution that is modelled nonparametrically. In addition, each fund regimes latent risk factor beta and alpha coefficients are also assumed to be random draws from unknown meta-distributions. We nonparametrically model each meta-distribution as an unknown distribution and place a hierarchical prior on it. A nonparametric heirarchical prior for the regime coefficients enable our change-point model to learn from past fund regimes risk factor betas and performance alphas to better predict their out-of-sample values. We find dynamic factor allocation and changing levels of skill for mutual funds. An arbitrary fund will have a relatively flat duration distribution for a regime. In other words, a fund is more or less likely to experience a change point.

CC1610: **Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator**
Presenter: **Michael Pitt**, University of Warwick, United Kingdom
Co-authors: Arnaud Doucet, George Deligiannidis, Robert Kohn
When an unbiased estimator of the likelihood is used within an Metropolis-Hastings scheme, it is necessary to trade off the number of samples used to evaluate the likelihood against the computing time. Many samples will result in a scheme which has similar properties to the case where the likelihood is exactly known but will be expensive. Few samples will result in faster estimation but at the expense of slower mixing of the Markov chain. We explore the relationship between the number of samples and the efficiency of the resulting Metropolis-Hastings estimates. Under the assumption that the distribution of the additive noise introduced by the log-likelihood estimator is independent of the point at which this log-likelihood is evaluated and other relatively mild assumptions, we provide guidelines on the number of samples to select for a general Metropolis-Hastings proposal. We illustrate on a complex stochastic volatility model that these assumptions are approximately satisfied experimentally and that the theoretical insights with regards to inefficiency and computational time hold true.

CO590 Room Senate EMPIRICAL MODEL DISCOVERY Chair: David Hendry

CO0159: **Improving global vector autoregressions**
Presenter: **Neil Ericsson**, Federal Reserve Board, United States
The recent financial crisis and Great Recession highlight the pervasiveness of international macro-financial linkages in world economy. Capturing those linkages in empirical economic models is central to much modeling, economic inference, forecasting, and policy analysis. Global vector autoregressions (GVARs) provide a framework for capturing those linkages. GVARs have several attractive features: a standardized economically appealing choice of variables for each country or region examined, a systematic treatment of long-run properties through cointegration analysis, flexible dynamic specification through vector error correction modeling, and a natural structure for capturing macro-financial relationships. The theoretical and empirical underpinnings for GVARs are re-examined, focusing on exogeneity assumptions, parameter constancy, and data aggregation. Refinements are proposed in these areas by utilizing recent developments in automated model selection and empirical model discovery, with the aim of achieving an even more robust approach to GVAR modeling. The substantive 26-country GVAR highlights these refinements.

CO0350: **Bias correction after selection with correlated variables**
Presenter: **Jennifer Castle**, Oxford University, United Kingdom
Co-authors: David Hendry, Jurgen Doornik
The aim is to develop bias correction for the conditional distributions of the estimated parameters of retained variables after model selection, such that approximately unbiased estimates of their coefficients are delivered. Building on previous work which demonstrates the approach for orthogonal regressors, we extend the analysis to correlated regressors. The selected model can be transformed to an orthogonal representation in which an operational formula can be applied. Consequently, the combined outcome of our results is that despite searching in a large model class, across different (unknown) states of nature, the finally selected model delivers nearly unbiased estimates and essentially unbiased standard errors for retained variables, with few adventitiously-significant effects – a performance close to commencing from the DGP. The theory is demonstrated by simulation evidence in a range of states of nature.

CO0775: **Automatic selection of multivariate dynamic econometric models**
Presenter: **Jurgen Doornik**, University of Oxford, United Kingdom
Co-authors: David Hendry
Automatic general-to-specific selection of univariate dynamic econometric models is now well established. This has been extended in several directions, including adding impulse dummies for every observation (IIS) and selection when there are more variables than observations. We consider extension of these procedures to the multivariate setting. The starting point is the unrestricted reduced form, captured by the vector-autoregression, possibly with additional unmodelled variables. At this first stage a variable either enters all equations or none. Cointegration properties can also be considered at this stage. The final stage can be the simultaneous equations model, where each variable is treated individually in each equation from the perspective of the entire model. We discuss the role of identification. Some applications illustrate how the procedure works in practice.

CO0954: **Detecting structural changes in linear models: A variable selection approach using multiplicative indicator saturation**
Presenter: **Morten Nyboe Tabor**, University of Copenhagen, Denmark
Co-authors: Oleg Kitov
We model structural breaks in regression coefficients of multivariate linear models using multiplicative indicator saturation (MIS) in Autometrics.

Baseline performance is assessed using Monte Carlo experiments. First, we show that there is almost no efficiency loss under the null of no breaks, irrespective of the number of variables in the system. For a bivariate case with multiple breaks of varying size and timing we find that MIS can detect the break points correctly even for zero-mean iid variables. Performance is improved for the non-zero mean autoregressive case and larger break sizes. Furthermore, using a recursive procedure, we assess the number of postbreak observations required for successful detection. Finally, we conduct empirical tests for structural change in inflation persistence and the Phillips curve for the UK. Our results confirm the presence of three breaks for both cases: 1965, 1971 and 1974. These are consistent with the post Bretton Woods argument and support the Lucas critique. Performance of MIS in simulations and empirical cases is compared with the BaiPerron procedure.

CO568 Room Holden MODELLING RISK Chair: Giovanni Barone-Adesi

CO0166: **Estimating the joint tail risk under filtered historical simulation**
Presenter: **Giovanni Barone-Adesi**, University of Lugano, Switzerland
Co-authors: Kostas Giannopoulos
The aim is to extend the use of filtered historical simulation in estimating the potential losses a CCP (central counterparty) would face from a multiple default in the EMIR framework. The proposed methodology provides a probabilistic estimation of defaulting of named members, the expected size of losses, i.e. the joint expected shortfall (JES), and confidence intervals around the JES. This in turn provides an estimate of financial resources needed to absorb multiple defaults. Our methodology is carrying a full re-pricing of all instruments in the portfolio. It takes into account positions that expire before the profits and losses (P&L) horizon.

CO0284: **A diagnostic criterion for approximate factor structure**
Presenter: **Patrick Gagliardini**, University of Lugano, Switzerland
A simple diagnostic criterion for approximate factor structure in large cross-sectional equity datasets is built. Given a model for asset returns with observable factors, the criterion checks whether the error terms are weakly cross-sectionally correlated or share at least one unobservable common factor. It only requires computing the largest eigenvalue of the empirical cross-sectional covariance matrix of the residuals of a large unbalanced panel. A general version of this criterion allows us to determine the number of omitted common factors. The panel data model accommodates both time-invariant and time-varying factor structures. The theory applies to generic random coefficient panel models under large cross-section and time-series dimensions. The empirical analysis runs on monthly returns for about ten thousand US stocks from January 1968 to December 2011 for several time-varying specifications. Among several multi-factor time-invariant models proposed in the literature, we cannot select a model with zero factors in the errors. On the opposite, we conclude for no omitted factor structure in the errors for several time-varying specifications.

CO0943: **A jump and smile ride: Continuous and jump variance risk premia in option pricing**
Presenter: **Fulvio Corsi**, Ca Foscari University Venice and City University London, Italy
Stochastic and time-varying volatility models typically fail to correctly price out-of-the-money (OTM) put options at short maturity. We extend Realized Volatility option pricing models by adding a jump component estimated from high-frequency data and the associated risk premium. The inclusion of jumps provides a rapidly moving volatility factor, which improves on the fitting properties under the physical measure. The change of measure is performed adopting a Stochastic Discount Factor (SDF) with three risk premia: equity, and two variance risk premia related to the continuous and jump components. On the one hand, employing an SDF with multiple risk premia further improve the flexibility under risk neutral dynamics while preserving analytical tractability. On the other hand, it provides new way of separately estimate the continuous and jump variance risk premia by coherently combining high-frequency returns and option data in a multi-factor option pricing model. The empirical analysis illustrates the contribution of the jump factor to the pricing performance of Standard and Poor's 500 index OTM options. Finally, we apply our multi-factor model to estimate the time evolution of the two components of the variance risk premium and test their ability to predict future market returns.

CO1011: **Bayesian semiparametric multivariate change point analysis**
Presenter: **Stefano Peluso**, Catholic University of Milan, Italy
Co-authors: Siddhartha Chib, Antonietta Mira
We develop a Bayesian semiparametric multiple multivariate change point model. Each regime parameter or subsets thereof can follow a dedicated change point process. Latent regime-specific Dirichlet process mixtures allow for a random distribution of recurrent regime parameters. A time-dependent transition matrix among the latent regimes is introduced through change points driven by a multilevel Reinforced Urn Process. The properties of the proposed model and of a previous one are studied theoretically through the analysis of the distribution of the interarrival times and the number of change points in a given time interval. A multivariate generalization of a previous algorithm permits the statistical analysis of multiple multivariate change points. The posterior sampling algorithm is applied to simulated and real data.

CO436 Room Jessel TIME-SERIES ECONOMETRICS Chair: Robert Kunst

CO0172: **Forecasting city arrivals with Google Analytics: The merits of big data shrinkage techniques and forecast combination**
Presenter: **Ulrich Gunter**, MODUL University Vienna, Austria
Co-authors: Irem Onder
The aim is to investigate the predictive ability of 10 Google Analytics web traffic indicators from the website of the DMO of Vienna for actual tourist arrivals. Tests show that Google Analytics indicators jointly Granger-cause arrivals to Vienna and vice versa. Hence, VAR becomes an appropriate model class. As the sample is small and to prevent over-parameterization, methods for big data shrinkage are applied to create rival forecast models to the classical VAR of dimension 11: Bayesian estimation of the VAR of dimension 11 (BVAR), reduction to a factor-augmented VAR of dimension 3 (FAVAR), and an application of Bayesian estimation to the FAVAR of dimension 3, the Bayesian FAVAR (BFAVAR). The forecast accuracy of these models is evaluated in terms of standard measures (RMSE and MAE) for horizons $h = 1,2,3,6$, and 12 relative to univariate benchmarks (MA(2), ETS, and naive-1). Results based on the single models show that FAVAR and BFAVAR generally outperform their multivariate competitors, but that overall MA(2) and ETS perform best except for one case. The classical VAR of order 11 and the naive-1 benchmark are significantly outperformed for $h = 1,2$, and 3. Four methods of forecast combination are evaluated: uniform combination, Bates-Granger weights, forecast encompassing tests, and a combination of the last two. For $h = 3,6$, and 12, the combined forecasts outperform the forecasts of the single models in four cases, while only in two cases uniform combination can be beaten.

CO0244: **Seasonal time deformation and periodic autoregressive time series models**
Presenter: **Philip Hans Franses**, Erasmus School of Economics, Netherlands
Periodic autoregressive time series models are models in which the autoregressive parameters vary across the seasons. There are various sources of periodicity, and these include economic behaviour and interpolation of data. We advance another potential source of periodicity which is improper aggregation. With simple examples we show that inappropriate aggregation of a monthly non-periodic process can lead to quarterly periodicity. We exploit this link between aggregation and periodicity by reversing the issue and trying to turn a periodic process into a non-periodic process by

specific temporal aggregation. We put forward a general theoretical result on the link between time deformation and periodicity, and we discuss various empirical examples.

CO0595: **Profit persistence and stock returns**
Presenter: **Michael Hauser**, Vienna University of Economics and Business, Austria
Co-authors: Adelina Gschwandtner
Further empirical evidence on the relationship between the product and the financial market is assembled. Drawing back on work in industrial organization, we analyze the relationship between profit persistence and factor adjusted stock returns looking at about 2000 listed US firms over the last 34 years. While the relationship between (current, lagged and unexpected) profits/earnings and returns has been extensively analyzed before, to our knowledge this is the first study to look at the relationship between stock returns and profit persistence. We interpret profit persistence as a result of market competition and innovation of the firm. It is shown that firm specific long-run profit persistence after correction for other additional economic fundamentals of the firm, has a positive impact on 4-factor adjusted returns and a negative impact on their volatility. Technically, we estimate the persistence characteristics of the firms in a first step by moving window regressions, and introduce them as variables with measurement errors in the return and volatility equations.

CO0749: **On the persistence of economic time series: A threshold approach**
Presenter: **Heiko Rachinger**, University of Vienna, Austria
Co-authors: Vanessa Berenguer-Rico
A threshold model is proposed for changes in persistence and volatility of economic time series. There is a large literature showing that the persistence and/or volatility of economic time series are not constant. In order to answer what changes the degree of persistence as well as the volatility of economic time series, we propose a Lagrange Multiplier (LM) test for theshold effects in the persistence and volatility of a time series. A sequential testing procedure allows us to solve a potential confounding problem about the sources of the changes. We derive the asymptotic distribution of the test statistic for a known and unknown threshold parameter. Further, we discuss estimation of the threshold and other parameters of the threshold model. Via Monte Carlo simulations, we analyze the finite-sample behaviour of the tests. Finally, we apply the methodology to several US macroeconomic series.

CO578 Room SH349 MEASUREMENT OF MARKET SPILLOVERS Chair: Matthias Fengler

CO0198: **Measuring spot variance spillovers when (co)variances are time-varying: The case of multivariate GARCH models**
Presenter: **Helmut Herwartz**, Georg-August-University Goettingen, Germany
Co-authors: Matthias Fengler
In highly integrated markets, news spreads at a fast pace and bedevils risk monitoring and optimal asset allocation. We therefore propose global and disaggregated measures of variance transmission that allow one to assess spillovers locally in time. Key to our approach is the vector ARMA representation of the second-order dynamics of the popular BEKK model. In an empirical application to a four-dimensional system of US asset classes – equity, fixed income, foreign exchange and commodities – we illustrate the second-order transmissions at various levels of (dis)aggregation. Moreover, we demonstrate that the proposed spillover indices are informative on the value-at-risk violations of portfolios composed of the considered asset classes.

CO0199: **Global equity market volatility spillovers: A broader role for the United States**
Presenter: **Daniel Buncic**, University of St Gallen, Switzerland
Co-authors: Katja Gisler
It has recently been shown that U.S. equity market returns carry valuable information to improve return forecasts in a large cross-section of international equity markets. We extend previous work on that and examine if U.S. based equity market information can be used to improve realized volatility forecasts in international equity markets. For that purpose, we obtain volatility data for the U.S. and 17 international equity markets from the Oxford Man Institute's realized library and augment for each foreign equity market the benchmark HAR model with lagged U.S. equity market volatility information. In-sample as well as out-of-sample evaluation results suggest a strong role for U.S. based volatility information. More specifically, apart from standard in-sample tests, which find U.S. volatility information to be highly significant, we show that this information can be used to substantially improve out-of-sample forecasts of realized volatility. Using large out-of-sample evaluation periods containing at least 2500 observations, we find that forecast improvements, as measured by the out-of-sample $R2$ (relative to a model that does not include U.S. based volatility information), can be as high as 12.83, 10.43 and 9.41 percent for the All Ordinaries, the Euro STOXX 50 and the CAC 40 at the one-step-ahead horizon and yield Clark-West adjusted t statistics of over 7.

CO0651: **Measuring the frequency dynamics of financial and macroeconomic connectedness**
Presenter: **Tomas Krehlik**, Charles University in Prague, Czech Republic
Co-authors: Jozef Barunik
A general framework is proposed for measuring frequency dynamics of connectedness in economic variables based on spectral representation of variance decompositions. We argue that the frequency dynamics is insightful when studying the connectedness of variables as shocks with heterogeneous frequency responses will create frequency dependent connections of different strength that remain hidden when time domain measures are used. Two applications support the usefulness of the discussion, guide a user to apply the methods in different situations, and contribute to the literature with important findings about sources of connectedness. Giving up the assumption of global stationarity of stock market data and approximating the dynamics locally, we document rich time-frequency dynamics of connectedness in US market risk in the first application. Controlling for common shocks due to common stochastic trends which dominate the connections, we identify connections of global economy at business cycle frequencies of 18 up to 96 months in the second application. In addition, we study the effects of cross-sectional dependence on the connectedness of variables.

CO1304: **Estimating global sovereign default risk connectedness**
Presenter: **Kamil Yilmaz**, Koc University, Turkey
We apply the Diebold-Yilmaz connectedness index methodology on sovereign credit default swaps (SCDSs) to estimate the network structure of global sovereign credit default risk. In particular, using the elastic net estimation method, we separately estimate networks of daily SCDS returns and volatilities for 38 countries between 2009 and 2014. Our results reveal striking differences between the network structures of returns and volatilities. In SCDS return connectedness networks, developing and developed countries stand apart in two big clusters. In the case of the SCDS volatility connectedness networks, however, we observe regional clusters among emerging market countries along with the developed-country cluster. Furthermore, we show that global factors are more important than domestic factors in the determination of SCDS returns and volatilities. In addition, we show that emerging market countries are the key generators of connectedness of sovereign default risk shocks while severely problematic countries as well as developed countries play relatively smaller roles.

CO506 Room Athlone MODELLING VOLATILITY Chair: Christos Savva

CO0235: **Risk-return trade-off for European stock markets**
Presenter: **Christos Savva**, Cyprus University of Technology, Cyprus
Factor models with macro-finance predictors are adopted to test the intertemporal risk-return relation for 13 European stock markets from 1986 to 2012. We use country specific, euro area, and US macro-finance factors to determine the conditional volatility and conditional return. We find that the risk-return trade-off is generally negative. The Markov switching model documents that there is time-variation in this trade-off that is linked to the state of the economy, but not the business cycles. Quantile regressions show that the risk-return trade-off is stronger at the lowest quantile of the conditional return.

CO1088: **Inflation volatility and the Euro: On the effect of the common currency**
Presenter: **Nektarios Michail**, Cyprus University of Technology, Cyprus
We employ a sample of Euro area countries within a Generalised Autoregressive Conditional Heteroskedasticity (GARCH) framework to examine whether the results of the recent literature about the decrease in inflation volatility after the introduction of the common currency hold. In addition we examine whether the impact is consistent or heterogeneous across countries and whether it remains the same in magnitude after taking into account recent data incorporating the European sovereign debt crisis.

CO0873: **The role of market indices in forecasting stocks volatility: A HAR framework using a mixed sampling approach**
Presenter: **Vasileios Pappas**, University of Bath, United Kingdom
Co-authors: Marwan Izzeldin
The aim is to examine the value added in forecasting high frequency stock data using a Heterogeneous Autoregressive (HAR) model augmented with market indicators (VIX, SPY and the SP500). Our empirics are based on high frequency data of 100 stocks representing 10 different sectors for the period 2000 to 2010. We allow for different sampling frequencies in both sides of the HAR regression specification as well as allowing for different market regimes. We find that the Augmented HAR (A-HAR) specification brings significant gains over the conventional HAR. We also outline the index and sampling frequency at which the maximum gains are attained (i.e. 900 seconds for the SP500 and 15 seconds for the stocks). The gains from A-HAR specification are more realised during the crisis episode where index information appears to be of more importance.

CO1025: **Volatility forecasting around the world**
Presenter: **Lazaros Symeonids**, University of East Anglia, United Kingdom
Co-authors: Apostolos Kourtis, Raphael Markellos
We investigate the performance of several popular volatility forecasting models across different countries, forecast horizons and economic conditions. To this end, we carry out an extensive comparison of the information content and predictive ability of implied, realized and GARCH volatility forecasts across 13 international equity indices. We find that model rankings remain roughly the same across countries, while they vary with the forecast horizon. The best model is the Heterogeneous Autoregressive (HAR) model in the daily horizon while implied volatility forecasts that account for the volatility risk premium are superior in the monthly horizon. GARCH forecasts are inferior to implied and realized volatility forecasts in almost all cases considered. Focusing on the recent financial crisis reveals that the forecasting ability of all models deteriorates in periods of market turmoil. We finally assess the economic value of the forecasting models within an international diversification framework. In this setting, we find that implied volatility forecasts can improve portfolio performance over historical methods.

CO566 Room Chancellor's Hall NONCAUSAL AND NON GAUSSIAN TIME SERIES MODELS Chair: Alain Hecq

CO0272: **Nonlinear dynamic interrelationships between real activity and stock returns**
Presenter: **Henri Nyberg**, University of Helsinki, Finland
Co-authors: Markku Lanne
The aim is to explore the differences between a linear Gaussian structural VAR (SVAR) and noncausal vector autoregressive VAR models in capturing a potentially nonlinear real activity-stock return-relationship. Unlike the conventional and commonly used linear SVAR model, the noncausal VAR model is capable of accommodating various nonlinear and non-Gaussian dependencies characteristic of stock returns, and, therefore, it is likely to produce real activity and stock market shocks different from those implied by a SVAR model. In quarterly U.S. data (1953-2012), we find evidence in favor of noncausality suggesting that the stock return is an insufficient proxy for news affecting real activity and stock prices. In addition, especially when conditioning on the state of the business cycle, GDP growth is found more important for the determination of stock returns than in a linear SVAR model.

CO0495: **Forecasting inflation in Europe with mixed causal-noncausal models**
Presenter: **Sean Telg**, Maastricht University, Netherlands
Co-authors: Alain Hecq, Lenard Lieb
In the literature, it has been claimed that inflation is a forward-looking variable rather than a backward-looking one. Recently, it has been argued that the inclusion of noncausal terms in the model is an adequate way to capture this type of behavior. For this reason, we forecast inflation in Europe with mixed causal-noncausal autoregressive (AR) models, which explicitly take dependence on the future into account. We compare the forecasting performance of various models, including the purely causal, mixed and purely noncausal specifications, to assess whether noncausality indeed improves the forecasting accuracy.

CO0557: **Modelling the demand of photovoltaic panels using mixed-causal autoregression**
Presenter: **Lenard Lieb**, Maastricht University, Netherlands
Co-authors: Alain Hecq
Renewable energies have been heavily subsidized over the last decade. In particular, subsidies for solar panels to enhance (private) demand have had a strong impact on the market, and triggered irregularities and phenomena such as increasing volatility and/or bubbles. We propose to use mixed-causal autoregressions to model the demand of solar panels in Belgium over the last decade. It has been shown previously that mixed-causal and noncausal models can parsimoniously capture some of those (nonlinear) features. We review the mostly theoretical literature on mixed-causal regressions and provide the applied researcher with a guideline for practical implementation for estimation and inference. We assess the finite sample properties of existing estimation strategies and suggest a robust alternative based on self-weighted least absolute deviation, allowing for the existence of infinite variance processes. A MATLAB toolkit is provided allowing the applied researcher to straightforwardly specify, estimate, and test mixed-causal autoregressive models.

CO1549: **Revisiting identification and estimation in structural VARMA models**
Presenter: **Alain Monfort**, ENSAE Paris, France
Co-authors: Christian Gourieroux
The basic assumption of a structural VARMA model (SVARMA) is that it is driven by a white noise whose components are uncorrelated (or

independent) and can be interpreted as economic shocks, called "structural" shocks. These models have to face two kinds of identification problems. The first identification problem is "static" and is due to the fact that there is an infinite number of linear transformations of a given random vector making its components uncorrelated. The second identification problem is "dynamic" and is a consequence of the fact that the SVARMA process may have a non invertible AR and/or MA matrix polynomial but, still, has the same second order properties as a VARMA process in which both the AR and MA matrix polynomials are invertible (the fundamental representation). Moreover the standard Box-Jenkins approach automatically estimates the fundamental representation and, therefore, may lead to misspecified Impulse Response Functions. The aim is to explain that these difficulties are mainly due to the Gaussian assumption underlying the Box-Jenkins type approaches, and that both identification challenges are solved in a non Gaussian framework. We also develop simple new parametric and semi-parametric estimation methods when there is nonfundamentalness in either the moving-average, or the autoregressive dynamics, and discuss the derivation of impulse response functions.

CO490 Room Bedford THE ECONOMETRICS OF CLIMATE CHANGE Chair: Tommaso Proietti

CO0408: **Econometric models of climate systems: The equivalence of two-component energy balance models and cointegrated VARs**
Presenter: **Felix Pretis**, University of Oxford, United Kingdom
Climate policy target variables including emissions and concentrations of greenhouse gases, as well as global mean temperatures are non-stationary time series invalidating the use of standard statistical inference procedures. Econometric cointegration analysis can be used to overcome some of these inferential difficulties, however, cointegration has been criticised in climate research for lacking a physical justification for its use. It will be shown that a physical two-component energy balance model of global mean climate is equivalent to a cointegrated system that can be mapped to a cointegrated vector autoregression, making it directly testable, and providing a physical justification for econometric methods in climate research. Doing so opens the door to investigating the empirical impacts of shifts from both natural and human sources, and enables a close linking of data-based macroeconomic models with climate systems. This approach finds statistical support of the model using global mean surface temperatures, 0-700m ocean heat content and radiative forcing (e.g. from greenhouse gases). The model results show that previous empirical estimates of the temperature response to the doubling of CO2 may be misleadingly low due to model mis-specification.

CO0648: **Modelling the interactions in paleoclimate data**
Presenter: **James Davidson**, University of Exeter, United Kingdom
Econometric methods are applied to model paleoclimate series for temperature, ice volume and and atmospheric concentrations of CO2 and CH4. These series, measured from Antarctic ice and ocean cores, are well-known to move together in the transitions between glacial and interglacial periods, but the dynamic relationship between the series is open to dispute. A further unresolved question is the precise role of Milankovitch theory, in which the glacial/interglacial cycles are correlated with orbital variations. We perform tests for Granger non-causality and also conduct a sensitivity analysis of the impulse-response characteristics of our model. Previous work with climate series has assumed nonstationarity and adopted a cointegration approach, but in a range of tests we find no evidence of integrated behaviour. We use conventional autoregressive methodology while allowing for conditional heteroscedasticity in the residuals, associated with the transitional periods.

CO0685: **Global mean temperatures and global CO2 concentrations: A seasonal state-space approach**
Presenter: **Eric Hillebrand**, Aarhus University and CREATES, Denmark
Co-authors: Tommaso Proietti
The aim is to study the statistical relation of monthly CO_2 concentrations measured at Mauna Loa, Hawaii, with monthly global mean temperatures published by the Goddard Institute for Space Studies (GISS). The central challenge is the trending nature of both series, which invalidates inference in a standard linear model. Both series exhibit strong seasonal patterns that need to be accounted for. We consider a state-space model that allows for seasonally varying deterministic and stochastic trends in temperatures and CO_2 concentrations, as well as seasonally varying autocorrelation and residual variances. The model can be summarized as containing a permanent and a transitory component for each series, where global warming is captured in the permanent component on which the seasons load differentially. The permanent component of CO_2 concentrations influences the permanent component of temperature. We conduct inference on the coupling coefficient and study different forecast scenarios.

CO1424: **Nonparametric estimation and bootstrap inference on the recent trends in atmospheric ethane (C2H6) above Europe**
Presenter: **Marina Friedrich**, Maastricht University, Netherlands
Co-authors: Whitney Bader, Bruno Franco, Bernard Lejeune, Emmanuel Mahieu, Hanno Reuvers, Stephan Smeekes, Jean-Pierre Urbain
Ethane is the most abundant non-methane hydrocarbon in the Earth's atmosphere and an important precursor of tropospheric ozone. Its monitoring is therefore crucial for the characterization of air quality and of the transport of tropospheric pollution. Ethane is also an indirect greenhouse gas, influencing the atmospheric lifetime of methane. The main sources of ethane are located in the northern hemisphere, and the dominating emissions are associated to production and transport of natural gas. A preliminary trend analysis was conducted using measurements performed in the Swiss Alps. Over the last two decades, the trend of ethane showed a decline of around 1% per year, thanks to a reduction of fugitive emissions of fossil fuel sources. However, a recent upturn potentially attributed to the massive exploitation of shale gas and tight oil reservoirs in North America was found. The goal is to investigate the presence and form of changes in trend functions using nonparametric techniques. The possible location of such changes is investigated. In addition, nonparametric estimation techniques are used to allow for nonlinear trend functions. Given the nonstandard nature of the measurements we rely on dependent wild bootstrap techniques to conduct inference on possible breaks in linear trends and on nonparametric trend functions.

CO410 Room Bloomsbury FINANCIAL REGULATION Chair: Dominique Guegan

CO0444: **The spectral stress VaR**
Presenter: **Kehan Li**, University Paris I Pantheon Sorbonne, France
Co-authors: Dominique Guegan
One of the key lessons of the crisis which began in 2007 has been the need to strengthen the risk coverage of the capital framework. In response, the Basel Committee in July 2009 completed a number of critical reforms to the Basel II framework which will raise capital requirements for the trading book and complex securitisation exposures, a major source of losses for many international active banks. One of the reforms is to introduce a stressed value-at-risk (VaR) capital requirement based on a continuous 12-month period of significant financial stress (Basel III). However the Basel framework does not specify a model to calculate the stressed VaR and leaves it up to the banks to develop an appropriate internal model to capture material risks they face. Consequently we propose a forward stress risk measure spectral stress VaR (SSVaR) as an implementation model of stressed VaR, by exploiting the asymptotic normality property of the distribution of estimator of VaRp. In particular to allow SSVaR incorporating the tail structure information we perform the spectral analysis to build it. Using a data set composed of operational risk factors we fit a panel of distributions to construct the SSVaR in order to stress it. Additionally we show how the SSVaR can be an indicator regarding the inner model robustness for the bank.

CC1222: **Probability density of future returns: A fully nonparametric heteroskedastic model**
Presenter: **Matthieu Garcin**, LabEx REFI - Natixis AM, France
Co-authors: Clement Goulet
A new model is proposed for estimating returns and volatility. The approach is based both on wavelet denoising technique and variational theory. It is assessed that the volatility can be expressed as a non-parametric functional form of past returns. Therefore, both returns and volatility can be forecasted. Confidence intervals for predicted returns as well as accurate probability densities for future returns can be built. This new technique outperforms classical time series theory. This model does not require the stationarity of the observed log-returns, it preserves the volatility stylised facts and it is based on a fully non-parametric form. This non-parametric form is obtained thanks to the multiplicative noise theory. An application on intraday and daily financial data is proposed.

CO1375: **Risk or regulatory capital: Bringing distributions back in the foreground**
Presenter: **Bertrand Hassani**, Pantheon Sorbonne, France
Co-authors: Dominique Guegan
We discuss the regulatory requirement (Basel Committee, ECB-SSM and EBA) to measure financial institutions' major risks, for instance Market, Credit and Operational, regarding the choice of the risk measures, the choice of the distributions used to model them and the level of confidence. We highlight and illustrate the paradoxes and the issues observed implementing an approach over another and the inconsistencies between the methodologies suggested and the goal to achieve. Some recommendations to the supervisor and proposed alternative procedures to measure the risks are made.

CO1489: **Empirical probability density function of Lyapunov exponents**
Presenter: **Clement Goulet**, Paris Sorbonne, France
We introduce a simple method to approximate the empirical distribution of Lyapunov exponent for one-dimensional discrete dynamical systems. We recall that its positivity provides a necessary condition for being chaotic. Hence, if a dynamical system has a positive Lyapunov exponent and if it has an attractor with fractal dimension, then forecasts can be done inside this attractor. Nevertheless, the estimation of Lyapunov exponent on observed dynamical systems often produce results close to zero and so it is hard to assess whether the dynamical system is chaotic or not. The approximation of the Lyapunov exponent distribution and the computation of confidence intervals overcome this limitation. The distribution approximation is done through Maximum Entropy bootstrapping technique. This technique does not require neither stationarity of the system nor any law assumption and preserve path dependency. To our knowledge this is the first time that such technique is used to approximate the empirical distribution of the Lyapunov exponent. We propose an application of our method on a denoised phase space generated by financial data.

CO542 Room Torrington FINANCE AND JOBS IN DYNAMIC MACRO MODELS Chair: Ekkehard Ernst

CO0480: **Countercyclical capital rules for small open economies**
Presenter: **Rossana Merola**, ILO International Labour Office, Switzerland
Co-authors: Daragh Clancy
Macro-financial feedback loops played a key role in both triggering and propagating the recent financial crisis. The Great Recession proved that macroeconomic policies were insufficient to ensure financial stability. This has been particularly true for small open economies within a monetary union, as they are constrained in the use of traditional stabilisation tools, such as nominal interest and exchange rates. We develop a DSGE model with a banking sector tailored for a small open economy in a monetary union. In our model, loans defaults are related to the value of collateralized assets (i.e. housing) and to wage income. We find that positive expectations for future house prices play a role in the accumulation of credit risk. The negative effects of this over extension of credit materialise when the bubble busts and these expectations prove to be overoptimistic. The resulting slowdown in economic activity, lower disposable incomes and a greater risk of unemployment, combined with devalued collateral, further exacerbate the increase in non-performing loans bank losses. In terms of policy advice, our simulations suggest that a proactive macro prudential rule responding to credit growth can help in smoothing economic fluctuations and promoting financial and macroeconomic stability. We also find that more aggressive action during the release phase can bolster the economy's ability to absorb a negative financial shock and damper its effects on the labour market.

CO0578: **Business cycle asymmetries and the labor market**
Presenter: **Christian Merkl**, Friedrich-Alexander-Unversitaet Erlangen-Nuernberg, Germany
Two business cycle facts for the United States are shown. First, the job-finding is very asymmetric over the business cycle and moves a lot more in recessions than in booms. Second, there is a positive correlation between the job-finding rate and the backed out matching efficiency. We provide an explanation for both business cycle facts by enhancing a search and matching model with idiosyncratic shocks for match formation and by solving the full nonlinear structure of the model. Understanding the sources of these nonlinearities is very important. In our calibration, government interventions (such as wage subsidies and government spending) are three times more effective in a heavy recession than in a boom. In addition, shifts of the matching efficiency are not necessarily a sign for higher structural unemployment.

CO0938: **Hiring uncertainty, investment and job creation: The role of financial frictions**
Presenter: **Ekkehard Ernst**, International Labour Organization, Switzerland
The aim is to analyze the role of financial market frictions on the effects of uncertainty shocks on growth and employment. A newly developed hiring uncertainty indicator is used to measure the difference between hiring intentions and actual job creation. The effect of this hiring uncertainty on macro and labour market dynamics is analysed both theoretically and empirically. The theoretical part considers a stochastic growth model with different shocks on the returns to capital and employment and analyses the role of financial market frictions in transmitting these shocks to capital accumulation and job creation. The empirical part analyses the importance of historical hiring uncertainty shocks for G7 countries by means of a panel SVAR approach. In particular, it is shown that in times of distressed financial markets when frictions are large, hiring uncertainty can exercise a large, negative effect on both growth and employment with long-lasting effects. A quantitative measure of these effects is provided for the average G7 country and estimates the role of hiring uncertainty during the recent crisis.

CO0941: **About monetary policy transmission in remittances-receiving countries**
Presenter: **Anne Oeking**, International Monetary Fund, United States
Co-authors: Ralph Chami, Ekkehard Ernst
We show that countries with higher remittance-to-GDP ratios exhibit weaker monetary policy transmission. Often, these countries are low-income or emerging market economies with fragile institutional and financial structures. However, the role played by remittances on monetary policy transmission in these countries has received little attention. Using panel data estimation, we analyze the impact of remittance flows on the bank lending channel. First, we derive some stylized facts regarding the banking sector and bank balance sheets in remittance-receiving countries. Second, we motivate the empirical exercise with a simple model. Finally, we test whether the monetary policy transmission channel from policy rates to bank lending rates is different in countries with higher remittances inflows. We show that banks in remittance-receiving countries have a stable and interest-insensitive source of funding through remittances with more long-term and stable bank deposits. This stable funding, however,

does not translate into higher private sector lending, as banks in these countries tend to hold more liquid and risk-free assets such as government securities. The excess liquidity generated by the remittance flows tends to make the interbank funding market less relevant, and remittances weaken the monetary transmission via the bank lending channel.

CO570 Room Court SPARSE MODELLING, SHRINKAGE AND REGULARIZATION Chair: Helga Wagner

CO0603: **Probabilistic low-rank matrix completion with adaptive spectral regularization algorithms**
Presenter: **Francois Caron**, University of Oxford, United Kingdom
A novel class of algorithms for low rank matrix completion is proposed. The approach builds on novel penalty functions on the singular values of the low rank matrix. By exploiting a mixture model representation of this penalty, we show that a suitably chosen set of latent variables enables to derive an Expectation-Maximization algorithm to obtain a Maximum A Posteriori estimate of the completed low rank matrix. The resulting algorithm is an iterative soft-thresholded algorithm which iteratively adapts the shrinkage coefficients associated to the singular values. The algorithm is simple to implement and can scale to large matrices. We provide numerical comparisons between our approach and recent alternatives showing the interest of the proposed approach for low rank matrix completion.

CO1175: **Achieving shrinkage in the time-varying parameter models framework**
Presenter: **Angela Bitto**, WU Wien, Austria
Co-authors: Sylvia Fruehwirth-Schnatter
We investigate shrinkage for time-varying parameter models based on the normal-gamma prior which has already been introduced for standard regression models. Our approach extends previous work in which the Bayesian Lasso prior has been considered. The Bayesian Lasso is a special case of the normal-gamma prior. We show how the normal-gamma prior can easily be extended to the time-varying parameter models and focus on inducing shrinkage on the square root of the variance of the prior of the error term in the non-centered state equation. We present both a univariate and a multivariate application. First we choose EU area inflation modelling based on the generalized Phillips curve, then we draw our attention to a multivariate time series with a time-varying covariance matrix and analyse DAX-30 data. Our findings suggest, that the normal-gamma prior bears advantages over the Bayesian Lasso prior in terms of statistical efficiency and performs significantly better when drawing attention to the predictive performance.

CO0777: **Sparse Bayesian modelling for categorical predictors**
Presenter: **Daniela Pauger**, Johannes Kepler University Linz, Austria
Co-authors: Helga Wagner, Gertraud Malsiner-Walli
The usual strategy to include a categorical covariate in a regression type model is to define one of the levels as baseline and to introduce dummy variables for all other levels. As this can result in a high-dimensional vector of regression effects, methods which allow sparser representation of the effect of categorical covariates are required. We achieve a sparse representation of the effect of a nominal predictor by defining informative prior distributions. The specification of a spike and slab prior on level effect differences allows classification of these differences as (practically) zero or non-zero. Thus, we can decide whether (1) a categorical predictor has no effect at all, (2) some (all) level effects are non-zero and/or (3) some (all) categories can be fused as they have essentially the same effect on the response. Additionally we consider a modification of the standard spike-and slab prior where the spike at zero is combined with a slab distribution which is a location mixture distribution. Model-based clustering of the effects during MCMC allows to detect levels which have essentially the same effect size. We demonstrate the performance of the developed methods in simulation studies and for real data.

CO1164: **The spike-and-slab LASSO**
Presenter: **Veronika Rockova**, University of Pennsylvania, United States
Co-authors: Edward George
Despite the wide adoption of spike-and-slab methodology for Bayesian variable selection, its potential for penalized likelihood estimation has largely been overlooked. We bridge this gap by cross-fertilizing these two paradigms with the Spike-and-Slab LASSO procedure for variable selection and parameter estimation in linear regression. We introduce a new class of self-adaptive penalty functions that arise from a fully Bayes spike-and-slab formulation, ultimately moving beyond the separable penalty framework. A virtue of these non-separable penalties is their ability to borrow strength across coordinates, adapt to en- semble sparsity information and exert multiplicity adjustment. The Spike-and-Slab LASSO procedure harvests efficient Bayesian EM and coordinate-wise implementations with a path- following scheme for dynamic posterior exploration. We show on simulated data that the fully Bayes penalty mimics oracle performance, providing a viable alternative to cross-validation. We develop theory for the separable and non-separable variants of the penalty, showing rate- optimality of the global mode as well as optimal posterior concentration when $p > n$. Thus, the modal estimates can be supplemented with meaningful uncertainty assessments.

CO625 Room Montague CREDIT RISK MODELLING Chair: Jonathan Crook

CO0666: **Reliable region predictions for automated valuation models**
Presenter: **Tony Bellotti**, Imperial College London, United Kingdom
Accurate property valuation is important for assessing credit risk in the mortgage market, allowing lenders to determine loan-to-value and estimate loss-given-default. Traditional property valuation using a surveyor is expensive and may not be accurate or entirely objective. Therefore, automated valuation models (AVM) are being developed to provide cheaper, objective valuations that can also allow dynamic updating of property values over the term of a mortgage. A useful feature of automated valuations is that they should give a range (or region) of plausible price estimates for each individual property, rather than a single point estimate. This would allow lenders to include conservatism in their credit risk assessments. Conformal Predictors (CP) are machine learning algorithms developed to provide just such region predictions. They have been shown to be reliable, in the sense that predictive accuracy can be precisely controlled using a pre-defined confidence level. We show how an AVM can be constructed with a CP, based on an underlying k-nearest neighbours approach. The AVM is tested on a well-known Boston House Price data set and a new London house price data set. We show that the region predictions are reliable and also investigate how the width of region predictions, ie the predictive uncertainty, is linked to property characteristics.

CC1301: **Predictability of emerging market yield spreads before and after Lehman Brothers: The role of macroeconomic volatility**
Presenter: **Ana-Maria Fuertes**, Cass Business School - City University London, United Kingdom
Co-authors: Alena Audzeyeva
We study the sources of emerging market credit spreads predictability using weekly data for four large sovereign Eurobond markets over two periods surrounding the Lehman Brothers' default. Information from the credit spread curve only is insufficient for the baseline model to outperform the random walk and slope regression benchmarks. Extensions with global and country macroeconomic factors outperform both benchmarks, particularly post-Lehman which indicates a closer alignment then with fundamentals. The analysis reveals a key predictive role for macroeconomic uncertainty measures such as the volatility of the global riskless short-interest rate and also the volatility of the country's trade balance.

CC1506: **Benchmarking state of the art classification algorithms for credit scoring**
Presenter: **Bart Baesens**, KU Leuven, Belgium
Co-authors: Stefan Lessmann, Hsinn-Vonn Seow, Lyn Thomas, Bart Baesens
More than 10 years have passed since a well-known benchmarking study of classification techniques for credit scoring has been published. One of its key findings was that simple techniques such as logistic regression and decision trees usually perform well when compared to more complex techniques such as neural networks and support vector machines. We will provide an updated perspective on these findings by considering the current state of the art in the field. More specifically, we will include recently suggest classification techniques, new performance metrics and innovative statistical evaluation frameworks. We will compare 41 classifiers in terms of 6 performance measures across 8 real- life data sets. The results might provide a new benchmark to catalyze further research in the field.

CO1813: **A new model for estimating exposure at default**
Presenter: **Jonathan Crook**, University of Edinburgh, United Kingdom
Co-authors: Mindy Leow
Conventional methods for estimating exposure at default use cross sectional models to predict a ratio involving credit balance and then manipulate this to predict balance at the time of default. But for revolving credit it is often the case that as an account approaches default the balance approaches and may even exceed the limit. We use panel data to exploit this relationship to derive a new method of predicting balance at default. We use panel data to model the monthly time path of balance and of limit over the life of an account. We then weight the predicted balance by the probability that in that month balance is less than limit and we weight the limit by the probability that the balance is greater than the limit. These probabilities are estimated using a survival model. The idea is to exploit the greater accuracy of limit models than of balance models. The model is tested on an out of time out of sample test sets. The model performs well in comparison with established methods.

CO556 Room Gordon ASSET PRICE BUBBLES Chair: Ivan Paya

CO1013: **The shine of precious metals around the global financial crisis**
Presenter: **Roderick McCrorie**, University of St Andrews, United Kingdom
Co-authors: Isabel Figuerola-Ferretti
We analyze the price behaviour of the precious metals gold, silver, platinum and palladium, before, during and in the aftermath of the 2007-08 financial crisis. Using the mildly explosive/multiple bubbles technology, we find significant, short periods of mildly explosive behaviour in the spot and futures prices of all four precious metals. Fewer such periods are detected using exchange-rate adjusted prices, and almost none when deflated prices are used. We assess whether these findings are indicative of bubble behaviour. Convenience yield is shown to have little efficacy in this regard; other fundamentals proxy variables and position data offer only very limited evidence against prices having been anything other than fundamentals-driven. Possible exceptions are in gold in the run-up to the highpoint of the crisis, and in silver and palladium around the launch of specific financial products. Some froth, however, is reported and discussed for each metal.

CO1178: **Improving the accuracy of asset price bubble start and end date estimators**
Presenter: **Robert Sollis**, Newcastle University, United Kingdom
Recent research has proposed using recursive right-tailed unit root tests to date the start and end of asset price bubbles. An alternative approach is proposed that utilises model-based minimum sum of squared residuals estimators combined with Bayesian Information Criterion model selection. Conditional on the presence of a bubble, the dating procedures suggested are shown to offer consistent estimation of the start and end dates of a fixed magnitude bubble, and can also be used to distinguish between different types of bubble process, i.e. a bubble that does or does not end in collapse, or a bubble that is ongoing at the end of the sample. Monte Carlo simulations show that the proposed dating approach out-performs the recursive unit root test methods for dating periods of explosive autoregressive behaviour in finite samples, particularly in terms of accurate identification of a bubble's end point. An empirical application involving Nasdaq stock prices is discussed.

CC0712: **Testing for speculative bubbles using spot and forward exchange rates: An application to the German hyperinflation**
Presenter: **Ivan Paya**, Lancaster University, United Kingdom
Co-authors: Efthymios Pavlidis, David Peel
The probabilistic structure of periodically collapsing bubbles creates a gap between future spot and forward exchange rates in small samples. We exploit this fact to propose two novel methods for detecting bubbles. The first method is based on recursive unit root tests and the second on rolling Fama regressions. Both methods do not rely on a particular model of asset price determination, they are robust to an explosive root in the process for market fundamentals, and are accompanied by a date-stamping strategy. As an empirical application, we analyze the German mark-US dollar exchange rate for the interwar German hyperinflation period.

CO1315: **Crude oil prices did not exhibit bubble behavior: Evidence using OVIX-adjusted WTI**
Presenter: **Isabel Figuerola-Ferretti**, ICADE, Spain
Co-authors: Roderick McCrorie, Ioannis Paraskevopoulos
A recent technology is used to test for bubble behavior in WTI crude oil front month futures prices over the last decade. Our sample encompasses both the pre-crisis period, in which there was a substantial run-up in crude oil prices, and the recent period in which prices have fallen significantly. Results using the raw series suggest there were two bubbles, a positive bubble in 2008 and a negative bubble from November 2014 to January 2015. The PSY test, however, makes the assumption of constant volatility across regimes. To adjust for this, we apply the test to the raw series deflated by CBOE crude oil VIX volatility, and show there no evidence to support bubble behavior in the deflated series. Our results suggest that when we account for forward looking option market volatility, there is no evidence to suggest there were bubbles in crude oil contrary to popular belief.

CO434 Room G21A GOODNESS-OF-FIT, MULTIPLE PREDICTORS AND MULTIVARIATE MODELS Chair: Lynda Khalaf

CO1791: **Dynamic panel analysis of market debt ratios**
Presenter: **Charles Saunders**, University of Western Ontario, Canada
Co-authors: Lynda Khalaf, Marie-Claude Beaulieu
Interest in dynamic panel data models for market debt ratios have emerged in recent years. The robustness of several methods to some common problems that affect standard estimators is examined. We first extend available results on the GMM estimator by applying recently proposed alternative methods. These include indirect inference, GMM approach, and X-differencing. The coverage of the confidence intervals for the various methods is also surveyed, including the Monte Carlo inversion approach to construct exact confidence sets.

CO1796: **Goodness-of-fit and model selection when the regressand is discrete**
Presenter: **Dalibor Stevanovic**, Universite du Quebec a Montreal, Canada
Co-authors: Rachidi Kotchoni
We propose a new goodness-of-fit measure adapted for regression models with a discrete regressand. Examples of such models include Probit

and Logit regressions for binary data and Poisson and Negative Binomial regressions for count data. The proposed goodness-of-fit measure is deduced from a Kullback-Leibler Information Criterion (KLIC) and decreasing in the distance between the probability distribution of the regressand predicted by the regression model and the corresponding marginal empirical distribution. Like a standard R-square, it lies between 0 and 1. For that reason, we called it the KLIC-R2. We show how to use Monte Carlo simulations or by parametric bootstrap to approximate the distribution of the KLIC-R2 conditionally on the observed realizations of the regressors. This distribution is obviously not pivotal and it must be simulated for each case. We illustrate the use of the KLIC-R2 to test the null hypothesis that a particular set of regressors has no predictive power against the most general alternative. Our Monte Carlo simulations show that the KLIC-R2 compares favorably to existing model selection criteria for discrete choice models.

CO0922: **Small-sample tests for multiple-predictor regressions**
Presenter: **Sermin Gungor**, Bank of Canada, Canada
Standard single-equation tests for predictability tend to over-reject when there is feedback from disturbances to future values of a predictor and the predictor variable is highly persistent. Our framework accounts for feedback to multiple persistent predictors by modelling them as a vector autoregression along with the predictive regression for the dependent variable, and by leaving free the joint distribution of system disturbances. We deal with the presence of nuisance parameters by establishing a pivotal bound on the null distribution of the full-information quasi-likelihood ratio statistic and we use Monte Carlo resampling techniques to obtain an exact small-sample test procedure.

CO454 Room Woburn NOWCASTING AND FORECASTING UNDER UNCERTAINTY I Chair: Katja Heinisch

CO0324: **Regional surveys: Identifying uncertainty and forecasting economic growth**
Presenter: **Katja Heinisch**, Halle Institute for Economic Research, Germany
Policy makers in regional institutions are more and more interested in a frequently and timely assessment and projection of economic growth in particular on regional GDP forecasts. However, regional data is only available with substantial delay for the German states from the German statistical office. Hence, particular weight is given to own survey data besides the official hard data to improve the regional GDP and gross value added forecasts. The aim is to analyze the performance of the IWH construction and industry surveys for the economic development and outlook of the East German economy. We examine whether these surveys are useful to improve the nowcast of the current and the forecast of the next quarter of GDP. Given the even number of response possibilities of these surveys all answers can be taken into account. This is also a big advantage compared to the ifo business climate index for East Germany where responses that correspond to unchanged are neglected. However, the overall performance of both indices is compared. In addition, the study uses the responses of the IWH survey to identify the uncertainty of the entrepreneurs for the assessment of the economic situation. Firms predictions in previous periods about expected changes will be compared with their assessment of realized changes. Derived from cross-sectional disagreement we can derive a measure of ex post forecast error uncertainty.

CO0770: **Nowcasting public finances in Italy**
Presenter: **Peter Claeys**, Vrije Universiteit Brussel, Belgium
Budget forecasts have become increasingly important as a tool of fiscal management to influence expectations of bond markets and the public at large. The inherent difficulty in projecting macroeconomic variables together with political bias thwart the accuracy of budget forecasts. The purpose is to make available efficient and accurate forecasts trends of fiscal deficits in those months in which official forecasts are not published. We improve accuracy by nowcasting public deficits for Italy over the period 1993-2012. We analyse monthly series of public finance (a cash indicator) to predict out-of-sample forecasts with leading indicators and internet search terms. The real-time deficit forecast from a VAR model beats a set of other private and public forecasts of the deficit. Deficits are hard to predict due to shifting economic conditions and political events. We test and compare predictive accuracy over time and although a weighted combined forecast is robust to breaks, it is hard to significantly improve over a simple RW model.

CO1051: **Lessons for forecasting unemployment in the U.S.: Use flow rates, mind the trend**
Presenter: **Murat Tasci**, Federal Reserve Bank of Cleveland, United States
The aim is to evaluate the ability of autoregressive models, professional forecasters, and models that leverage unemployment flows to forecast the unemployment rate. We pay particular attention to flows-based approaches to generalize whether data on unemployment flows is useful in forecasting the unemployment rate. We find that any approach that leverages unemployment inflow and outflow rates performs well in the near term. Over longer forecast horizons, we find a useful framework, even though it was designed to be mainly a tool to uncover long-run labor market dynamics such as the "natural" rate. Its usefulness is amplified at specific points in the business cycle when unemployment rate is away from the longer-run natural rate. Judgemental forecasts from professional economists tend to be the single best predictor of future unemployment rates. However, combining those guesses with flows-based approaches yields significant gains in forecasting accuracy.

CO0778: **On business confidence as an indicator for industrial production: Evidence from the EC survey**
Presenter: **Marco Malgarini**, ANVUR, Italy
Co-authors: Stefano Fantacone, Petya Garalova, Eleonora Mazzoni
Business surveys are usually considered as good indicators for industrial production, being not revised, not in need to be filtered and released in advance with respect to quantitative data. However, evidence of a possible break in the relationship among survey data and industrial production (IP) has recently emerged: we look closely at this relationship using Eurostat and EC data for EU countries, analyzing the rolling correlation among soft and hard data and looking at coherence of the two series at cyclical turning points. We also check for the capability of business surveys of causing IP in the sense of Granger and evaluate the evolution of this relationship over time with rolling methods. Emerging differences in cyclical behavior may be either linked to statistical problems in sampling selection during the crisis, or to changes in the way agents form expectations. If long term perceptions of normal levels of output are lower than in the past, it is well possible for opinion variables to show a favorable trend even if the underlying quantitative variable does not show remarkable changes.

EI010 Room CLO B01 SPECIAL SESSION ON ROBUSTNESS FOR FUNCTIONAL AND COMPLEX DATA Chair: Graciela Boente

EI0333: **Fast estimation of the geometric median in Hilbert spaces: An application to robust online principal components analysis**
Presenter: **Herve Cardot**, Universite de Bourgogne, France
It is not unusual anymore to have to analyze very large samples of high dimensional data, that also may be observed sequentially. In such a large sample/high dimensional data context outliers may be hard to detect automatically and considering robust indices of location instead of (eventually trimmed) mean values can be interesting. The geometric median, which is a natural extension of the median in metric spaces is an interesting candidate for estimating the center of a distribution in a robust way. A very simple and efficient recursive estimator of the geometric median in Hilbert spaces is studied. It allows for automatic update and does not require to store all the data. It is also shown that its asymptotic distribution is the same as the minimizer of the empirical risk. An application to robust PCA, via the online estimation of the median "covariation" matrix, is given. Illustrations on both simulated data and real data confirm the interest of using this new robust dimension reduction technique.

EI0967: **Robust analysis of high-dimensional functional data**
Presenter: **Juan Romo**, Universidad Carlos III de Madrid, Spain
Co-authors: Ana Arribas-Gil
Functional data analysis refers not only to one-dimensional functional observations, but also to multivariate samples of functions or even to high-dimensional functional data sets. Robust analysis of functional data can be based on the concept of depth, that allows us to establish the notion of centrality and extremality and provides fundamentals for testing or classification. We analyze depth for high-dimensional functional data and apply these ideas to simulated high-dimensional functional samples and to real high-dimensional functional observations.

EI1064: **Robustness against cellwise outliers**
Presenter: **Peter Rousseeuw**, KU Leuven, Belgium
A multivariate dataset consists of n observations in p dimensions, and is often stored in an n by p matrix X. Robust statistics has mostly focused on identifying and downweighting outlying rows of X, called rowwise or casewise outliers. However, downweighting an entire row if only one (or a few) of its cells are deviating entails a huge loss of information. Also, in high-dimensional data the majority of the rows may contain a few contaminated cells, which yields a loss of robustness as well. Recently new robust methods have been developed for datasets with missing values and with cellwise outliers, also called elementwise outliers. Several methods of this type will be studied and compared in terms of their robustness as well as their statistical and computational efficiency. Simulation results will be shown as well as real data examples.

EO182 Room MAL 540 STATISTICS FOR COSMOLOGICAL DATA Chair: Domenico Marinucci

EO0219: **The challenge of weak lensing**
Presenter: **Thomas Kitching**, UCL, United Kingdom
Weak lensing distorts observed galaxy shapes by changing their size and ellipticity at a level of one part in a thousand. By measuring these small distortions statistically and in three dimensions both the geometry of the Universe and the growth of structure can be inferred. Several groundbreaking surveys are being planned that will map the majority of the sky to measure this effect, for example the ESA Euclid mission. We will review the state-of-the-art in image analysis inferring ellipticities and sizes from very noisy date; cosmological parameter inference using three-dimensional spherical harmonic analyses; and simulations using large-scale n-body and hydrodynamical methods. We will focus on the outstanding challenges in this field, and motivate the need for a close collaboration between the cosmology, computational and statistical communities.

EO0396: **Sparsity in astrophysics: Astrostatistics meets astroinformatics**
Presenter: **Jason McEwen**, University College London, United Kingdom
Astrostatistics has become a well-established sub-field, where powerful statistical methods are developed and applied to extract scientific information from astrophysical observations. In particular, Bayesian methods have now found wide-spread application in astrophysics. Astroinformatics, on the other hand, is an evolving but less mature sub-field, where informatics techniques provide a powerful alternative approach for extracting scientific information from observational data. Informatics techniques have close links with information theory, signal processing and computational harmonic analysis, and have been demonstrated to be very effective. Wavelet methods, for example, allow one to probe both spatial- and scale-dependent signal characteristics simultaneously. Such techniques are very effective in studying physical phenomena since many physical processes are manifest on particular physical scales, while also spatially localised. Recent developments in this domain have led to the theory of compressive sensing, a revolutionary breakthrough in the field of sampling theory, which exploits the sparsity of natural signals. The aim is to introduce compressive sensing from both the synthesis and analysis perspectives, highlighting statistical connections, and discuss the application of such techniques in astrophysics.

EO0654: **Cosmostatistics in a random Universe**
Presenter: **Andrew Jaffe**, Imperial College, United Kingdom
Cosmological inferences are inherently probabilistic: we can only hope to directly observe a fraction (perhaps vanishingly small) of the Universe. Furthermore, our models are usually themselves probabilistic: they predict only the statistical distribution of the matter and energy in the Universe. In recent years, the amount of data has increased by orders of magnitude and cosmologists have developed analysis methods in the face of these limitations. We review these methods and some of the latest results, paying specific attention to data from the Cosmic Microwave Background.

EO0990: **Wavelets for cosmological data analysis: From the sphere to the 3-dimensional case**
Presenter: **Domenico Marinucci**, University of Rome Tor Vergata, Italy
Co-authors: Claudio Durastanti, Yabebal Fantaye
We start from a review of the construction of spherical wavelets and their use for Cosmological data analysis, focussing in particular on needlets. We discuss their statistical and analytic properties, and then we review a number of recent and less recent applications, such as the investigation for non-Gaussianity and asymmetries in Cosmic Microwave Background data and the search for point sources. We then consider extensions to three-dimensional settings, reviewing in particular the recent proposal of 3d radial needlets and their potential applications to future Large Scale Structure surveys.

EO138 Room MAL 421 CLUSTERING MIXED DATA Chair: Julien Jacques

EO0261: **Mixture model of Gaussian copulas to cluster mixed-type data**
Presenter: **Matthieu Marbac**, McMaster University, Canada
A mixture model of Gaussian copulas is proposed to cluster mixed data. This approach allows us to straightforwardly define simple multivariate intra-class dependency models while preserving classical distributions for the one-dimensional margins of each component in order to facilitate the model interpretation. Moreover, the intra-class dependencies are taken into account by the Gaussian copulas which provide one robust correlation coefficient per couple of variables and per class. This model generalizes different existing models defined for homogeneous or mixed variables. The Bayesian inference is performed via a Metropolis-within-Gibbs sampler. The model is illustrated by a real data set clustering.

EO0248: **Kernel discriminant analysis with parsimonious Gaussian process models**
Presenter: **Charles Bouveyron**, University Paris Descartes, France
A family of parsimonious Gaussian process models is presented which allows to build, from a finite sample, a model-based classifier in an infinite dimensional space. The proposed parsimonious models are obtained by constraining the eigen-decomposition of the Gaussian processes modeling each class. This allows, in particular, to use non-linear mapping functions which project the observations into infinite dimensional spaces. It is also demonstrated that the building of the classifier can be directly done from the observation space through a kernel function. The proposed classification method is thus able to classify data of various types such as categorical data, functional data or networks. Furthermore, it is possible to classify mixed data by combining different kernels. The methodology is as well extended to the unsupervised classification case. Experimental results on various data sets demonstrate the effectiveness of the proposed method.

EO0656: **Model-based clustering with mixed/missing data using the new software MixtComp**
Presenter: **Christophe Biernacki**, Inria, France

The "Big Data" paradigm involves large and complex data sets where the clustering task plays a central role for data exploration. For this purpose, model-based clustering has demonstrated many theoretical and practical successes in a various number of fields. MixtComp is a new software, written in C++, implementing model-based clustering for multivariate missing/binned/mixed data under the conditional independence assumption. Current implemented mixed data are continuous (Gaussian), categorical (multinomial), integer (Poisson) and ordinal (specific model) ones. However, architecture of MixtComp is designed for incremental insertion of new kinds of data (ranks, functional, directional...) and related models. Model estimation is performed by a Stochastic EM algorithm (SEM) and several classical model selection criteria are available (BIC, ICL). Currently, MixtComp is not freely provided as an R package but is freely available through a specific user-friendly web interface (https://modal-research.lille.inria.fr/BigStat/) and its output corresponds to an R object directly usable in the R environment. Beyond its clustering task, it also allows us to perform imputation of missing/binned data (with associated confidence intervals) by using the mixture model ability for density estimation as well.

EO1117: **The moneyball effect: How statistical analysis can transform the sports industry**
Presenter: **Yang Tang**, McMaster University, Canada
Co-authors: Paul McNicholas

In the past few years, professional sports teams have been relying more on computer-driven statistical analysis. This is sometimes known as the "Moneyball effect". Psychometric data is available for the Canadian women rugby sevens team, including daily psychometric and training variables on 32 female athletes over a six-month period. Sixteen variables are considered: three continuous variables, eight ordinal variables, three categorical variables, and two integer variables. Model-based approaches via a mixture model are used to distinguish groups as well as characterize them. We analyze the data using mixture modelling approaches to define distinct psychological status. We further use a data mining technique, association rule mining, to investigate relationships between psychological status and physical performance within each group. We expect a continuous learning process that could give better outcomes thereby contributing to a transformation within the sports industry.

EO136 Room MAL B33 METHODS FOR THE ANALYSIS OF SEMI-COMPETING RISKS DATA Chair: Sebastien Haneuse

EO0359: **Estimating restricted mean job tenures in semi-competing risks data compensating victims of discrimination**
Presenter: **Qing Pan**, George Washington University, United States
Co-authors: Joseph Gastwirth

When plaintiffs prevail in a discrimination case, a major component of the calculation of economic loss is the length of time they would have been in the higher position had they been treated fairly. This problem is complicated by the fact that one's eligibility for promotion is subject to termination by retirement and both the promotion and retirement processes may be affected by discriminatory practices. The semi-competing risk data are decomposed into the marginal retirement process and the promotion process conditional on that retirement has not occurred. Predictions for compensation purpose are made by utilizing the expected promotion and retirement probabilities of similarly qualified members of the non-discriminated group. The restricted mean durations of three periods are estimated - the time an employee would be at the lower position, at the higher level and in retirement. The proposed restricted mean job duration estimators are shown to be robust in the presence of an independent frailty term. Data from the reverse discrimination case, Alexander v. Milwaukee, where white male lieutenants were discriminated in promotion to captain are reanalyzed. While the appellate court upheld liability, it reversed the original damage calculations, which heavily depended on the time a plaintiff would have been in each position. The results obtained by the proposed method are compared to those of the original jury. Substantial differences in both directions are observed.

EO0493: **A joint frailty-copula model between tumour progression and death for meta-analysis**
Presenter: **Virginie Rondeau**, University of Bordeaux INSERM, France
Co-authors: Takeshi Emura, Masahiro Nakatochi, Kenta Murotani

Dependent censoring often arises in biomedical studies when time to tumour progression (e.g., relapse of cancer) is censored by an informative terminal event (e.g., death). For meta-analysis combining existing studies, a joint survival model between tumour progression and death has been considered under semicompeting risks, which induces dependence through the study-specific frailty. Copulas are utilized to generalize the joint frailty model by introducing additional source of dependence arising from intra-subject association between tumour progression and death. The practical value of the new model is particularly evident for meta-analyses in which only a few covariates are consistently measured across studies and hence there exist residual dependence. The covariate effects are formulated through the marginal Cox proportional hazards, and the baseline hazards are nonparametrically modeled on a basis of splines. The estimator is then obtained by maximizing a penalized log-likelihood function. We also show that the present methodologies are easily modified for the competing risks setup, and are generalized to accommodate left-truncation. Simulations are performed to examine the performance of the proposed estimator. The method is applied to a meta-analysis for assessing a recently suggested biomarker CXCL12 for survival in ovarian cancer patients. We implement our proposed methods in R joint.Cox package (version 1.1).

EO0944: **Stabilised Aalen-Johansen estimator of the transition probabilities to protect against too small risk sets**
Presenter: **Arthur Allignol**, Ulm University, Germany
Co-authors: Sarah Friedrich, Jan Beyersmann

Left-truncation (delayed entry) arises if study entry occurs after time 0. Failure to account for left-truncation leads to length-bias, i.e., the number of individuals which are considered to be under observation at time 0 is falsely inflated leading to biased estimates of the target quantity. Survival analysis techniques naturally account for left-truncation. However the number of individual at risk at the beginning of the follow-up might be very small, leading to highly variable and thus potentially unreliable estimates. A modified version of the Kaplan-Meier estimator has been previously proposed that tackles these issues by discarding contributions from too small risk sets. The work was later extended by proposing a modified Aalen-Johansen estimator in the the competing risks setting. Based on the latter, we introduce a modified Aalen-Johansen estimator for general Markov multistate models. This is relevant even in the absence of left-truncation. For instance, in an illness-death model without recovery, initial state 0, intermediate illness state 1 and absorbing death state 2, there will be internal left-truncation due to 0 to 1 transitions. A simulation study and an example illustrate the use of this estimator.

EO0895: **Effects of unobserved heterogeneity on methods for analysing prevalent cohort and current duration designs**
Presenter: **Niels Keiding**, University of Copenhagen, Denmark

Time to pregnancy is the duration from the time a couple starts trying to become pregnant until they succeed. It is considered one of the most direct methods to measure natural fecundity in humans. Statistical tools for designing and analysing time to pregnancy studies belong to the general area of survival analysis, but several features require special attention. Recruiting at incidence is difficult to carry out, so that prospective follow-up is more realistically achieved in a prevalent cohort design. Retrospective (pregnancy-based) designs, widely used in this area, do not allow efficiently including couples remaining childless. The current duration design starts from a cross-sectional sample of couples currently trying to become pregnant, using the backward recurrence time as basis for the estimation of time to pregnancy or time to the end of the pregnancy attempt. On the basis of a simulation study, it is studied the effect of unmeasured population heterogeneity on estimates of the distribution of time to pregnancy and of the effect of risk factors. The generality of such properties for prevalent cohort and current duration designs are briefly discussed.

EO288 Room CLO 203 RECENT DEVELOPMENT ON SINGLE INDEX MODELS Chair: Jianhui Zhou

EO0406: **Polynomial spline estimation for generalized partially linear single-index models**
Presenter: **Lily Wang**, Iowa State University, United States
Co-authors: Guanqun Cao
The estimation for the generalized partially linear single-index model is studied, where the systematic component in the model has a flexible semiparametric form with a general link function. An efficient and practical approach is proposed to estimate the single-index link function, single-index coefficients as well as the coefficients in the linear component of the model. The estimation procedure is developed by applying quasi-likelihood and polynomial spline smoothing. Large sample properties of the estimators are derived and the convergence rate of each component of the model is obtained. Asymptotic normality is established for the coefficients in both the single-index and linear components. The polynomial spline approach is computationally stable and efficient in practice. Both simulated and real data examples are used to illustrate our proposed methodology.

EO0490: **A new estimation procedure for single index models**
Presenter: **Xiangrong Yin**, University of Kentucky, United States
A novel and effect algorithm is introduced for solving single index models, especially when data has large p-small n structure. Implementation and Monte Carlo results as well as a real data analysis are presented. Although we focus on single-index models, extension to multiple index models is also discussed.

EO1172: **A new inference approach for varying-coefficient single-index models**
Presenter: **Valentin Patilea**, CREST-ENSAI, France
Semiparametric single-index assumptions are convenient and widely used dimension reduction in regression analysis. In a mean regression setup, the single-index model (SIM) assumption means that the conditional expectation of the response given the vector of covariates is the same as the conditional expectation of the response given an index, that is a linear combination of the covariates. In a conditional distribution modeling, under the SIM assumption the conditional law of a response given the covariate vector coincides with the conditional law given an index. We consider extended SIM setups where the coefficients of the linear combination could depend on an additional covariate. Such extensions allow us to capture some dynamic features in the data while preserving an appealing modeling strategy. We propose a new estimation approach of the varying- coefficients in mean SIM and SIM for the conditional law. The estimator is obtained by minimization of a distance criterion based on kernel smoothing. Consistency and asymptotic normality results are established. The new methodology is illustrated by simulations and real data applications.

EO0554: **A single index model for censored quantile regression**
Presenter: **Jianhui Zhou**, University of Virginia, United States
Co-authors: Miao Lu
Quantile regression has been getting more attention in survival analysis recently due to its interpretability. For possible nonlinear relationship between survival time and risk factors, we study a single index model for censored quantile regression, and employ the local linear approximation for the unknown link function. To account for censoring, we consider the redistribution-of-mass to obtain a weighted quantile regression estimator. The developed estimator can be penalized for variable selection purpose. The asymptotic properties of the developed estimators are investigated. The performance of the developed estimation and variable selection methods is illustrated in simulation studies, and the methods are applied to a real data example.

EO184 Room CLO 101 OBJECT ORIENTED DATA ANALYSIS I Chair: John Kent

EO0414: **Principal nested shape spaces, with applications to molecular dynamics data**
Presenter: **Ian Dryden**, University of Nottingham, United Kingdom
Molecular dynamics simulations produce large datasets of temporal sequences of molecules, such as flexible proteins. It is of interest to summarize the shape evolution of the molecules in a succinct, low-dimensional representation. However, Euclidean techniques such as principal components analysis (PCA) can be problematic as the data may lie far from a flat manifold. Principal nested spheres can lead to striking insights which may be missed using PCA. We provide some fast fitting algorithms and apply the methodology to a large set of 100 runs of 3D protein simulations, investigating biochemical function in applications in Pharmaceutical Sciences.

EO0416: **Backward nested subspaces: Asymptotics, two-sample tests and applications**
Presenter: **Benjamin Eltzner**, University Goettingen, Germany
Dimension reduction is an important tool in multivariate statistical analysis, aiming at a description of the relevant modes of variation in a data set. For data on manifolds or stratified spaces, dimension reduction becomes more involved as forward and backward methods must be distinguished. We introduce the notion of backward nested subspaces as a generalization of principal nested spheres. For such general families of estimated nested subspaces we provide asymptotic results and inferential tools. We illustrate the power of our methods by simulations and applications to data.

EO0464: **Marker invariant analysis of gait data and its applications**
Presenter: **Fabian Telschow**, University of Goettingen, Germany
In biomechanical analysis of motions of the human knee joint, curves over time in the space $SO(3)$ of 3D rotations are the data objects. In non-invasive practice, such curves are obtained by following markers placed on thigh and shank. Even if placed by experts, slightly differing locations account for additional statistical variation that makes it difficult to accomplish the task of identifying individuals as well as of identifying changes in gait patterns due to degenerative effects, say. To accomplish these tasks, it turns out that paths in $SO(3)$ modulo its isometry group form the natural object space for which we develop novel inferential tools, among others simultaneous confidence regions.

EO1157: **Inference for functional data using the adjusted p-value function**
Presenter: **Simone Vantini**, Politecnico di Milano, Italy
Co-authors: Alessia Pini
Inference for functional data embedded in $L^2(a,b)$ is considered, with particular emphasis on the domain selection problem (i.e., detecting those portions of the domain where a given functional null hypothesis is rejected). We will present a general and fully non-parametric approach to achieve that target that is based on the introduction of two new inferential tools: the unadjusted and the adjusted p-value functions. After providing their definitions, we will describe their inferential properties in terms of control of the Type-I error probability and of consistency (point-wise and interval-wise, respectively). Finally, to show the flexibility of the methodology we will provide an overview on some applications in which the *unadjusted* and the *adjusted p-value functions* have been used to face different testing problems such to answer specific research questions pointed out by experts: two-population test for pair-wise comparison of the tongue movements in different experimental settings; functional analysis of variance of reflectance spectra for selecting frequency bands for remote monitoring of laser welding; functional-on-scalar linear model of body part trajectories for the long-term assessment of therapies to fix Anterior Cruciate Ligament injuries.

EO635 Room MAL B20 BAYESIAN SEMI- AND NONPARAMETRIC MODELLING I Chair: Li Ma

EO0448: **A general framework for Bayes structured linear models**
Presenter: **Chao Gao**, Yale University, United States
A unified approach will be given to both Bayes high dimensional statistics and Bayes nonparametrics in a general framework of structured linear models. With the proposed two-step model selection prior, a general theorem of posterior contraction will be presented under an abstract setting. The main theorem can be used to derive new results on optimal posterior contraction under many complex model settings including stochastic block model, graphon estimation and dictionary learning. It can also be used to re-derive optimal posterior contraction for problems such as sparse linear regression and nonparametric aggregation, which improve upon previous Bayes results for these problems. The key of the success lies in the proposed two-step prior distribution. The prior on the parameters is an elliptical Laplace distribution that is capable to model signals with large magnitude, and the prior on the models involves an important correction factor that compensates the effect of the normalizing constant of the elliptical Laplace distribution.

EO0868: **Sparsity party in a haystack**
Presenter: **Jean-Bernard Salomond**, Universite Paris Dauphine, France
Co-authors: Stephanie van der Pas, Johannes Schmidt-Hieber
Sparse models have grown more and more popular in the past years. In the Bayesian setting, the first theoretical results were proved for the spike and slab priors. However, from a practical point of view, these priors are not convenient for computational reasons. In the meanwhile, a class of so called shrinkage priors have been developed. Although one do not recover exact zeroes as with the spike and slab, they are easier to implement and give good results in practice. Among these priors, scale mixture many of the most used can be written as a scale mixture of Gaussian which makes them particularly easy to implement. We propose general conditions on scale mixtures of Gaussian in the sparse Gaussian sequence setting, such that the posterior achieve the minimax risk which also leads to minimax estimator for the signal. These conditions give some general guidelines to choose a shrinkage prior for estimation under sparsity assumption.

EO0874: **Priors for the frequentist, consistency beyond Schwartz**
Presenter: **Bas Kleijn**, University of Amsterdam, Netherlands
Schwartz theorem on posterior consistency is generalized in several respects. First, while testability remains a requirement, it is shown to be equivalent to Doobs Bayesian form of posterior concentration. This type of testability reduces to a measurability condition rather than the topological requirements most frequentists would expect. Second, the frequentist using Bayesian tools requires a stronger form of posterior consistency and imposes a property akin to (but weaker than) contiguity to strengthen Doobs prior-a.s. consistency. It is shown that Schwartzs Kullback-Leibler condition for the prior is sufficient, as well as Freedmans tailfreeness property and related forms of equicontinuity. For rates of convergence and models for dependent samples Le Cams First Lemma is extended, implying sufficiency of the Ghosh-Ghosal-vdVaart construction as well as more recent criteria based on Hellinger transforms. Examples include the stochastic block model, spike-and-slab-type priors for sparse sequences and the Neyman-Scott paradox.

EO0983: **Posterior concentration rate of a class of multivariate density estimators based on adaptive partitioning**
Presenter: **Linxi Liu**, Stanford University, United States
Co-authors: Wing Hung Wong
We study a class of non-parametric density estimators under Bayesian settings. The estimators are piecewise constant functions on binary partitions. We analyze the concentration rate of the posterior distribution under a suitable prior, and demonstrate that the rate does not directly depend on the dimension of the problem. This is as an extension of a companion work where the convergence rate of a related sieve MLE was established. Compared to the sieve MLE, the main advantage of the Bayesian method is that it can adapt to the unknown complexity of the true density function, thus achieving the optimal convergence rate without artificial conditions on the density.

EO206 Room MAL B35 STOCHASTIC PROCESSES WITH APPLICATIONS Chair: Simone Padoan

EO0524: **Stochastic boundaries and other techniques for stochastic PDE models**
Presenter: **Finn Lindgren**, University of Bath, United Kingdom
Recurring issues when constructing spatial and spatio-temporal Gaussian Markov random field models is how to ensure model consistency for different resolutions, and how to impose consistent boundary conditions. Both of these issues can be handled by considering continuous domain stochastic PDE models, that turn into Markov random fields on the coefficients of finite dimensional basis function expansion representations. Such computationally efficient representations can be used for Kriging of large space-time data on complicated domains, and as building blocks in complex latent Gaussian models, such as point process models for animal detections in ecology.

EO0641: **Efficient likelihood-based inference for the Brown-Resnick process**
Presenter: **Emeric Thibaud**, Colorado State University, United States
Max-stable processes are the only non-trivial limits of properly rescaled pointwise maxima of random processes. The Brown-Resnick process is a parametric max-stable model which has proven to be well-suited for modeling extremes of environmental processes. In most applications, the full likelihood of this model is unobtainable, and inference has been based on composite likelihoods, resulting in a loss in efficiency, and preventing the use of standard Bayesian methods. We present a new Bayesian approach to estimate the parameters of a Brown-Resnick process, and we discuss its computational challenges. We illustrate the method with an application to extreme low temperatures.

EO0815: **A nonparametric model for stationary time series**
Presenter: **Isadora Antoniano-Villalobos**, Bocconi University, Italy
Co-authors: Stephen Walker
Stationary processes have been used as statistical models for dependent quantities evolving in time. Stationarity is a desirable model property, however, the need to define a stationary density limits the capacity of such models to incorporate the diversity of the data arising in many real life phenomena. Alternative models have been proposed, usually resulting in a compromise, sacrificing the ability to establish properties of estimators, in favor of greater modeling flexibility. A characterization is provided for a family of time-homogeneous processes with nonparametric stationary densities, which retain the desirable statistical properties for inference, while achieving substantial modeling flexibility, matching those achievable with certain non-stationary models. For the sake of clarity, attention is restricted to first order Markov processes. Posterior simulation involves an intractable normalizing constant. Therefore, a latent extension of the model is presented, which enables exact inference through a trans-dimensional MCMC method. The capabilities of the model are presented through an application.

EO0834: **Extremes of Skew-Symmetric distributions**
Presenter: **Boris Beranger**, Universite Pierre and Marie Curie and University of New South Wales, Australia
Co-authors: Simone Padoan, Scott Sisson
In environmental, economic or financial fields, the data of real applications can exhibit highly asymmetric distributions. In risk management it is important to analyze the frequency that extreme events such as heat waves, market crashes, etc., occur. Such real processes are high-dimensional by nature. Estimating the dependence of extreme events is crucial for predicting future phenomena, that can have a large impact on real life. A simple way of dealing with asymmetrically distributed data is to use the so-called Skew-Symmetric distributions such as the skew-normal and skew-t. These distributions are applied to the context of extreme value theory in order to model the dependence. By defining a non-stationary skew-normal process, which allows the easy handling of positive definite, non-stationary covariance functions, we derive a new family of max-stable processes - the extremal-skew-t process. We provide the spectral representation and the resulting angular densities of the extremal-skew-t process, and illustrate its practical implementation. Finally, an application to wind speed data over 83 locations across the USA is provided.

EO322 Room MAL 415 METHODOLOGY AND APPLICATIONS OF LATENT VARIABLE MODELS Chair: Silvia Pandolfi

EO0550: **Estimating discrete latent models for two-way data arrays: A composite likelihood approach**
Presenter: **Prabhani Kuruppumullage Don**, Dana-Faber Cancer Institute, United States
Co-authors: Francesco Bartolucci, Francesca Chiaromonte, Bruce Lindsay
Composite likelihood is a likelihood modification useful in instances where maximum likelihood estimation (MLE) is computationally infeasible. We present two discrete latent variable models for two-way data arrays, in which MLE is intractable due to the complex structures of the models. The first model aims to sort a two-way array of observations in blocks, each corresponding to a fixed row cluster and a fixed columns cluster (block mixture model). The second model aims to simultaneously produce clusters along one of the data dimensions and contiguous groups along the other (clustering by segmentation). In both cases, we construct composite likelihoods as a computationally tractable alternative to the full likelihood. We also discuss how to evaluate the number of components for each model, and demonstrate the performance of our methods via simulations. Finally, we illustrate the use of our approach through applications to genomic data.

EO0579: **A multivariate latent variable model for the analysis of health status over time**
Presenter: **Silvia Cagnone**, University of Bologna, Italy
Co-authors: Cinzia Viroli
A latent variable model for the analysis of multivariate mixed longitudinal data is proposed. It extends a previous factor mixture model to longitudinal data. The model is based on the introduction of two hidden variables: a continuous latent variable for modeling the association among the observed variables at each time point and a latent discrete variable that follows a first-order Markov chain with the aim of taking into account the unobserved heterogeneity. The aim of the proposed model is twofold: it allows us to perform dimension reduction when data are of mixed type and it performs model based clustering in the latent space. We derive an EM algorithm for the maximum likelihood estimation of the model parameters. The method is illustrated by an application to a longitudinal dataset on health status.

EO1000: **A latent variable approach for dealing with heaping and "too many to count" in domestic violence counts**
Presenter: **Brian Francis**, Lancaster University, United Kingdom
Co-authors: Sylvia Walby, Jude Towers
One recent theory suggests that there are different types of domestic violence. These types are defined by the nature of the behaviour of one partner towards the other. It is hypothesized that one form of domestic violence is likely to lead to repeated acts of violence within a relationship, perhaps escalating in severity. This suggests that a mixture of two count processes may be a good way to model such data. We discuss recent work which has attempted to determine whether there are a number of distinct groups of domestic violence victims, or whether there is a single distribution with overdispersion. Such mixture models however, need to take account not only of capping but of heaping and a "too many to count" category. Heaping is the tendency for respondents to report yearly incidents to the nearest rounded number. This will affect the mixture model in introducing modal peaks in the incident distribution. Dealing with both of these leads to a mixture of mixtures problem, where the heaping and too many to count processes mixes on top of a mixture of Poisson or negative binomials. Results will be presented from an analysis of ten years of data on intimate personal violence from the Crime Survey of England and Wales, and analysis problems are identified.

EO1078: **A discrete-valued latent stochastic process for the estimation of credit migration matrices**
Presenter: **Fulvia Pennoni**, University of Milano-Bicocca, Italy
Co-authors: Giulia Elisei
We focus on the problem of the estimation of the credit migration matrices which are a widely used instrument in the context of risk management to account for the credit portfolio. We propose to apply a latent Markov model in which the possible movement directions are modeled as a continuous or discrete-valued stochastic process following a non-homogenous Markov chain of first or second order. It is a novel approach in this context, which outperform the existing models mainly based on an observed Markov process. The latter do not satisfactory represent the nature of the credit ratings which is mainly governed by underlying forces. We illustrate the traditional techniques employed in this field and their limitations in comparison to our proposal by applying them to simulated data which reproduce the credit ratings histories of Standard and Poor rated firms and sovereigns around the world from 1982 to 2014. We also show how the model may be of interest for prediction of default.

EO040 Room CLO 102 TOPICS IN DIRECTIONAL STATISTICS Chair: Toshihiro Abe

EO0647: **A circular autocorrelation of stationary circular Markov processes**
Presenter: **Hiroaki Ogata**, Tokyo Metropolitan University, Japan
Co-authors: Toshihiro Abe, Takayuki Shiohama, Hiroyuki Taniai
The stationary Markov process is considered and its circular autocorrelation function is investigated. More specifically, a transition density of the stationary Markov circular process is defined by two circular distributions, and we elucidate structure of the circular autocorrelation when the one distribution is uniform and the other is arbitrary. The asymptotic properties of parametric and nonparametric estimators of the circular autocorrelation function are derived. Simulation results are also given. Furthermore, we consider the bivariate process of trigonometric functions and give the explicit form of its spectral density matrix.

EO0857: **Modeling circular Markov processes with time varying autocorrelation**
Presenter: **Takayuki Shiohama**, Tokyo University of Science, Japan
Co-authors: Toshihiro Abe, Hiroaki Ogata, Hiroyuki Taniai
A simple model with time varying autocorrelation for circular Markov processes is proposed. For this purpose, we introduce a time varying concentration parameter in the underlying circular Markov processes. Maximum likelihood estimation as well as the estimation via nonlinear and non-Gaussian state space models are considered. The proposed models are used to illustrate the non-linear relationships between the wind directions and the wind speeds through by time-varying autocorrelations.

EO0250: **The WeiSSVM: A tractable, parsimonious and highly flexible model for cylindrical data**
Presenter: **Christophe Ley**, University of Ghent, Belgium
Co-authors: Toshihiro Abe
The aim is to describe cylindrical distributions obtained by combining the sine-skewed von Mises distribution (circular part) with the Weibull distribution (linear part). This new model, the WeiSSVM, enjoys numerous advantages: simple normalizing constant and hence very tractable density, parameter-parsimony and interpretability, good circular-linear dependence structure, easy random number generation thanks to known marginal/conditional distributions, flexibility illustrated via excellent fitting abilities, and a straightforward extension to the case of directional-linear data. Inferential issues, such as independence testing, can easily be tackled with the WeiSSVM, which we will apply on two distinct real data sets.

EO0921: **Exact and approximate inference for toroidal diffusions**
Presenter: **Eduardo Garcia-Portugues**, University of Copenhagen, Denmark
Co-authors: Michael Sorensen, Kanti Mardia, Thomas Hamelryck
Inference methods are studied for diffusions on the hypertorus. The diffusions constructed have well-known stationary distributions, such as the von Mises or the Wrapped Normal, and can be considered as the analogues of the celebrated Ornstein-Uhlenbeck process on the real line. New pseudo-likelihood procedures based on the Euler and Ozaki discretization schemes are proposed. In addition, approximated likelihood inference is carried out by a suitable imitation of the transition probability densities. The efficiency of these procedures is compared with respect to exact likelihood in a simulation study under a variety of scenarios. Applications to the modelling of backbone angles in molecular dynamics simulations of proteins are shown, together with a procedure to simulate from arbitrary toroidal distributions.

EO294 Room CLO 306 STATISTICAL EVALUATION OF MEDICAL DIAGNOSTIC TESTS Chair: Christos T Nakas

EO0674: **Bayesian bootstrap inference for the ROC surface**
Presenter: **Vanda Inacio**, Pontificia Universidad Catolica Chile, Chile
Co-authors: Miguel de Carvalho
The receiver operating characteristic (ROC) surface is a popular tool for evaluating the accuracy of medical diagnostic tests that classify individuals into one of three ordered classes. We propose a fully nonparametric method based on the Bayesian bootstrap for conducting inferences for the ROC surface and its functionals, such as the volume under the surface. The proposed estimator is based on a simple, yet interesting, representation of the ROC surface in terms of placement variables. The performance of the proposed approach is assessed through a simulation study and a real data application is also provided.

EO0724: **Assessment of diagnostic markers subject to limit of detection**
Presenter: **Maria del Carmen Pardo**, Complutense University of Madrid, Spain
Co-authors: Christos T Nakas, Alba Franco-Pereira
Diagnostic accuracy of potential biomarkers is routinely assessed using ROC curve methodology. The area under the ROC curve (AUC) is the most widely used index of diagnostic accuracy. Frequently, marker measurements are subject to limits of detection (LOD) due to laboratory instrument precision. As a consequence, measurements below a certain level are undetectable. Various strategies have been proposed for the assessment of diagnostic markers that are subject to LOD since, simply ignoring observations below the LOD leads to negatively biased estimates of the AUC. We propose a nonparametric approach based on two-sample likelihood ratio tests for right censored data for the assessment of the diagnostic accuracy of a biomarker subject to LOD. We compare with existing methods through a large simulation study. We illustrate our methods on a potential biomarker associated with the non-dipping phenomenon for patients with sleep apnoea syndrome.

EO1055: **Statistical evaluation of low cost non-inferior imaging diagnostic tests**
Presenter: **Joong-Ho Won**, Seoul National University, Korea, South
Co-authors: Ying Lu
Typical medical imaging research seeks the most accurate cutting edge diagnostic and predictive techniques. Relatively less attentions are given to the alternatives that may have clinically acceptable accuracy but offer benefits in cost, less invasiveness, or less radiation, etc. The problem becomes more important in light of the healthcare reform that wants to contain the expenditure while improve the access. Validation of such techniques has several difficulties. First, different from techniques based on blood or specimen, imaging techniques often have no historical samples to study and require new study to collect data. For rare events, it can be costly. Second, the rapid technology evolution requires such validation studies to be short in order to keep the evaluation relevant. Using an example of dual X- ray absorptiometry (DXA), the standard diagnostic test for osteoporosis, and quantitative ultrasound (QUS), a low cost alternative, we will illustrate why we should be interested such a low-cost alternatives, present a tree-structured subgroup analysis algorithm to identify those who will benefit most from such low cost QUS for diagnosis for prevalent osteoporotic spinal fracture and propose a new statistical design that extends traditional case-control study by including short-term follow-up to estimate prospective the odds ratio for hip fractures.

EO1852: **Inference issues for true-class fractions in 2D and 3D ROC analysis**
Presenter: **Christos T Nakas**, University of Thessaly, Greece
The three-class approach is used for progressive disorders when researchers want to classify subjects as members of one of three categories based on a continuous diagnostic marker. The optimal cut-off points required for this classification are often chosen to maximize the generalized Youden index. The effectiveness of these chosen cut-off points can be evaluated by estimating their corresponding true class fractions and their associated confidence regions. In the two-class case, parametric and nonparametric methods were investigated for the construction of confidence regions for the pair of the Youden-index-based optimal sensitivity and specificity fractions that take into account the correlation introduced between sensitivity and specificity when the optimal cut-off point is estimated from the data. A parametric approach based on the BoxCox transformation to normality often works well otherwise a procedure using logspline density estimation can be used. The true class fractions that correspond to the optimal cut-off points estimated by the generalized Youden index are correlated similarly to the two-class case. We present pitfalls in the assumptions of correlation between true-class fractions in the 2-class case and a generalisation of previous methods to the three-class case, where ROC surface methodology can be employed for the evaluation of the discriminatory capacity of a diagnostic marker.

EO180 Room MAL G15 INFERENCE IN MULTIPLE AND MATRIX GRAPHICAL MODELS Chair: Anna Gottard

EO0706: **A Bayesian approach for array-variate biological networks**
Presenter: **Francesco Stingo**, UT MD Anderson, United States
Co-authors: Yang Ni, Veerabhadran Baladandayuthapani
Multidimensional data constituted by measurements along multiple axes have emerged across many scientific areas such as genomics and cancer surveillance. Traditional multivariate approaches are unsuitable for such highly structured data due to inefficiency, loss of power and lack of interpretability. We illustrate a novel class of multidimensional graphical models that includes both directed and undirected graphs as well as arbitrary combinations of these.

EO0479: **Differential networks: Interpretation of an edge**
Presenter: **David Macleod**, London School of Hygiene and Tropical Medicine, United Kingdom
A differential network refers to an analysis method which compares two networks each representing one of two subgroups. For example we can compare the metabolomic network formed by a set of diseased individuals with the network formed by a set of disease free individuals, with an aim of identifying structural differences which could provide information about the aetiology of disease. The primary method of estimating this network is to estimate the difference in partial correlations of each pair of nodes in each of the two subgroups, with a suitably large difference leading to the inclusion of an edge in the differential network. Most of the recent literature concerning this method focuses on the method of estimating the edges in the network. However, little time has been spent on interpreting what an edge in a network of this type means. A simulation study has been performed, investigating how the method responds to a particular type of data where a pair of nodes in the network are causes of disease to begin to develop our understanding of what a differential network represents.

EO1367: **Topic-adjusted visibility metric for scientific articles**
Presenter: **Siew Li Linda Tan**, National University of Singapore, United Kingdom
Co-authors: Aik Hui Chan, Tian Zheng
Measuring the impact of scientific articles is important for evaluating the research output of individual scientists, academic institutions and journals. While citations are raw data for constructing impact measures, there exist biases and potential issues if factors affecting citation patterns are not properly accounted for. We address the problem of field variation and introduce an article level metric useful for evaluating individual articles' visibility. This measure derives from joint probabilistic modeling of the content in the articles and the citations amongst them using latent Dirichlet allocation (LDA) and the mixed membership stochastic blockmodel (MMSB). Our proposed model provides a visibility metric for individual articles adjusted for field variation in citation rates, a structural understanding of citation behavior in different fields, and article recommendations which take into account article visibility and citation patterns. We develop an efficient algorithm for model fitting using variational methods. To scale up to large networks, we develop an online variant using stochastic gradient methods and case-control likelihood approximation. We apply our methods to the benchmark KDD Cup 2003 data set with approximately 30,000 high energy physics papers.

EO1210: **Efficient model search over DAG causal hypotheses about brain connectivity within a given population**
Presenter: **James Quartermaine Smith**, University of Warwick, United Kingdom
Directed DAGs of probabilistic dynamic processes are currently now being used as frameworks for describing the dynamic relationships between centres of activity in a given unit. For example one recent application of these methods can be found in applications of Markov graphics to model he dynamically evolving relationships between the regions of a brain - as measured by fMRI experiments - as a particular individual engages in various activities. The space of models of this type is obviously huge and so scoring the efficacy of different explanations in terms of their fit to this highly multivariate time series data is a challenging one. However an important related issue is how inferences can be made concerning the nature of the connectivities between centres of activity lying in subpopulations of individuals - for example those suffering from a particular medical condition as opposed to those who do not. The aim is to discuss some new efficient search methodologies - originally designed for the analysis of models concerning single subjects - can be adapted for use in the study of these even larger composite systems. We will then discuss some of the inferential and interpretational challenges these methodologies still face.

EO060 Room MAL B29 APPLICATIONS OF EMPIRICAL MEASURES AND EMPIRICAL PROCESSES Chair: Eric Beutner

EO0760: **Large deviations for weighted empirical measures arising in importance sampling**
Presenter: **Henrik Hult**, KTH Royal Institute of Technology, Sweden
Importance sampling is a popular method for efficient computation of various properties of a distribution such as probabilities, expectations, quantiles, etc. The output of an importance sampling algorithm can be represented as a weighted empirical measure, where the weights are given by the likelihood ratio between the original distribution and the sampling distribution. The efficiency of an importance sampling algorithm is studied by means of large deviations for the weighted empirical measure. The main result, which is stated as a Laplace principle for the weighted empirical measure arising in importance sampling, can be viewed as a weighted version of Sanov's theorem. The main theorem is applied to quantify the performance of an importance sampling algorithm over a collection of subsets of a given target set as well as quantile estimates. The proof of the main theorem relies on a weak convergence approach to large deviations.

EO0223: **Concordance-assisted learning for estimating optimal individualized treatment regimes**
Presenter: **Wenbin Lu**, North Carolina State University, United States
A new concordance-assisted learning (CAL) is presented for estimating optimal individualized treatment regimes. First, we will introduce a type of concordance function for prescribing treatment and propose a robust rank regression method for estimating the concordance function. Then, we will discuss the proposed CAL methods for estimating optimal treatment regimes that maximize the concordance function, named prescriptive index, and for searching the optimal threshold. Moreover, we will discuss the convergence rates and asymptotic distributions of the proposed estimators for parameters in the prescriptive index and the optimal threshold. Finally, we will present some simulations and an application to an AIDS data to illustrate the practical use and effectiveness of the proposed methodology.

EO0345: **Donsker and Glivenko-Cantelli theorems for a class of processes generalizing the empirical process**
Presenter: **Davit Varron**, University of Franche-Comte, France
The aim is to present a general limit theorem for sequences of random discrete probability measures for which the structural assumption is that the point masses are conditionally i.i.d. given the weights. Such a class of random discrete measures not only encompasses the empirical measure, but also encompasses the normalised homogenous completely random measures, stick breaking random measures (or priors). We shall state a Donsker and a Glivenko Cantelli theorem for thoses radom measures, as processes indexed by a class of functions admitting a uniform entropy integral. As a by product of our results, we will provide an alternative proof of the posterior consistency and the Bernstein von Mises phenomenon (in strong topologies) of the Dirichlet process prior.

EO0471: **A refined notion of qualitative robustness for estimators derived from empirical measures**
Presenter: **Henryk Zaehle**, Saarland University, Germany

A refinement of Hampel's notion of qualitative robustness for empirical plug-in estimators is proposed. First, we present criteria for the refined notion of robustness. Then, we apply these criteria to some specific plug-in estimators including empirical risk measures and maximum likelihood estimators. Finally, we explain how the refined notion of robustness can be used to reasonably classify plug-in estimators w.r.t. their degrees of robustness.

EO312 Room CLO 204 RECENT DEVELOPMENTS IN OPTIMAL DESIGN OF EXPERIMENTS Chair: Chiara Tommasi

EO0774: **MV-optimization in binary response models**
Presenter: **Victor Casero-Alonso**, University of Castilla-La Mancha, Spain
Co-authors: Jesus Lopez-Fidalgo, Ben Torsney

MV-optimality will be considered in order to minimize the greater variance of the estimates of the parameters, obtaining maximum information for fitting the model. In particular, binary response models are studied. These models can be seen as weighted linear regression models when the weight function has a finite integral. A methodology and a software tool for obtaining MV-optimal designs in a compact interval are given. Some illustrative examples are provided for some standard weight functions where the design space is a general interval $[a,b]$ on the real line. A representation of graphical MV-optimal designs is made in the Euclidean plane taking a and b as the axes.

EO0931: **Optimal discrimination design between copula models**
Presenter: **Elisa Perrone**, Institute of Science and Technology, Austria
Co-authors: Werner Mueller

Copula modeling is largely employed in many areas of applied statistics. However, the design of related experiments is still a neglected aspect. The issue related to the choice of the copula is analyzed by using discrimination design techniques. A first look into the problem is given by considering the usage of Ds-optimality in the sense of model discrimination. Moreover, we report some examples to highlight the flexibility of such approach. Finally, the application of other discrimination criteria is also discussed.

EO0894: **Optimal design, Lagrangian and linear model theories: A fusion**
Presenter: **Ben Torsney**, Glasgow, United Kingdom

We consider the problem of optimizing a criterion of several variables, subject to them satisfying several linear equality constraints. Lagrangian Theory requires that at an optimum all partial derivatives be exactly linear in a set of Lagrange Multipliers. It seems we can argue that the partial derivatives, viewed as response variables, must exactly satisfy a Linear Model with the Lagrange Multipliers as parameters. This then is a model without errors implying a fitted model with zero residuals. The residuals appear to play the role of directional derivatives in optimal design. Further, if all variables are nonnegative, we can exploit the multiplicative algorithm for finding optimal design weights.

EO1244: **Approximate optimal design for generalised linear mixed effects models**
Presenter: **Noha Youssef**, American University in Cairo, Egypt

Finding an optimal design for a generalised linear mixed effects model involves the computation of Fisher information matrix. Fisher information matrix requires the calculations of the second partial derivatives and expectations. We consider the cases where Fisher information matrix does not have a closed form or the evaluation of these expectations are time consuming. Three alternatives for approximating Fisher information matrix of are explored. The three alternatives are Laplace transformation, Gaussian quadrature and the Stochastic approximation optimization algorithms. Numerical examples are presented to compare the efficiency of the three optimal designs obtained via the three different methods.

EO096 Room MAL B34 DEPENDENCE MODELS AND COPULAS I Chair: Fabrizio Durante

EO0872: **Conditional, partial and average copulas and association measures**
Presenter: **Irene Gijbels**, Katholieke Universiteit Leuven, Belgium

Classical examples of association measures between two random variables include Pearson's correlation coefficient, Kendall's tau and Spearman's rho. For the situation where another variable influences the dependence between the pair, so-called partial association measures, such as a partial Pearson's correlation coefficient and a partial Kendall's tau, have been proposed. In recent years conditional association measures have been studied, such as a conditional Kendall's tau. Such an association measure can be expressed in terms of a conditional copula. Even in case the dependence structure between two variables is influenced by a third variable, one still wants to be able to summarize the dependence structure of the pair by one single number, taking into account the additional influence. We discuss two different ways to do this. We study conditional, partial and average copulas and association measures, discuss non- and semiparametric estimation of these, and investigate their asymptotic behaviour. Examples are given to illustrate the use of the concepts and methods.

EO0771: **Conditional copula simulation for systemic risk stress testing**
Presenter: **Claudia Czado**, Technische Universitaet Muenchen, Germany

Since the financial crisis of 2007-2009 there is an active debate of regulators and academic researchers on systemic risk, with the aim of preventing similar crises in the future or at least reducing their impact. A major determinant of systemic risk is the interconnectedness of the international financial market. We propose to analyze interdependencies in the financial market using copulas. In particular we use the class of vine copulas, which overcome limitations of the popular elliptical and Archimedean copulas. To investigate contagion effects among financial institutions, we develop methods for stress testing by exploiting the underlying dependence structure. In a case study of 38 major international institutions, 20 insurers and 18 banks, we then analyze interdependencies of CDS spreads and perform a systemic risk stress test using single and multiple stress factors. The specified dependence model and the results from the stress test provide new insights into the interconnectedness of banks and insurers.

EO0925: **Working capital and market risk**
Presenter: **Alexandra Dias**, University of Leicester, United Kingdom
Co-authors: Silvia Pazzi

The estimation of market risk is essential in financial risk management. The measure of risk mostly used in practice still is Value-at-Risk (VaR). Most of the methodologies available to estimate VaR rely on modelling the price process disregarding other information concerning the firm. We investigate the relation between equity VaR and the liquidity of the firm. We use as a measure of the liquidity of a firm the Working Capital to Total Assets Ratio (WCR). The WCR enters in the well-known Z-score model for the probability of bankruptcy. We find that the dependence structure between VaR and WCR can be modelled by a t-copula. Hence, the relation between VaR and WCR is linear with tail dependence. We find that the correlation between VaR and WCR is positive although we would expect higher risk to be associated with lower liquidity. We find the explanation for this puzzling result based on the quality of the accrual components of the working capital. We show that the strong dependence found between VaR and WCR via a t-copula can be used to improve the estimation of market risk.

EO0808: **On sharp inequalities between Kendall's tau and Spearman's rho**
Presenter: **Wolfgang Trutschnig**, University of Salzburg, Austria

Using properties of shuffles of copulas and tools from combinatorics we solve the long-standing open question about the exact region Ω determined by all possible values of Kendalls τ and Spearmans ρ. In particular, we prove that the well-known inequality between Kendall's tau and Spearman's rho is not sharp outside a countable set, give a simple analytic characterization of Ω in terms of a continuous, strictly increasing piecewise concave function, and show that is compact and simply connected, but not convex. The results also show that for each $(x,y) \in \Omega$ there are mutually completely dependent random variables X,Y whose τ and ρ values coincide with x and y respectively.

EO314 Room MAL 414 HIGH DIMENSIONS AND SMALL SAMPLE SIZES IN MULTIVARIATE INFERENCE Chair: Markus Pauly

EO0877: **Bootstrapping the de-sparsified Lasso**
Presenter: **Ruben Dezeure**, ETH Zurich, Switzerland
Co-authors: Peter Buehlmann, Cun-Hui Zhang

Assigning statistical significance in high-dimensional linear models has been a very active research area as of late. Some very different approaches have been taken with differing assumptions and empirical performance. We focus on the de-sparsified Lasso, highlight some strengths and weaknesses, and discuss bootstrapping procedures. The implications of model misspecification will be discussed.

EO0893: **Regularized M-estimation of multivariate scatter**
Presenter: **Lutz Duembgen**, University of Bern, Switzerland

At first we discuss briefly the minimization of arbitrary smooth functionals on the space of symmetric, positive definite matrices. In this context we introduce briefly the notion of geodesic convexity. Then we apply these considerations to regularized M-functionals of multivariate scatter. We conclude with some theoretical considerations and numerical examples.

EO0451: **Non- and semiparametric ways to infer from multivariate data: Asymptotics, and approximations**
Presenter: **Arne Bathke**, University of Salzburg, Austria
Co-authors: Solomon Harrar

The aim is to consider the case of drawing inference from multivariate data in factorial designs using non- or semiparametric approaches, not assuming multivariate normality. Different asymptotic frameworks are considered, in addition to approximations for cases where the total number of subjects is small to moderate. Non- and semiparametric models each have their advantages and disadvantages which are discussed.

EO0168: **Permuting repeated measures data despite all the dependencies**
Presenter: **Markus Pauly**, University of Ulm, Germany
Co-authors: Edgar Brunner, Sarah Friedrich

In a repeated measures design, the subjects are repeatedly observed at different occasions, e.g. at different time points. Classical repeated measures models assume that the observation vectors are independent with normally distributed error terms and a common covariance matrix for all groups. In medical and biological research, however, these two assumptions are often not met and may inflate the type-I error rates of the corresponding procedures. We present a different approach working under covariance heterogeneity and without postulating any specific underlying distribution. This procedure is based on an asymptotic pivotal statistic and its small sample behavior is improved by means of an adequate permutation technique. The latter leads to astonishingly successful results despite all the dependencies in the repeated measures design which is shown in extensive simulations. Moreover, the theoretical properties of the method are analyzed and it is applied to a practical data set.

EO222 Room MAL B36 STATISTICAL METHODS FOR BIG DATA AND ANTIFRAUD ANALYSIS Chair: Domenico Perrotta

EO1185: **Goodness-of-fit testing for the Newcomb-Benford law**
Presenter: **Andrea Cerioli**, University of Parma, Italy
Co-authors: Domenico Perrotta, Lucio Barabesi, Andrea Cerasa

The Newcomb-Benford law for digit sequences has recently attracted interest in anti-fraud analysis. However, most of its applications rely either on diagnostic checks of the data, or on informal decision rules. We suggest a new way of testing the Newcomb-Benford law that turns out to be particularly attractive for the detection of frauds in customs data collected from international trade. Our approach has two major advantages. The first one is that we control the rate of false rejections at each stage of the procedure, as required in anti-fraud applications. The second improvement is that our testing procedure leads to exact significance levels and does not rely on large-sample approximations.

EC1310: **Analysing large datasets with the forward search in SAS**
Presenter: **Francesca Torti**, European Commission, Italy
Co-authors: Marco Riani

The application of robust methods to international trade data may present serious scalability problems because the sample size n typically ranges from few tens to several hundreds of thousands units. This is particularly true for the Forward Search (FS), which needs to build a series of subsets of size increasing from few units (say v, i.e. the number of data variables) to n units. For this reason,we have implemented a SAS package for the FS that complements the official MATLAB implementation FSDA. We illustrate the main features of the new SAS FS package on a number of challenging international trade datasets provided by the customs services of some Member States. The illustration will also cover the typical interactive graphical tools for exploratory data analysis offered by FSDA.

EO1325: **Robustness for multilevel models: Fraud detection with the forward search**
Presenter: **Fabrizio Laurini**, University of Parma, Italy
Co-authors: Aldo Corbellini

Several methods using multiple regression or classification tools are commonly adopted to identify outliers which are, perhaps, the most important statistical units for anti-fraud detection. For data in the European Union, which are analysed, the presence of clusters of several firms and several countries, may hide structures and information, making standard and classical tools often unreliable. Moreover, even the parameters estimation of classical models can be severely biased by influential observations or outliers. A methodological solution is to exploit the natural hierarchical structure of multilevel models to take into account the time-varying evolution of quantities traded, and their price, for each country. Multilevel models, however, are not robust as they simply generalise linear models and ANOVA. A forward search algorithm is presented to make parameter estimation robust in the presence of outliers and avoiding masking and swamping, leading to a more accurate identification of suspicious firms. The influence of outliers, if any is inside the dataset, will be monitored at each step of the sequential procedure, which is the key element of the forward search. Preliminary results on simulated data have highlighted the benefit of adopting the forward search algorithm, which can reveal masked outliers, influential observations and show hidden structures. An application to real data is also illustrated.

EO1637: **Automatized proposals based on trimming and constraints for mixture modelling and clustering estimation**
Presenter: **Agustin Mayo-Iscar**, Universidad de Valladolid, Spain
Co-authors: Luis Angel Garcia-Escudero, Alfonso Gordaliza, Carlos Matran

Nowadays there is a huge increase of new sources of data that require automated tools for classifying them. It is expected that the corresponding data sets contain contaminated observations in addition to genuine observations belonging to real populations. Trimming and restrictions have shown efficacy for robustifying mixture modelling or clustering estimators designed for classifying data from different settings. Their joint application allows to avoid both the influence of contaminated observations and spurious components or clusters in the estimation. In order to illustrate the performance of these proposals, we apply procedures based on them to the automated analysis of the monthly import flows corresponding to different products in the European Union Trade Data.

EO058 Room MAL 402 CHANGE POINT ANALYSIS Chair: Pierre Bertrand

EO1294: **On parameter estimation and hyphothesis testing for Poisson processes in case of a change-point with variable jump size**
Presenter: **Serguei Dachian**, Universite Lille, France
Co-authors: Lin Yang

A model of Poissonian observations having a jump (change-point) in the intensity function is considered. Two cases are studied. The first one corresponds to the situation when the jump size converges to a non-zero limit, while in the second one the limit is zero. The limiting likelihood ratios in these two cases are quite different. The properties of the maximum likelihood and Bayesian estimators, as well as those of the general likelihood ratio, Wald's and Bayesian tests are deduced from the convergence of normalized likelihood ratios.

EO1461: **Model selection for multifractional Brownian motion**
Presenter: **Pierre Bertrand**, University Clermont-Ferrand, France
Co-authors: Marie Eliette Dury

The multifractional Brownian motion (mBm) can be viewed as a generalization of the fractional Brownian motion (fBm) where the Hurst index H is replaced by a time-varying function $H(t)$. A time-varying Hurst index is encountered in different kinds of applications: In quantitative finance for instance, it has been shown that the Hurst index estimated on sliding windows is varying with time between 0.45 and 0.65. Theoretical explanations are developed by economists. To sum up, arbitrage opportunity for fBm are possible when the Hurst index H is constant and known in advance, but no more when Hurst index is time-varying and random. Moreover, periods with Hurst index that significantly differs from $H = 1/2$, which corresponds to efficiency of the market, can be explained by behaviourial finance. For such a time-varying Hurst index, the methods of estimation developed up to now localize the estimation of Hurst index on a small vicinity, for models that become more and more sophisticated, e.g. the Hurst index being itself a stochastic process. Actually, we can not know whether fluctuations reflect reality or are just artifact of the statistics. Our aim is to provide a simplest possible model with a time-varying Hurst index. To sum up, the naive multifractional estimator has too many fluctuations that appear as a statistical artifact. Then it should be asymptotically rejected. Moreover we propose a way to choose the simplest possible function $H(t)$.

EO1490: **MUTATIS Software presentation**
Presenter: **Guillaume Paugam**, Blaise Pascal University, France
Co-authors: Pierre Bertrand

We created MUTATIS as a way of applying statistical methods of change point analysis to large datasets of time series. MUTATIS stands for Multi-Threaded Analysis of Time Series. This user-friendly piece of software was created to provide medical and academic personnel with our statistical tools for processing time series. It is written in Java in order to facilitate interoperability - i.e. platform-independence - and its transition towards an Android application. In particular, it was developed as part of the Do Well B. project, and can process large datasets of heartbeat time series, with a focus on ease-of-use and speed. It has the capability of cleaning the input files to remove artefacts, detect change points for any given criterion (mean, variance, slope, etc.), extract high and low frequency energies, and generate reports and graphics amongst other things. MUTATIS tries to find the balance between simplicity and efficiency, with a very clean interface and optional configuration files that can easily be tweaked and adjusted. The use of threads greatly decreases the processing times of the software, using up to 4 cores fully. Although MUTATIS was first created to address offline analysis, it is also now able to detect change points with an online algorithm, meaning the method can and likely will be used to process heartbeat series in real time via a Personal Health System.

EO1585: **Change point detection by filtered derivative with $p-$Value: Choice of the extra-parameters and the impact on MISE**
Presenter: **Doha Hadouni**, Blaise Pascal, France
Co-authors: Pierre Bertrand

The Filtered Derivative with $p-$Value method (FDpV) is a two-step procedure for change point analysis. In the first step, we use the Filtered Derivative function (FD) to select a set of potential change points, using its extra-parameters - namely the threshold for detection, and the sliding window size. In the second one, we calculate the $p-$value for each change point in order to only retain the true positives and discard the false positives. We deal with off-line change point detection using the FDpV method. We give a way to estimate the optimal extra-parameters of the function FD, in order to have the fewest possible false positives (false alarms) and non-detected change points (ND). Thus, the estimated potential change points may differ slightly from the theoretically correct ones. After setting the extra-parameters, we need to know which criterion (the absence of detection or the false alarm) has more impact on the Mean Integrated Square Error (MISE). Which leads us to calculate the MISE in both cases (false alarm case and the case of non detected change point). Finally, we simulate some examples with a Monte Carlo method so we can better understand the positive and negative ways the parametrisation can affect the results.

EC039 Room MAL 539 CONTRIBUTIONS ON REGRESSION ANALYSIS Chair: Juhyun Park

EC1441: **On the accuracy of conics and quadric surfaces parameters estimators**
Presenter: **Eva Fiserova**, Palacky University, Czech Republic

Fitting quadratic curves and quadric surfaces to given data points is a fundamental task in many fields like engineering, astronomy, physics, biology, quality control, image processing, etc. The classical approach for fitting is geometric fit based on minimization of geometric distances from observed data points to the fitted curve/surface. We focus on solving the problem of geometric fit by the linear regression model with nonlinear constraints. The constraints are represented by the general equation of the certain curve/surface. In order to obtain approximate linear regression model, this nonlinear constraints are being linearized by the first-order Taylor expansion. The iterative estimation procedure provides locally best linear unbiased estimates of the unknown algebraic parameters of the considered curve/surface and also estimates of their uncertainties. Subsequently, we are capable to express the estimates of the geometric parameters like the centre, angle of rotation, and semi-axes lengths and their uncertainties. The aim is to show the iterative algorithm for curve/surface fitting. Furthermore, the results of performed simulation study of the algebraic and geometric conics/surfaces parameters estimators accuracy will be presented.

EC1579: **Adaptive nonparametric instrumental regression in the presence of dependence**

Presenter: **Nicolas Asin**, Universite Catholique de Louvain, Belgium

Co-authors: Jan Johannes

We consider in nonparametric instrumental regression the estimation of the structural function, which models the dependence of a response Y on the variations of an endogenous explanatory variable Z in the presence of an instrument W. Given an i.i.d. sample of (Y,Z,W) a lower bound is derived for a maximal weighted mean integrated squared error assuming the structural function belongs to an ellipsoid linked to the conditional expectation operator of Z given W. We propose an estimator of the structural function based on a dimension reduction and an additional thresholding. Assuming either an i.i.d. sample of (Y,Z,W) or a sufficiently weak dependence characterized by fast decreasing mixing coefficients it is shown that the estimator is minimax rate-optimal considering smoothness assumptions. However, the proposed estimator requires an optimal choice of a dimension parameter depending on certain characteristics of the structural function and the conditional expectation operator of Z given W. As these are unknown in practice, we investigate a fully data-driven choice of the tuning parameter which combines model selection and Lepskis method. It is shown that the adaptive estimator with data-driven choice of the dimension parameter can attain the lower minimax risk bound up to a constant.

EC1451: **Tailor the estimator to data**

Presenter: **Jan Amos Visek**, Charles University in Prague, Czech Republic

The fact that the general proof of consistency of the Least Trimmed Squares appeared 20 years after the proposal of estimator indicated how demanding the studies of the its asymptotics were. It was due to 0 - 1 objective function. Hence attempts to remove the problem, the *S*-estimator (SE) and the *L*east Weighted Squares (LWS), depressed the influence of suspicious points smoothly, each by its way. They preserved the high breakdown point and the scale- and regression-equivariance and allowed easier study of their asymptotics. They coped with heteroscedasticity and increased the flexibility of estimator. Moreover, neither LWS is a special case of SE nor SE is a special case of LWS. Unifying them in the *S*-weighted estimators (SWE) we have hopefully removed cons while inherited pros from both, mainly allowing even for an unbounded objective function together with the weights tailored for given data. The asymptotics of SWE will be presented together with patterns of results of a simulation study performed by simple algorithm.

EC1560: **Jackknifed estimators for the negative binomial regression model**

Presenter: **Semra Turkan**, Hacettepe University, Turkey

Co-authors: Gamze Ozel

The negative binomial (NB) regression is very popular in applied researches when analyzing the count data. However, multicollinearity problem arises for the NB regression model when the independent variables are highly intercorrelated. Shrinkage estimator is a commonly applied solution to the general problem caused by multicollinearity. Recently, the ridge regression (RR) estimators and some methods for estimating the ridge parameter k in the NB regression have been proposed. The Jackknifed estimators are proposed to remedy the multicollinearity and reduce the bias. A simulation study is provided to evaluate the performance of estimators. Both mean squared error (MSE) and the percentage relative error (PRE) are considered as the performance criteria. The simulated result indicated that some of proposed Jackknifed estimators should be preferred to the ML method and ridge estimators to reduce MSE and bias.

Saturday 12.12.2015 14:25 - 16:05 Parallel Session E – CFE-CMStatistics

CO494 Room MAL B33 FORECASTING IN CENTRAL BANKS Chair: Francesco Ravazzolo

CO0210: **Forecasting process in Polish central bank**
Presenter: **Aleksandra Halka**, Narodowy Bank Polski, Poland
It is acknowledged that the monetary policy must be forward-looking, as there are substantial lags between the policy makers' actions and their impact on the economy. That is why central banks are focused on the forecasts of the key macroeconomic variables rather than on its current values. The aim is to disclose the features of the forecasting process and the related institutional framework at the central bank of Poland (Narodowy Bank Polski NBP). We present the range of models used for both medium- and short-term forecasting with the emphasis on the latter ones. We point out the role of the expert judgment in the forecasting process and the presentation of uncertainty surrounding the forecasts. We also raise the issue of incorporation of the short-term forecasts into the medium-term projections derived from both structural (NECMOD) and DSGE (SOE-PL) models. Therefore the final forecast may be viewed as the outcome of model and expert procedures, detailed considerations of short-term shocks, and medium-term trends. We summarize with the evaluation of the forecast's performance, both, between models used in the bank and against the outcomes of the external forecasts.

CO0650: **Nowcasting GDP with adaptive masking and ridge regression**
Presenter: **Jonas Hallgren**, Royal Institute of Technology KTH, Sweden
Co-authors: Johannes Siven, Erik Alpkvist, Ard den Reijer, Timo Koski
Nowcasting refers to methods for forecasting the current state of the economy and developments in the short term. As National Accounts statistics are published with a delay, more up-to-date indicators are used to determine the current level of GDP growth.The so-called ragged edge structure of macroeconomic data as formed by the availability of the indicators is important. A dynamic data masking approach is proposed, in which old data vintages are forced to dynamically adapt to the availability structure of the data at each nowcast generating moment. The adaptation is done by masking out data, such that the remaining data historically reflect the availability structure coherent with the current nowcast occasion. We then apply ridge regressions on the masked data to generate the nowcast. Our proposed approach thus employs both shrinkage and masking as a way of dealing with high dimensional data. We demonstrate in a backtest on macroeconomic data with real-time vintages that our method outperforms the dynamic factor model approach. Dynamic masking gives ease of implementation, a solid theoretical foundation, flexibility in modeling, and encouraging results; we view it therefore as a relevant addition to the nowcasting toolbox.

CO0867: **Norges Banks system for short-term forecasting of macroeconomic variables**
Presenter: **Anne Sofie Jore**, Norges Bank, Norway
Co-authors: Knut Are Aastveit, Francesco Ravazzolo, Claudia Foroni
Norges Banks short-term forecasts are based on a number of statistical and econometric models and judgment. A broad information set about the economic situation is part of the analysis. No single model can provide a complete description of reality. Different models have different properties. Theory and experience show that a weighted average of different model-based forecasts is often more accurate than forecasts provided by individual models. Norges Bank has therefore developed a system, SAM (System for Averaging Models), for averaging forecasts for inflation and mainland GDP provided by different models.

CO1063: **Short-term forecasting inflation and GDP at the Bank of Italy**
Presenter: **Antonio Conti**, Universite Libre de Bruxelles, Italy
We present some of the models employed at the Economic Outlook Unit to forecast the behavior of economic activity and inflation in Italy and in the Euro area. The range of models is pretty wide, covering bridge techniques, Bayesian VARs, Factor and Vector Error Correction Models. After a brief survey of these models, we show some examples of forecasting procedures, focusing on both the use of coincident indicators and the interaction with the ECB for the Narrow Inflation Projection Exercise.

CO390 Room MAL B34 MULTIVARIATE GARCH AND DYNAMIC CORRELATION MODELS Chair: Jean-David Fermanian

CO0238: **Dynamic correlation models based on vines: The model and applications**
Presenter: **Jean-David Fermanian**, Ensae-Crest, France
A new method for generating dynamics of conditional correlation matrices between asset returns is developed. These correlation matrices will be parameterized by their partial correlations, whose structure will be described by an oriented graph called "vine". Since partial correlation processes can be fitted and simulated separately, our approach is more parsimonious and flexible than other usual techniques, particularly the Dynamic Conditional Correlation family. We introduce a vine-GARCH class of processes and describe a quasi-maximum likelihood estimation procedure. We evaluate the performances of our models empirically and compare them with DCC-type specifications, through simulated experiments and by exploiting a database of daily stock returns.

CO1275: **Dynamic correlation models based on vines: Asymptotic theory**
Presenter: **Benjamin Poignard**, ENSAE - CREST, France
Co-authors: Jean-David Fermanian
The proposed vine-GARCH dynamics are estimated by the quasi-maximum likelihood method. We prove the consistency and asymptotic normality of the quasi-maximum likelihood estimator of the parameters obtained in a two-step procedure.

CO1286: **Macroeconomic variables, long run correlation, and optimal portfolio allocation the case of DCC-type models**
Presenter: **Matthias Fengler**, University of Sankt Gallen, Switzerland
Co-authors: Jean-David Fermanian, Hassan Malongo
We study the impact of macroeconomic variables on correlation dynamics. To this end, we suggest two DCC-type models: (i) a DCC model that includes exogenous variables in the long-term second-order moment matrix; (ii) a regime-switching DCC model with a time-varying transition probability matrix that depends on macro variables. Using different portfolio construction techniques, we assess the out-of-sample performance of the models and investigate the relevance of macro-variables for optimal portfolio allocation.

CO1347: **Macro-finance determinants of the long-run stockbond correlation: The DCC-MIDAS specification**
Presenter: **Hossein Asgharian**, Lund University, Sweden
Co-authors: Charlotte Christiansen, Ai Jun Hou
We investigate long-run stockbond correlation using a model that combines the dynamic conditional correlation model with the mixed-data sampling approach and allows long-run correlation to be affected by macro-finance factors (historical and forecasts). We use macro-finance factors related to inflation and interest rates, illiquidity, state of the economy, and market uncertainty. Macro-finance factors, particularly their forecasts, are good at forecasting long-run stockbond correlation. Supporting the flight-to-quality phenomenon, long-run correlation tends to be small and negative when the economy is weak.

CO637 Room MAL G15 NOWCASTING AND FORECASTING UNDER UNCERTAINTY II Chair: Gian Luigi Mazzi

CO0290: **Evaluating point and density forecasts from an estimated DSGE: The role of off-model information over the crisis**
Presenter: **Riccardo M Masolo**, Bank of England and Centre for Macroeconomics, United Kingdom
Co-authors: Matthew Waldron, Lena Koerber, Nicholas Fawcett
The purpose is to investigate the real-time forecast performance of the Bank of Englands main DSGE model, COMPASS, before, during and after the financial crisis with reference to statistical and judgemental benchmarks. A general finding is that COMPASS's relative forecast performance improves as the forecast horizon is extended (as does that of the statistical suite of forecasting models). The performance of forecasts from all three sources deteriorates substantially following the financial crisis. The deterioration is particularly marked for the DSGE models GDP forecasts. One possible explanation for that, and a key difference between DSGE models and judgemental forecasts, is that judgemental forecasts are implicitly conditioned on a broader information set, including faster-moving indicators that may be particularly informative when the state of the economy is evolving rapidly, as in periods of financial distress. Consistent with that interpretation, GDP forecasts from a version of the DSGE model augmentedto include a survey measure of short-term GDP growth expectations are competitive with the judgemental forecasts at all horizons in the post-crisis period. More generally, a key theme is that both the type of off-model information and the method used to apply it are key determinants of DSGE model forecast accuracy.

CO0310: **Comparing Alternative Methods of Combining Density Forecasts - with an Application to US inflation and GDP Growth**
Presenter: **James Mitchell**, University of Warwick, United Kingdom
Co-authors: Ana Galvao, Anthony Garratt
The performance of a wide range of alternative density forecast combination or pooling methods is explored via Monte Carlo and in an application forecasting US output growth and inflation using a large number of macroeconomic predictors.

CO0653: **A daily indicator of economic growth**
Presenter: **Claudia Foroni**, Norges Bank, Norway
Co-authors: Gian Luigi Mazzi, Fabrizio Venditti, Valentina Aprigliano, Massimiliano Marcellino
Alternative methods to construct a daily indicator of growth are studied. We aim for an indicator that (i) provides reliable predictions (ii) can be easily updated at the daily frequency (iii) gives interpretable signals. Using a large panel of daily and monthly data for the Euro area we explore the performance of three classes of models: bridge, MIDAS and UMIDAS models and different forecast combination strategies. Forecasts obtained from UMIDAS models, combined with inverse MSE weights, best satisfy the required criteria.

CO1014: **Density nowcasts of Euro area real GDP growth and pooling**
Presenter: **Marta Banbura**, European Central Bank, Germany
Co-authors: Lorena Saiz
In contrast to combination of point forecasts, the combination of density forecasts does not necessarily lead to better accuracy compared to individual densities. This is because the resulting combined density tends to have a higher dispersion, in particular in the linear pooling case. We review different strategies to combine individual density forecasts from a suite of models typically used for nowcasting the Euro area real GDP growth. The individual density now/forecasts are obtained by using Bayesian estimation combined with a simulation smoother. The individual densities are combined using different approaches, ranging from a general non-linear pooling (i.e. beta-transform linear pooling) to linear and logarithmic pooling, and with different weighting schemes: dynamic and static weights (e.g. optimal weights, equal weights, based on recursive log-score and continuous ranked probability score (CRPS)).

CO478 Room MAL 421 VOLATILITY MODELING AND CORRELATION IN FINANCIAL MARKETS Chair: Giampiero Gallo

CO0299: **Public news flow in intraday component models model for trading activity and volatility**
Presenter: **Adam Clements**, Queensland University of Technology, Australia
Co-authors: Joanne Fuller, Vasilios Papalexiou
There is a long history of literature examining the factors underlying the volatility of asset returns. Both news arrivals and trading activity have been identified as having significant impacts on volatility. Models for the intraday trading activity and volatility of equity returns are proposed. Utilising the Dynamic Conditional Score framework component models for intraday order flow and volatility are developed. In doing so, the role of overnight and intraday news arrivals on both order flow and volatility are examined, along with the link between order flow and volatility. It is found that along with order flow, overnight news arrivals (and their associated sentiment) are the significant driving factors of intraday volatility. Given models for order flow, forecast of volatility are generated conditional on the forecasts of order flow and overnight news. It is found that such an approach provides superior forecasts relative to a simple model for volatility.

CO0696: **Markov switching models with fuzzy regimes**
Presenter: **Giampiero Gallo**, Universita di Firenze, Italy
Co-authors: Edoardo Otranto
Realized volatility of financial time series is characterized by alternating turmoil and quiet periods, suggesting to consider changes in regime with different magnitude in the specification of the model. The question as of whether these changes should be abrupt or smooth remains open. In a recent work we have shown that the realized kernel volatility of the S&P500 index is well represented by modifications of the Asymmetric Multiplicative Error Models (AMEM) where a Markov Switching (MS)AMEM model with three regimes performs well in-sample, whereas a Smooth Transition (ST)AMEM seems to have the best performance out-of-sample. We combine now the two approaches, providing a new class of models where the parameters of a certain regime are subject to smooth transition changes. These models capture the possibility that regimes may overlap with one another (thus we label them fuzzy). We compare the performance of these models against the MS-AMEM and ST-AMEM, keeping the no-regime AMEM and HAR as benchmarks. The empirical application is carried out on the volatility of four US indices (S&P500, Russell2000, Dow Jones30, Nasdaq100): the superiority of the fuzzy regime models is established on the basis of several criteria and loss functions.

CO0849: **Modelling volatility dynamics of FX rates around macroeconomic news**
Presenter: **Fabrizio Lillo**, Scuola Normale Superiore, Italy
Co-authors: Vladimir Filimonov, Marcello Rambaldi
Financial time series are severely affected by the arrival of news and statistical modeling of volatility dynamics in such non-stationary periods is quite challenging. We propose an econometrical framework to model the tick-by-tick dynamics of mid-quote FX rates around scheduled macroeconomic announcements. We build on the Hawkes point process that combines an exogenous noise with an endogenous self-exciting mechanism due to the past price history. We extend this model with extra anticipation and reaction components to describe arrival of macro-economic releases. We show that the model is able to capture an increase of trading activity after (and even before) the news, both when the news has a sizable effect on volatility and when this effect is negligible. By considering the time of the news as a free parameter, our model can identify the exact time when the news affects the price dynamics, fully leveraging ultra high frequency data. This method can be used to detect and measure contagion

effects in the systems of interdependent securities (e.g. how a news on US economy affects the EUR/JPY rate) and to identify abrupt increases in high-frequency market activity, not necessarily triggered by news, but associated with abrupt price changes or with market manipulations.

CO1177: **Reduction and composite likelihood estimation of non-scalar multivariate volatility models**
Presenter: **Juan-Pablo Ortega**, CNRS Universite de Franche-Comte, France
Co-authors: Luc Bauwens, Lyudmila Grigoryeva
Multivariate volatility models are widely used in the description of the dynamics of time-varying asset correlations and covariances. Among the well-known drawbacks of many of these parametric families one can name the so called curse of dimensionality and the nonlinear parameter constraints that need to be imposed at the time of estimation and that are difficult to handle. We implement a composite likelihood (CL) estimation method for several non-scalar DCC and DVEC model specifications. The use of the CL approach motivates the in-depth study of different model reduction questions and the analysis of the closedness of the considered families under the reduction operation. The availability of both the QML and CL estimation tools makes possible the empirical out-of-sample performance comparison of the non-scalar DCC and DVEC models under study. We discuss an important estimation bias issue related to the use of covariance targeting and its impact on the empirical performance of the considered multivariate volatility models.

CO550 Room MAL 402 PORTFOLIO SELECTION AND ASSET PRICING AND MODELLING Chair: Abderrahim Taamouti

CO0327: **Parametric portfolio policies with common volatility dynamics**
Presenter: **Abderrahim Taamouti**, Durham University Business School, United Kingdom
A parametric portfolio policy function is considered that incorporates common stock volatility dynamics to optimally determine portfolio weights. Reducing dimension of the traditional portfolio selection problem significantly, only a number of policy parameters corresponding to first- and second-order characteristics are estimated based on a standard method-of-moments technique. The method, allowing for the calculation of portfolio weight and return statistics, is illustrated with an empirical application to 30 U.S. industries to study the economic activity before and after the recent financial crisis.

CO0328: **Transformed diffusions and copulas: Identification, inference and VIX derivative pricing**
Presenter: **Ruijun Bu**, University of Liverpool, United Kingdom
Co-authors: Kaddour Hadri, Dennis Kristensen
A new semiparametric approach is proposed for modelling continuous-time nonlinear univariate diffusions. In this approach, the observed process is assumed to be an unknown and unspecified transformation of an underlying parametric diffusion (UPD). This modelling strategy yields a general class of semiparametric copula-based diffusion models, where the dynamic copula is parametric and the marginal distribution is nonparametric. This new class of diffusions can generate rich nonlinear dynamics while at the same time admit closed-form transition densities. We provide primitive conditions for the identification of the UPD parameters together with the unknown transformation function from discrete samples. Root-n consistent semiparametric likelihood-based estimators of the UPD parameters as well as kernel-based drift and diffusion estimators are constructed, which are shown to be normally distributed in large samples. A simulation study investigates the finite sample performance of our estimators in the context of modelling US short-term interest rates. Closed-form formulae for pricing VIX futures and VIX options based on our modelling approach are also derived.

CO1024: **Term structure of forward moments and predictability of asset returns**
Presenter: **Dennis Philip**, Durham University, United Kingdom
Co-authors: Panayiotis Andreou, Anastasios Kagkadis, Abderrahim Taamouti
We rely on the recently established aggregation property of the second and third moments of returns to construct forward moments extracted from option prices. We show that according to standard affine no-arbitrage models, the forward moments should exhibit a factor structure, while the equity risk premium should also be an affine function of the same state variables. In light of this, we use dimensionality reduction techniques to extract, from the forward moments, the common factors that maximize the covariance between these factors and the equity premium. We show empirically that a small number of factors can explain the equity premium, both in-sample and out-of-sample, better than most traditional predictors. Moreover, we document that the inclusion of forward skewness into the analysis improves the asset return predictability and thus show that forward moments encapsulate important information about future market returns.

CC0605: **Bootstrapping reduced rank realized covariance matrices**
Presenter: **Julian Williams**, Durham University, United Kingdom
Co-authors: Abdullah Alshami, Prosper Dovonon, Abderrahim Taamouti
Estimating the ex-post quadratic variation of asset prices from high frequency data is an important tool in asset pricing. Recent results have derived the asymptotic and sample characteristics for i.i.d. bootstrapping of a bi-variate covariance matrix from high frequency returns and compute critical statistics for realized regressions. We demonstrate a bootstrap procedure for covariance matrices of dimension greater than two and implement a test for rank-deficiency. We then apply this technique to the problem of imputing multivariate hedging ratios from the cross section of futures prices measured at ultra-high (millisecond) time frequencies. We demonstrate the robustness of our approach to mild miss-specification and perform an in sample analysis of the hedging efficacy against the naive long run hedge and a Multi-variate GARCH hedge estimated at the daily frequency.

CO442 Room MAL B29 CORPORATE FINANCE Chair: Christodoulos Louca

CO0680: **The influence of national culture on the capital structure of SMEs**
Presenter: **Ioannis Tsalavoutas**, University of Glasgow, United Kingdom
Co-authors: Gillian Fairbairn, Darren Henry
Variations in entrepreneurial attitudes towards risk and control imply a link between the capital structure of SMEs and national culture. We investigate this unexplored relationship by using two of Schwartzs latest cultural dimensions (Hierarchy and Embeddedness) and a large panel data sample from seven countries, covering the period 2006 to 2008. Our results show that Hierarchy is negatively related to debt levels not only for the full sample, but also across the sub-samples of micro, small and medium firms. This suggests that managers who operate in cultures where wealth, social power, and authority are important cultural values use less debt. Embeddedness is also negatively related to debt levels of small and medium firms. This suggests that relatively-smaller companies in cultures which inter alia value family security and self-discipline tend to use less debt. Further testing shows that national culture can affect long-term, short-term debt, and the choice between the two differently. While the results for Hierarchy show a consistent, negative relationship between this dimension and both types of debt, the results for Embeddedness vary depending on the size of the firm and the duration of the debt source. Our findings contribute to prior literature by providing empirical evidence of national cultures influence on the capital structure of SMEs through the managers behaviour towards risk and control.

CO1018: **Terrorism, emotions, and corporate policies**
Presenter: **Constantinos Antoniou**, University of Warwick, United Kingdom
Co-authors: Alok Kumar, Anastasios Maligkris
The aim is to examine whether emotion-related biases induced by extraneous negative events affect corporate decision-making. Specifically, we conjecture that corporate managers located near major terrorist attacks will experience negative emotions, which would induce them to adopt more conservative corporate policies. Consistent with our conjecture, we demonstrate that firms located near terrorist events increase their cash holdings, and reduce their R&D expenditure and their long-term leverage around the events. These effects are temporary, become weaker as the distance between the firm and the event location increases, and are mainly concentrated among firms managed by younger CEO. Using multiple media proxies to capture the saliency of negative events, we also find that events with greater media exposure are associated with larger changes in corporate policies.

CO0985: **On stock returns and default risk: New international evidence**
Presenter: **Alexandros Kostakis**, Manchester Business School, United Kingdom
A unique dataset is constructed with bankruptcy filings for a large sample of non-U.S. firms in 14 developed markets which sheds new light on the cross-sectional relation between default risk and stock returns. Using a flexible approach to estimate default risk probabilities, we offer conclusive evidence supporting the existence of an economically and statistically significant positive default risk premium in international markets. This finding is robust to different portfolio weighting schemes, data filters, sample periods, and holding period definitions, and it holds using both in-sample estimates of default probabilities during the 1992-2010 period and out-of sample estimates during the 2000-2010 period. We also show that the magnitude of the default risk premium is contingent upon several firm characteristics.

CO1491: **CEO career horizon and stock price crashes**
Presenter: **Christodoulos Louca**, Cyprus University of Technology, Cyprus
Co-authors: Andreas Petrou, Panayiotis Andreou
Prior studies focus on the effects of firm-level factors that relate to stock price crashes and ignore CEO characteristics. The aim is to fill this gap in the literature by providing evidence that firms managed by CEOs with longer career horizon experience more stock price crashes, including crashes caused by revelation of negative news coming from breaks in positive earnings strings. Further, this relation is heightened in the presence of managerial discretion as captured by either CEO duality or a diversified organizational structure. Overall, the findings suggest that incentives to hoard negative information arise from CEO career horizon and create agency problems that lead subsequently to stock price crashes. In addition, they raise important considerations for managerial discretion, and more broadly corporate governance system, at various stages of CEO career horizon.

CO530 Room MAL 414 TOPICS IN TIME SERIES AND PANEL DATA ECONOMETRICS Chair: Martin Wagner

CO0728: **Cointegrating multivariate polynomial regressions: Fully modified OLS estimation and inference**
Presenter: **Oliver Stypka**, TU Dortmund, Germany
Co-authors: Martin Wagner
A fully modified OLS (FM-OLS) estimator is developed for cointegrating multivariate polynomial regressions, i.e., regressions that include as explanatory variables deterministic variables, integrated processes and products of non-negative integer powers of the integrated processes. The stationary errors are allowed to be serially correlated and the regressors are allowed to be endogenous. The FM-OLS estimator is extended from cointegrating polynomial regressions to cointegrating multivariate polynomial regressions, with the difference being the inclusion of cross-products of powers of the integrated processes, which overcomes the additive separability restriction typically used in nonlinear cointegration analysis for the polynomial case. The FM-OLS estimator has a zero-mean Gaussian mixture limiting distribution that allows for standard asymptotic inference. In addition to hypothesis testing on the parameters also Wald and LM specification tests are derived, as well as a KPSS-type test for cointegration. The theoretical analysis is complemented by a simulation study. Since the developed estimator immediately leads to a RESET-type specification test, we also compare in the simulation section the performance of our FM-OLS RESET test with the more restrictive RESET-type tests as well as the integrated modified OLS.

CO0930: **Large initial values and time series tests of the convergence hypothesis**
Presenter: **Michael Scholz**, University of Graz, Austria
Co-authors: Martin Wagner
Time-series based tests of the convergence hypothesis have been widely used in applied macroeconomics for over two decades. The occurrence of large initial conditions and their consequences for the asymptotic behavior of standard unit-root tests like the Augmented-Dickey-Fuller (ADF) test were ignored so far in the empirical growth literature. The ADF test is analyzed in the presence of large initial values and derives the asymptotic behavior in the stationary, the unit-root, and the near integrated case. In a finite sample simulation study we find that the limiting distributions with the initial value component are well approximated by the theoretical results but differ dramatically from standard ADF distributions. As a consequence, we observe severe size and power distortions of the standard ADF test causing to many rejections of the convergence hypothesis. The economic implications of these findings are demonstrated in an application using output series from the Penn World Tables over 1950-2011.

CO1226: **Integrated modified OLS estimation of spatially correlated cointegrated systems**
Presenter: **Leopold Soegner**, Institute for Advanced Studies, Austria
Co-authors: Timothy Vogelsang, Martin Wagner
We consider a panel of cointegrating regressions linked by a spatial lag. As is usual in the cointegration literature we allow for regressor endogeneity, error serial correlation and the presence of deterministic trend components. In addition we allow for cross-sectional dependence via two channels: First, we allow the errors to be cross-sectionally correlated. Second, of course, cross-sectional dependence occurs through a spatial lag of the dependent variables entering the cointegrating relationships. The integrated modified OLS estimator is extended from the single equation setting to our setting of cross-sectionally spatially correlated cointegrating regressions. We also consider fixed-b inference in this setting.

CO1427: **Estimator averaging for improving efficiency**
Presenter: **Hanno Reuvers**, Maastricht University, Netherlands
There are usually several estimators available to estimate the parameters of a statistical model. The requirement of these parameter estimates to be consistent does not limit the possibilities much further. Frequently, the next criterion to discriminate between the estimators is their efficiency, say the mean squared error. An important question to ask is whether there is an estimator which dominates the others in terms of efficiency, regardless of the parameters in the true DGP and regardless of the distributional assumptions on the error term. If such an estimator does not exist, then one might consider using several estimators and combining them into a final estimate. We suggest a weighted estimator in which the weights are chosen based on the minimization of the asymptotic mean squared error. As possible applications of our method we discuss estimator averaging in regression models and (vector) autoregressions. Monte Carlo simulations show that the resulting estimator outperforms the single estimates in large areas of the parameter space.

CO588 Room MAL B20 NEW BAYESIAN METHODS AND ECONOMETRIC APPLICATIONS Chair: Michael Smith

CO0917: **The Poisson gamma belief network**
Presenter: **Mingyuan Zhou**, University of Texas at Austin, United States
Co-authors: Yulai Cong, Bo Chen
A key issue in deep learning is to define an appropriate network structure, including both the depth of the network and the width of each hidden layer, which may be naturally addressed with completely random measures. We propose the Poisson gamma belief network (PGBN), which factorizes each of its layers into the product of a connection weight matrix and the nonnegative real hidden units of the next layer, to infer a multilayer representation of high-dimensional count vectors. We use the gamma-negative binomial process together with a layer-wise training strategy to infer the network structure, allowing the width of each hidden layer to grow without bound. We further demonstrate that with a fixed budget on the width of the first layer, the PGBN can increase its number of hidden layers to boost its performance. We propose an efficient upward-downward Gibbs sampler to jointly train all the hidden layers of the PGBN, with example results on text analysis illustrating its efficacy and unique properties.

CO0597: **Shape-constrained nonparametric high-dimensional function estimation using Bayesian additive regression trees**
Presenter: **Tom Shively**, University of Texas at Austin, United States
Co-authors: Hugh Chipman, Edward George, Robert McCulloch
Bayesian additive regression tree (BART) models are a flexible method for nonparametrically estimating a high-dimensional function. We show how to extend BART models to incorporate shape constraints such as monotonicity. Such constraints arise naturally in many disciplines. For example, economic theory states that demand for a product is a monotonically decreasing function of its own price and a monotonically increasing function of its competitors prices. The imposition of shape constraints often results in much better function estimates than unconstrained methods provide. The stochastic search method we use to find promising BART models is different from the one used for unconstrained models because it is no longer possible to integrate out the function values at each node in a constrained model. In particular, decisions to split or collapse terminal nodes are made by conditioning on the function values at all other nodes. Also, new terminal node function values are generated individually or in pairs subject to the shape constraint. We show via simulation that for a wide range of high-dimensional functions the resulting constrained function estimates have good properties and are often a considerable improvement over unconstrained estimates.

CO0850: **Dynamic quantile function models**
Presenter: **Richard Gerlach**, University of Sydney, Australia
Co-authors: Wilson Ye Chen, Gareth Peters, Scott Sisson
Modelling the time-varying distributions of financial returns has been an interest to many authors in recent decades. The increasing availability of high-frequency data has presented new challenges, namely, effectively making use of the noisy information contained in the intra-daily observations at a reasonable computational cost. Borrowing ideas from symbolic data analysis (SDA), we consider data representations beyond scalars and vectors. Specifically, we consider a quantile function as an observation, and propose a class of dynamic models for quantile-function-valued time series. Direct modelling of quantile functions can be more convenient in applications where the quantity of interest is a quantile. We present a method whereby a likelihood function can be defined for quantile function-valued data, and develop an MCMC algorithm for simulating from the posterior distribution. In the empirical study, we model the time series of quantile functions of high frequency financial returns, and demostrate the usefulness of our method by forecasting one-step-ahead the extreme quantiles of intra-daily returns. Furthermore, through a simple empirical scaling rule, we are able to forecast one-step-ahead the Value-at-Risk of daily returns.

CO0846: **Econometric modeling of regional electricity spot prices in the Australian market**
Presenter: **Michael Smith**, University of Melbourne, Australia
Wholesale electricity markets are increasingly integrated via high voltage interconnectors, and inter-regional trade in electricity is growing. To model this, we employ a spatial equilibrium model of price formation, where constraints on inter-regional flows result in three distinct equilibria in prices. We use this to motivate an econometric model for the distribution of observed electricity spot prices that captures many of their unique empirical characteristics. The econometric model features supply and inter-regional trade cost functions, which are estimated using Bayesian monotonic regression smoothing methodology. A copula multivariate time series model is employed to capture additional dependence in regional prices. The model has the advantage of preserving the heavy right-hand tail in the predictive densities of price. We fit the model to hourly spot price data in the five interconnected regions of the Australian national electricity market. The fitted model is then used to measure how both supply and price shocks in one region are transmitted to the distribution of prices in all regions in subsequent periods.

CO444 Room MAL B36 APPLIED FINANCIAL ECONOMETRICS Chair: Richard Luger

CO0958: **Linear two-stage-least-squares estimators for GARCH processes**
Presenter: **Todd Prono**, American University, United States
It is shown how standard, two stage least squares (TSLS) techniques apply to GARCH(p,1) model estimation. The instruments are first-order lags, which are valid in the usual sense so long as the series being modeled displays skewness and that skewness does not link to a leverage effect in the conditional variance. These TSLS estimators are shown to be strongly consistent and to weakly converge in distribution to stable, though infinite variance, mixtures of Poisson and point processes. Both consistency and distributional convergence only require (slightly stronger than) third moment existence; a theoretical requirement that is both in line with empirical findings that support well-defined third moments of many financial time series like equity and FX returns and is decidedly weaker than the necessary moment-existence criterion underlying standard root-T asymptotics. A Monte Carlo study is conducted in order to assess, for instance, the optimal size of the instrument vector, recognizing that many instruments are available. The estimators are also applied to intra-day spot returns on the Japanese Yen, Euro, and Swiss Franc in order to gauge both in-sample fit as well as out-of-sample forecasting power. Benchmarking these results is the familiar (and popular) quasi maximum likelihood estimator.

CO0973: **The macroeconomic drivers of bond market risk**
Presenter: **Gregory Bauer**, Bank of Canada, Canada
We use the matrix-logarithm model of the covariance matrix to estimate time variation in the realized quarterly covariance matric of bond market volatility. The matrix is generated from a new data set of daily Government of Canada bond yields starting in 1936, which, to the best of our knowledge, is the longest daily bond series available. The model uses latent factors which are functions of both lagged volatility and macroeconomic fundamentals. We explore how time variation in the structure of the Phillips curve drives different risk factors.

CO0928: **Affine term structure modes with observable volatility**
Presenter: **Antonio Diez de los Rios**, Bank of Canada, Canada
A new class of tractable affine term structure models is introduced in which (i) bond yields are driven by a low dimensional structure, (ii) bond yields display conditional heteroskedasticity, (iii) bond market volatility and bond yields can move independently (unspanned volatility), and (iv) we

allow for the covariance of the factors to change signs. In this model, the pricing factors are observable portfolios of yields and their corresponding realized covariances which allow us to avoid non-linear filtering methods, thus greatly easing its estimation.

CO1083: **Markov-switching quantile autoregression: A Gibbs sampling approach**
Presenter: **Richard Luger**, Laval University, Canada
Co-authors: Xiaochun Liu
We extend the class of linear quantile autoregression models by allowing for Markov-switching effects in the location of the conditional quantiles. We also propose a Gibbs sampling algorithm for posterior inference by using data augmentation and a location-scale mixture representation of the asymmetric Laplace distribution. An estimate of the marginal likelihood is available as a by-product of the procedure, since all complete conditional densities used in the Gibbs sampler have closed-form expressions. We use the marginal likelihood estimates across different probability levels to determine the order of a stepwise re-estimation procedure which solves the well-known quantile crossing problem. Our method to enforce non-crossing quantiles can be applied in any quantile regression model with endogenous or exogenous covariates, and whether Markov-switching effects are allowed for or not.

CO514 Room MAL B35 ECONOMETRIC CHALLENGES IN RISK MANAGEMENT Chair: Laura Spierdijk

CO0407: **Chain ladders with incomplete observations**
Presenter: **Ruud Koning**, University of Groningen, Netherlands
Chain ladders have been used to determine reserves for paying claims that take a long time to process. How much money do we need to reserve for next year to cover losses incurred in this year? In recent literature, payments patterns are considered to be random, and relevant models have been formulated as Generalised Linear Models. Development patterns have been modelled by a negative binomial distribution. However, usually datasets used to estimate the parameters of these GLM distributions are small, and partially incomplete. It may be helpful to incorporate covariates to model the development pattern at the claim level, and determine reserves by aggregating over claims. This approach is not directly applicable if some of the observations are incomplete. We propose an extension to this approach that does allow for incompletely observed covariates. Also, we discuss some methodology needed to estimate the statistical uncertainty associated with this approach.

CO0433: **Dynamic estimation of substitution elasticies with an application to US banks**
Presenter: **Laura Spierdijk**, University of Groningen, Netherlands
Co-authors: Laura Spierdijk, Sherrill Shaffer, Tim Considine
Substitution elasticities quantify the extent to which the demand for inputs responds to changes in input prices. They are considered particularly relevant from the perspective of cost and risk management. Changes in firms' substitution elasticities have typically been associated with behavioral shifts in response to economic and regulatory changes. We show how a dynamic logit demand (DLD) system can be used to estimate short-run and long-run substitution elasticities, as well as the speed at which input price changes are incorporated in the demand for inputs (known as the lag time). In contrast to the existing literature, our approach does not make a priori assumptions about the variable or quasi- fixed nature of input factors. We illustrate the approach in an empirical application that analyzes US commercial banks' substitution elasticities and lag times before and after the onset of the crisis. We conclude that a dynamic approach is crucial to accurately capture the drop in input factor substitutability after the crisis.

CO0638: **Micro-level stochastic loss reserving models fortime-discrete data**
Presenter: **Katrien Antonio**, University of Amsterdam and KU Leuven, Netherlands
Co-authors: Els Godecharle, Robin Van Oirbeek
The focus is on the development of future cash flows for outstanding insurance claims. We develop a model on a claim-by-claim basis using time discrete data, leading to the construction of so-called micro-level stochastic loss reserving models. To this end, a multiple state framework is proposed, such that the claim development process can be reconstructed as a series of transitions between a given set of states. The transition probabilities between the latter states are modeled by means of aproportional odds model, hereby allowing the transition probabilities to depend on, if necessary, time-varying covariates. For each payment in the claim development process, we model its claim size distribution using an approach based on splicing: the respective claim size distribution is divided in different slices and by means of a multinomial distribution the probability is determined that the payment corresponding to a certain claim pertains to the slice of interest. The different slices themselves are modeled by means of truncated gamlss models and the best fitting truncated distribution is determined from the large library of the gamlss error distributions by means of the Akaike Information Criterion (AIC). The multinomial distribution and the truncated gamlss distributions of all slices are all allowed to depend on time-varying covariates.

CO1580: **Term structure extrapolation and asymptotic forward rates**
Presenter: **Michel Vellekoop**, University of Amsterdam, Netherlands
We investigate different inter- and extrapolation methods for term structures under different constraints in order to generate market-consistent estimates which describe the asymptotic behavior of forward rates. Our starting point is the method proposed by Smith and Wilson, which is used by the European insurance supervisor EIOPA to generate forward curves which converge to an a priori given constant. We use the characterization of the Smith-Wilson class of interpolating functions as the solution to a functional optimization problem to extend their approach in such a way that forward rates will converge to a value which is not given a priori, but an outcome of the optimization process. Precise conditions are stated which guarantee that the optimization problems involved are well-posed on appropriately chosen function spaces. As a result, a well- defined optimal asymptotic forward rate can be derived directly from prices and cashflows of traded instruments. This allows practitioners to use raw market data to extract information about long term forward rates, as we will show in a study which analyzes historical EURIBOR swap data.

CG017 Room MAL 539 CONTRIBUTIONS ON BAYESIAN METHODS IN ECONOMICS AND FINANCE Chair: Jonathan Stroud

CC1426: **A statistical approach to sequential portfolio optimization with multi-step forecasting**
Presenter: **Kaoru Irie**, Duke University, United States
Co-authors: Mike West
In the context of sequential forecasting and portfolio optimization, we introduce a novel approach to Bayesian analysis based on mapping a specified loss function minimization problem to that of finding the mode of a posterior distribution in a synthetic statistical model. Computational methods for exploring distributions can then be applied to solve the original optimization problem. We do this in the context of novel portfolio utility functions that extend traditional Markowitz-type methods to multiple-step ahead investments with explicit penalties for transaction costs. Various forms introduced and explored include sparsity-inducing penalties on portfolio turnover and asymmetric penalties on the deviation from the target return, which yield interesting classes of synthetic statistical models of state-space forms with non-Gaussian structure. The resulting computational problems are addressed using combinations of EM methods, MCMC and analytic filtering and smoothing. Significant practical benefits in application to financial portfolio are realized in applied studies of FX, commodity and stock index time series, based on sequential forecasting using customized dynamic dependency network models.

CC1519: **Learning about informed trading risk through the dynamic volume - spread relationship**
Presenter: **Samuel Dua Oduro**, University of Kent, United Kingdom
Co-authors: Jim Griffin, Jaideep Oberoi
Illiquidity caused by some investors trading on their information advantage (informed trading) has been a subject of market microstructure research in the last few decades. Researchers have employed information-based models that use observed or inferred order flow to investigate this problem. However, these studies provide static estimates of the effect of informed trading on liquidity. Varied evidence of relationship between volume, bid-ask spread and volatility has been documented in the extant literature. In particular, theory suggests that bid-ask spread and volume are jointly driven by investors' valuation of the asset. The complex relationship between these variables are time variant, since the informed trading component of order flow changes as trading takes place. Thus, volume and bid-ask spread may provide insight on the time-varying composition of economic agents trading an asset. By modelling the joint relationship between volume and bid-ask spread dynamically in a state space model, we decompose volume into two components with corresponding effects on bid-ask spread. The structure of the model and estimation methodology enhances the sequential processing and incorporation of past volume and bid-ask spread as conditioning information. Results obtained from the model applied to two NYSE traded assets, IBM and Ashland Oil are presented.

CC1658: **IMF programs and sensitivity to external shocks: A Bayesian VAR approach**
Presenter: **Mirela Sorina Miescu**, Queen Mary University of London, United Kingdom
The aim is to assess whether participation into IMF programs significantly diminishes the vulnerability to external shocks. Nowadays, one of the primary purposes of the IMF is to insure global stability, giving advice to member countries on the financial and economic policies that promote stability, help avoiding crises and smooth the adjustment to external shocks. Employing a Bayesian VAR model to obtain a measure for the exposure to external shocks, we close our analysis with an Instrumental Variable approach and we show that participation in the IMF arrangements significantly diminishes the sensitivity to exogenous shocks. Despite the many critics concerning the effects of the IMF loans on the recipient country economy, our results provide evidence that the the Fund is efficient in helping member countries to build a strong economic resilience. These results are of considerable interest in a context where shocks and crises have become a systemic feature of the global economy affecting both developed and developing countries.

CC1443: **Forecasting performance of Bayesian DCS models using asymmetric conditional distributions**
Presenter: **Blazej Mazur**, Cracow University of Economics, Poland
The objective is to examine predictive performance of univariate Bayesian dynamic conditional score models (or generalized autoregressive score models) based on conditional sampling distributions allowing for flexible asymmetry modelling. Distributions with at least two parameters that control asymmetry are considered flexible, implying potentially different asymmetry pattern around mode and in tails. The distributions are obtained by the two-piece method, including $t-$distribution with different degrees of freedom in each tail. DCS-type dynamics is often assumed for conditional location or dispersion parameters, a potential gain from considering time-varying conditional asymmetry is considered. It implies time-varying asymmetry of the predictive distribution. DCS-type specifications offer numerical advantages over models with latent processes within recursive prediction setup. Bayesian inference is used since it does not require parametric restrictions ensuring asymptotic normality of the ML estimator. Empirical applications involve monthly/quarterly macroeconomic data (inflation) and daily series of financial returns. Out-of-sample (density) predictive properties of different specifications are compared using log-score and CRPS criteria, PITs are informally examined. Differences regarding inference on deflation probability (macro) and risk measures (VaR and ES for financial data) are discussed.

CG375 Room MAL 415 CONTRIBUTIONS ON NON-STARIONARITY AND NON-LINEARITY Chair: Bent Nielsen

CC1510: **Cumulated sum of squares statistics for non-linear and non-stationary regressions**
Presenter: **Vanessa Berenguer Rico**, University of Oxford, United Kingdom
Co-authors: Bent Nielsen
We show that the cumulated sum of squares test has a standard Brownian bridge-type asymptotic distribution in non-linear regression models with non-stationary regressors. This contrasts with cumulated sum tests which have been studied previously and where the asymptotic distribution involves nuisance quantities. Through simulation we show that the power is comparable in a wide range of situations.

CC1098: **Singular spectrum analysis based neural network for crude oil price forecasting**
Presenter: **Milan Csaba Badics**, Corvinus University of Budapest, Hungary
Prediction of crude oil price movement is regarded as one of the most challenging tasks of time series forecasting since oil prices are noisy, non-stationary and nonlinear. Due to the fact that crude oil price series are affected by a variety of factors, traditional statistics-based forecasting approaches fail to produce good prediction performance. For solving these problems, a new composite model forecasting the world crude oil spot price value, based on the combination of a singular spectrum analysis (SSA) and a standard neural network (ANN), is presented. For this purpose, the original crude oil spot price series was first decomposed into terms of the trend, the market fluctuations and the noise. Then a three-layer feedforward neural network was developed to forecast these components (except noise) 1 day ahead. Finally the final forecast results of crude oil price were calculated with summation of each forecast components. The proposed approach first utilizes the advantage of SSA for crude oil price time series. For verification purpose, two main crude oil price series, WTI and Brent, were used to test the effectiveness of the proposed methodology. Empirical results demonstrate that the SSA-ANN method has a strong forecasting capability for crude oil prices and outweighs the forecasts in accuracy of both the EMD-ANN and the ANN model with non-filtered forecasting variables in terms of mean square errors (MSE) and directional accuracy (DA).

CC1475: **Nonlinear leverage effects in asset returns: Evidence from the U.S. and Japanese stock markets**
Presenter: **Teruo Nakatsuma**, Keio University, Japan
Co-authors: Kenichiro McAlinn, Asahi Ushio
The leverage effect in asset returns is the main focus. In the literature of financial econometrics, the term leverage effect refers to a stylized fact that a decline in the asset value will tend to increase the future volatility of the asset return. In the context of the stochastic volatility (SV) model, the leverage effect is often expressed as a negative linear relationship between the present asset return and the future log volatility. We propose to relax this linearity by approximating a nonlinear relationship between the asset return and the log volatility with a Hermite polynomial. We also developed a particle filtering method to estimate the proposed SV model with nonlinear leverage effects. To examine whether any nonlinear leverage effects exist in the stock markets, we estimated SV models with nonlinear leverage effects for stocks included in S&P 500, the NASDAQ composite index or the Nikkei 225 index. We chose the most appropriate order of the Hermite polynomial for each stock with the marginal likelihood. The empirical results strongly support the nonlinearity of the leverage effect for many stocks. They also suggest that the orders of Hermite polynomials possibly differ across markets as well as segments of stocks.

CC1290: **A tractable framework for analyzing a class of nonstationary Markov models**
Presenter: **Inna Tsener**, University of the Balearic Islands, Spain
Co-authors: Serguei Maliar, John Taylor , Lilia Maliar
We study a class of infinite-horizon nonlinear dynamic economic models in which preferences, technology and laws of motion for exogenous variables can change over time either deterministically or stochastically, according to a Markov process with time-varying transition probabilities, or both. The studied models are nonstationary in the sense that the decision and value functions are time-dependent, and they cannot be generally solved by conventional solution methods. We introduce a quantitative framework, called extended function path (EFP), for calibrating, solving, simulating and estimating such models. We apply EFP to analyze a collection of challenging applications that do not admit stationary Markov equilibria, including growth models with anticipated parameters shifts and drifts, unbalanced growth under capital augmenting technological progress, anticipated regime switches, deterministically time-varying volatility and seasonal fluctuations. Also, we show an example of estimation and calibration of parameters in an unbalanced growth model using data on the U.S. economy. Examples of MATLAB code are provided.

EO160 Room Court BAYESIAN ANALYSIS OF MISSING DATA AND LONGITUDINAL DATA Chair: Jaeyong Lee

EO0200: **Bayesian pattern mixture models for the analysis of repeated attempt designs**
Presenter: **Michael Daniels**, University of Texas at Austin, United States
It is not uncommon in follow-up studies to make multiple attempts to collect a measurement after baseline. Recording whether these attempts are successful or not provides useful information for the purposes of assessing the missing at random (MAR) assumption and facilitating missing not at random (MNAR) modeling. This is because measurements from subjects who provide this data after multiple failed attempts may differ from those who provide the measurement after fewer attempts. This type of 'continuum of resistance' to providing a measurement has hitherto been modeled in a selection model framework, where the outcome data is modeled jointly with the success or failure of the attempts given these outcomes. We present a Bayesian pattern mixture approach to model this type of data. We re-analyze the repeated attempt data from a trial that was previously analyzed using a selection model approach. Our pattern mixture model is more flexible and transparent than the models that have previously been used to model repeated attempt data and allows for sensitivity analysis and informative priors.

EO1243: **Bayesian latent pattern mixture models for handling attrition in panel studies with refreshment samples**
Presenter: **Yajuan Si**, University of Wisconsin-Madison, United States
Co-authors: Jerome Reiter, Sunshine Hillygus
Many panel studies collect refreshment samples-new, randomly sampled respondents who complete the questionnaire at the same time as a subsequent wave of the panel. With appropriate modeling, these samples can be leveraged to correct inferences for biases caused by non-ignorable attrition. We present such a model when the panel includes many categorical survey variables. The model relies on a Bayesian latent pattern mixture model, in which an indicator for attrition and the survey variables are modeled jointly via a latent class model. We allow the multinomial probabilities within classes to depend on the attrition indicator, which offers additional flexibility over standard applications of latent class models. We present results of simulation studies that illustrate the benefits of this flexibility. We apply the model to correct attrition bias in an analysis of data from the 2007-2008 Associated Press/Yahoo News election panel study.

EO1271: **A sensitivity analysis model for longitudinal data with nonignorable intermittent missingness**
Presenter: **Chenguang Wang**, Johns Hopkins University, United States
Missingness in longitudinal studies could be monotone (dropout) or intermittent. Despite the broad discussion of monotone missing data analysis methods, handling non-ignorable intermittent missing data is methodologically and computationally challenging and, as a result, statistics literature is relatively limited. Typical multiple imputation approaches do not assume that the imputation model is compatible with the inference model, which has been termed "uncongenial". We propose an innovative approach to first model the observed response and missingness data with Bayesian nonparametric models. We then introduce sensitivity parameters to identify the full data model. As an example, we apply the proposed approach to a sleep drug clinical trial.

EO0151: **Analysis of longitudinal ordinal data with general random effects covariance matrix**
Presenter: **Keunbaik Lee**, Sungkyunkwan University, Korea, South
Co-authors: Jiyeong Kim
To analyze the longitudinal categorical data, we typically use generalized linear mixed models (GLMMs). In the models, the random effects covariance matrix is used to demonstrate both the subject-specific and time variations, and the covariance matrix may also be heterogenous. However, the structure of the covariance matrix is assumed to be homogeneous and restricted because of the high dimension and the positive definiteness of the matrix. To release these assumptions two Cholesky decomposition methods were proposed in linear mixed models: the modified Cholesky (Pourahmadi, 1999) and the moving average Cholesky (Zhang and Leng, 2012) decompositions. In this paper we propose a cumulative logit random effects model with heterogeneous random effects covariance matrix for longitudinal ordinal data. We also exploit the two decompositions to model the random effects covariance matrix, and compare estimated parameters using the two decompositions. Methods are illustrated with a lung cancer data set.

EO100 Room Senate ADVANCES IN ROBUST DATA ANALYSIS Chair: Alfonso Gordaliza

EO0227: **A spatial influence function and local outliers: Applications with geographical information systems**
Presenter: **Alfonso Garcia-Perez**, UNED, Spain
Co-authors: Yolanda Cabrero-Ortega
An outlier is usually considered as an observation well separated from the majority of the data. If we add the spatial coordinates to the observations we show in a couple of examples that apparent outliers are not outliers in fact and, conversely, outliers appear where, apparently, there were none. After motivating the need of using geographical coordinates of the observations, we define a Spatial Influence Function as a very useful tool to detect local (spatial) outliers and study its main properties. The definition involves, in practice, a classical GAM fit. Local outliers are observations that differ from their neighbors. We consider that these differences must be understood with respect to a robust GAM, i.e., the difference between the observed value and the predicted one, indicates the extent to which the observed data is consistent with its neighbor. Georeferenced residuals will determine areas of local outliers. The exploration of these zones in finding local outliers will be done with the help of QGIS, a Geographical Information System that interacts with R. We apply these ideas to several examples.

EO1002: **Highly robust and efficient negative binomial regression**
Presenter: **Alfio Marazzi**, University of Lausanne, Switzerland
Co-authors: Michael Amiguet, Victor Yohai
We consider the regression model $Y|x \sim F_{\mu(x),\alpha}$, where $F_{\mu,\alpha}$ is the negative binomial distribution with $\mu(x) = E(Y|x) = h(\beta^{\mathrm{T}}x)$ and dispersion α; x is a covariate vector, β a vector of coefficients, h a link function. Given a sample (y_i, x_i), we are looking for highly robust and efficient estimators of β and α based on three steps. In the first step the maximum rank correlation estimator is used to consistently estimate the slopes up to a scale factor.

The scale factor, the intercept, and the dispersion parameter are consistently estimated using a combination of M-estimates. In the second step, outliers are identified observing that, if u_i is uniformly distributed, then the tail probabilities $F_{\mu(x_i),\alpha}(y_i) - u_i f_{\mu(x_i),\alpha}(y_i)$ are uniformly distributed ($f_{\mu,\alpha}$ is the density of $F_{\mu,\alpha}$). For the final step, an adaptation of the adaptively truncated maximum likelihood regression estimate is used.

EO1203: **Robust switching models and fuzzy clustering: A robust approach based on trimming**
Presenter: **Francesco Dotto**, Sapienza - University of Rome, Italy
Co-authors: Alessio Farcomeni, Luis Angel Garcia-Escudero, Agustin Mayo-Iscar
We propose a robust switching regression model based on fuzzy clustering. We efficiently estimate the underlying linear relationships between a dependent variable and a set of predictors in each cluster. Estimation is based on an iterative trimming procedure. During each iteration we trim the observations with largest current regression error, and the remaining are used to estimate the parameters of the linear model within each cluster. Fuzzy weights computation is based on the current cluster-specific residuals. A constraint on the relative variability of the residuals protects against the occurrence of spurious maximizers of the objective function. The procedure is robust to Tukey-Huber contamination and the use of fuzzy weights allows to recover the underlying linear structures even with moderate cluster overlap.

EO1676: **Tools for simulating and contaminating clustering data**
Presenter: **Domenico Perrotta**, EC - Joint Research Centre, Italy
Co-authors: Marco Riani, Andrea Cerioli, Francesca Torti
Flexible generation methods for data, outliers and noise are necessary to evaluate the performance of clustering algorithms. MixSim is a framework which simulates data from normal mixture distributions on the basis of pre-specified synthesis statistics on an overlap measure, defined as sum of pairwise misclassification probabilities. We have extended MixSim to control additional overlapping statistics and departures from homogeneity and sphericity among groups. The output is a more flexible data generation framework, to better address modern robust clustering scenarios. We have studied the properties and the implications that this new way of simulating clustering data entails in terms of coverage of space, goodness of fit to theoretical distributions, and degree of convergence to nominal values. Everything is addressed in the multivariate and regression contexts, the latter being an extension that we made to also simulate mixtures of regression structures. The traditional and new MixSim features can be tested using our MATLAB implementation of the framework, perfectly integrated in our FSDA toolbox together with state of the art robust clustering algorithms and principled routines for their evaluation and calibration. A spin off of our work is a general complex routine, ported from C language to MATLAB, to compute the distribution function of a mixture of non central χ^2 random variables which is at the core of MixSim and has its own interest for many test statistics.

EO048 Room CLO 306 STATISTICS FOR GENOMICS Chair: Christophe Ambroise

EO0270: **DINGO: Differential network analysis in genomics**
Presenter: **Kim-Anh Do**, University of Texas MD Anderson Cancer Center, United States
The aim is to propose a differential network analysis in genomics (DINGO) model for estimating group-specific networks as well as making inference on the differential networks. DINGO jointly estimates the group-specific conditional dependencies by decomposing them into a global and group-specific components. The delineation of these components allows a more refined picture of major driver and passenger events in the elucidation of cancer progression and development. We highlight the performance of this model through simulations and application to The Cancer Genome Atlas (TCGA) glioblastoma data.

EO0714: **Transcriptional landscape reconstruction from high-throughput sequencing count data via state space models**
Presenter: **Pierre Nicolas**, INRA, France
Co-authors: Bogdan Mirauta, Hugues Richard
The most common RNA-Seq strategy consists of random shearing, amplification, and high-throughput sequencing, of the RNA fraction. Methods to analyze transcription level variations along the genome from the read count profiles generated by this global RNA-Seq protocol are needed. We developed statistical approaches to estimate the local transcription levels and to identify transcript borders along the genome. Our transcriptional landscape reconstruction relies on a state-space model to describe transcription level variations in terms of abrupt shifts and more progressive drifts. A new emission model is introduced to capture not only the read count variance inside a transcript but also its short-range autocorrelation and the fraction of positions with zero-counts. The estimation relies on a Markov-Chain Monte Carlo approach involving a Sequential Monte Carlo algorithm, the Particle Gibbs.

EO0734: **Bayesian variable selection for binary outcomes in high dimensional settings**
Presenter: **Wenyi Wang**, The University of Texas MD Anderson Cancer Center, United States
Co-authors: Amir Nikooienejad, Valen Johnson
One of the important areas of high dimensional data analysis is variable selection where one might want to find the most contributing variables among the myriad of them. This important problem becomes more difficult when the number of observations is much less than the number of variables. For instance, in bioinformatics and in cancer genomics research where thousands of genes play the role of covariates with a small number of samples. During the past decade, many methods have been proposed such as ISIS and Adaptive LASSO. Recently it has been proposed to use non-local prior densities on Bayesian model parameters and showed it outperforms the commonly used penalized likelihood methods for continuous outcomes. We utilize these newly proposed non-local priors and introduce a new Bayesian method of variable selection named MOMLogit for the binary outcomes. We have tested this algorithm on simulation data, and demonstrated it performs well under different parameter settings. As our method outperforms the ISIS methods under common conditions, we expect our method to have a significant impact on related applications such as in bioinformatics and computational biology.

EO1265: **Fast tree inference with weighted fusion penalties**
Presenter: **Guillem Rigaill**, Universite Evry, France
Given a data set with many features observed in a large number of conditions, it is desirable to aggregate conditions which are similar to ease the interpretation and extract the main characteristics of the data. A fusion penalty framework is presented to address this question when the number of conditions is large. If the fusion penalty is encoded by an l_q-norm, we prove for uniform weights that the path of solutions is a tree. For the l_1- norm the path is piecewise linear and we derive a homotopy algorithm to recover exactly the whole tree structure. For weighted l_1-fusion penalties, we demonstrate that distance-decreasing weights lead to balanced tree structures. For a subclass of these weights that we call "exponentially adaptive", we derive an $O(n\log(n))$ homotopy algorithm and we prove an asymptotic oracle property. This guarantees that we recover the underlying structure of the data efficiently both from a statistical and a computational point of view. We provide a fast implementation of the homotopy algorithm for the single feature case, as well as an efficient embedded cross-validation procedure that takes advantage of the tree structure of the path of solutions. Our proposal outperforms its competing procedures on simulations both in terms of timings and prediction accuracy. As an example we consider phenotypic data: given one or several traits, we reconstruct a balanced tree structure and assess its agreement with the known taxonomy.

EO645 Room Chancellor's Hall DEPENDENCE MODELS AND COPULAS II Chair: Wolfgang Trutschnig

EO0308: **Paths and indices of maximal tail dependence**
Presenter: **Ricardas Zitikis**, University of Western Ontario, Canada
Co-authors: Edward Furman, Jianxi Su
The aim is to demonstrate that the existing methods for measuring tail dependence in copulas may sometimes underestimate the extent of extreme co-movements of dependent risks. This phenomenon holds in the context of both symmetric and asymmetric copulas. As a remedy, we introduce a notion of paths of maximal (tail) dependence and utilize the notion to propose several new indices of tail dependence.

EO0413: **Goodness-of-fit test for specification of semiparametric copula dependence models**
Presenter: **Ostap Okhrin**, Dresden University of Technology, Germany
Co-authors: Shulin Zhang, Peter Song, Qian Zhou
Goodness-of-fit tests for semiparametric copula models are considered. Our contribution is two-fold: we first propose a new test constructed via the comparison between in-sample and out-of-sample pseudo-likelihoods. Under the null hypothesis that the copula model is correctly specified, we show that the proposed test statistic converges in probability to a constant equal to the dimension of the parameter space. We establish the asymptotic normality and investigate the local power of the test. We also extend the proposed test to the specification test of a class of multivariate time series models, and propose a new bootstrap procedure to establish the finite-sample null distribution, which is shown to have better control of type I error than the commonly used bootstrap. Second, we introduce a hybrid mechanism to combine several test statistics, so that the resulting test will make a desirable test power among the involved tests. This hybrid method is particularly appealing when there exists no single dominant optimal test. We conduct comprehensive simulation experiments to compare the proposed new test and hybrid approach with two of the best blanket tests in the literature. For illustration, we apply the proposed tests to analyze two real datasets.

EO1072: **A three-stage estimation of copula-based VAR model**
Presenter: **Giorgia Rivieccio**, Parthenope University, Italy
Co-authors: Giovanni De Luca
The 2008 financial crisis, concretized in a financial and economic default contagion, showed the inadequacy of linear co-movement measures, focusing the attention of researchers on nonlinear and asymmetric dependence structures, which take into account tail dependence. Vector AutoRegressive (VAR) models are able to capture the dynamic behavior of multiple time series, but their use is limited in a linear and symmetric framework. In order to describe the serial and cross-sectional dependence among time series, the copula tool can be a practical method. We introduce a non-linear VAR model based on copula functions, which results more efficient of the classical VAR approach and allows to consider tail dependence estimates. To this end, we extend the three-stage pseudo maximum likelihood estimation (3SPMLE) in order to simplify computational burden in high dimensions when copula-based VAR model is applied. The method is applied to stationary Markov time series, under some assumptions which include a time-invariant copula as well as marginal distributions. We explore, via simulated and real data, the performance of the model compared to the classical VAR, giving the implications of misspecified assumptions for margins and joint distribution and providing tail dependence measures of economic variables involved in the analysis.

EO0681: **On the conditional value at risk (CoVaR) and tail dependence of copulas**
Presenter: **Piotr Jaworski**, University of Warsaw, Poland
Conditional Value at Risk (CoVaR) is a measure of contagion of market X on market Y. The idea is to measure Value at Risk at level β of Y conditioned on X, where random variables X and Y are modelling the returns of market indices for some fixed period of time. We concentrate on the case when the condition putted on X is simply overpassing by the loss from X its Value at Risk at given level α, i.e. $CoVaR_{\beta,\alpha}(Y|X) = VaR_{\beta}((Y|-X>=VaR_{\alpha}(X)))$. Let C be a copula describing the dependence between X and Y. We show how to determine $CoVaR(Y|X)$ basing on the tail behaviour of copula C.

EO168 Room CLO 203 RECENT ADVANCES IN STATISTICAL MODELING AND COMPUTATION Chair: Tsung-I Lin

EO0361: **Hypothesis test for the equality of two population mean vectors for doubly multivariate data**
Presenter: **Anuradha Roy**, The University of Texas at San Antonio, United States
Co-authors: Ivan Zezula, Daniel Klein, Ricardo Leiva
A new hypothesis testing of the equality of mean vectors in two populations using Block T-square statistic for doubly multivariate data on q response variables at p positions in block exchangeable covariance matrix setting is proposed. The minimum sample size needed for our new test is only $q+1$, unlike $pq+1$ in Hotelling's T-square test. The new test is very efficient in small sample scenario, when the number of observations is not adequate to estimate the unknown unstructured variance-covariance matrix. Block exchangeable covariance structure is a realistic assumption in many doubly multivariate data, and it substantially reduces the number of estimated parameters. The motivation for testing p q-variate measurements concurrently (doubly multivariately) rather than testing q-variate measurements separately (just multivariately) p times, is discussed. A natural extension of the Hotelling's T-square statistic, the Block T-square (BT-square) statistic, a convolution of two Hotelling's T-square statistics, is obtained. BT-square statistic uses unbiased estimates of the component matrices of the orthogonally transformed block exchangeable covariance structure that is present in a dataset. A convolution of two Hotelling's T-square distributions is not a Hotelling's T-square distribution. The proposed method is implemented with a dataset from medical studies.

EO0963: **Bayesian semiparametric longitudinal modeling for censored data under skewness**
Presenter: **Mauricio Castro**, Universidad de Concepcion, Chile
In biomedical studies, the analysis of longitudinal data based on Gaussian assumptions is a very common practice. Nevertheless, in many situations, the observed responses are skewed in nature and the use of symmetric mixed effects models can be questionable. Moreover, the responses might be subjected to some upper and lower quantification limits depending on the diagnostic assays used for their detection and also they might have an unspecified non-linear relation with some covariates such as time. To address this situation, a semiparametric Bayesian model for longitudinal censored data using the skew- normal/independent (SNI) family of distributions is considered. The SNI family is an attractive class of skew-symmetric and heavy-tailed distributions that captures, at the same time, the skewness and the heavy-tailed characteristics presented in the data. Therefore, this class of distribution provides an appealing robust alternative to the routine use of symmetric distributions for longitudinal data. The proposed method also uses pth-degree spline smooth functions to approximate the nonlinear functions and considers responses that are either left- or right-censored. An efficient Markov chain Monte Carlo (MCMC) algorithm is introduced to carry out posterior inference and a case deletion influence diagnostics based on the q-divergence measures is presented for examining the robust aspects of the SNI class against outlying and influential observations.

EO1144: **Heavy-tailed longitudinal regression models for censored data**
Presenter: **Victor Hugo Lachos Davila**, UNICAMP, Brazil
Co-authors: Larissa Avila Matos, Luis Mauricio Castro Cepero
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 structure for a censored linear model based on the scale mixtures of skew normal distributions. To compensate for the autocorrelation existing among irregularly observed measures, a damped exponential correlation structure is employed. We propose an exact estimation procedure to obtain the maximum likelihood estimates of the fixed effects and variance components, using a stochastic approximation of the EM (SAEM)algorithm. This approach permits estimation of the parameters of a heavy-tailed longitudinal regression model when censoring is present in an easy and fast way, obtaining as a by-product the standard errors of the fixed effects and the log-likelihood function. The methodology is illustrated through an application to an HIV-AIDS study and several simulation studies.

EO0358: **Truncated multivariate t distributions and its applications**
Presenter: **Tsung-I Lin**, National Chung Hsing University, Taiwan
The use of truncated distributions arises often in a wide variety of scientific problems. In the literature, there are a lot of sampling schemes and proposals developed for various specific truncated distributions. So far, however, the study of the truncated multivariate t (TMVT) distribution is rarely discussed. We present general formulae in analytic matrix expressions for computing the first two moments of the TMVT distribution under the double truncation. Results for the left and right truncation can be viewed as special cases. The slice sampling algorithm is used to generate random variates from the TMVT distribution by introducing auxiliary variables. This strategic approach can result in a series of full conditional densities that are of uniform distributions. The R package TTmoment has been recently developed for a demonstration of its implementation. Several practical examples and real applications are shown to illustrate the feasibility of the proposed results.

EO140 Room Bedford MULTIPLE TESTS I Chair: Arnold Janssen

EO0381: **Martingale approach for multiple testing and FDR control**
Presenter: **Arnold Janssen**, Heinrich-Heine University Duesseldorf, Germany
Under martingale dependence the FDR of various multiple tests can exactly be calculated. The results are key tool in order to discuss finite sample FDR control of these tests. Some of these results are also new when the p-values are independent. Adaptive multiple tests with data dependent critical values will also be discussed.

EO0520: **Robust estimation and FDR control of adaptive multiple tests**
Presenter: **Philipp Heesen**, Heinrich-Heine-University Duesseldorf, Germany
Co-authors: Arnold Janssen
A number of multiple tests which control the false discovery rate (FDR) have been introduced. On this basis, adaptive multiple tests have been developed which improve the exhaustion of the predetermined FDR level by estimating the portion of true null hypotheses. This estimation has originally been designed for true null hypotheses whose corresponding p-values are distributed according to the uniform distribution on $(0,1)$. Although many of these adaptive tests still control the FDR at finite sample size for null hypotheses with p-values which are stochastically larger than the uniform distribution, the exhaustion of the predetermined FDR level may deteriorate since the estimation of the portion of true null hypotheses may deteriorate. New robust estimators for the portion of true null hypotheses and related quantities are presented which lead to new adaptive test procedures of Storey type. Moreover, these test procedures still control the FDR at finite sample size.

EO0818: **Multiple hypothesis testing in multivariate copula models**
Presenter: **Thorsten Dickhaus**, University of Bremen, Germany
We are concerned with simultaneous testing of a family of null hypotheses under a single statistical model. We assume that the individual tests are carried out by means of (marginal) $p-$values and that these $p-$values, regarded as random variables, are dependent. Two popular type I error measures in multiple testing are the family-wise error rate (FWER) and the false discovery rate (FDR). Firstly, we express the threshold of an FWER-controlling simultaneous test procedure (STP) in terms of the copula function of the family of $p-$values, assuming that each of these $p-$values is marginally uniformly distributed on the unit interval under the corresponding null hypothesis. This offers the opportunity to exploit the rich literature on copula-based modeling of multivariate dependency structures for the construction of STPs in non-Gaussian situations. The second part deals with Archimedean copulae in multiple testing in the case that the distributional transforms of the $p-$values are elements of a sequence of exchangeable random variables. We utilize analytic properties of Archimedean copulae to derive sharp bounds for the FDR of the linear step-up test for such $p-$values.

EO0898: **On the Simes test under dependence**
Presenter: **Helmut Finner**, Deutsches Diabetes-Zentrum, Germany
Co-authors: Klaus Strassburger
A modified Bonferroni test procedure for testing an overall null hypothesis in multiple testing problems, nowadays referred to as the Simes test, was proposed time ago. The underlying approach may be considered as a basic step in the development of many new test procedures and new error rate criteria as for example control of the false discovery rate. A key issue is the validity of the Simes test and the underlying Simes inequality under dependence. Although it has been proved that the Simes inequality is valid under suitable assumptions on dependence structures, important cases are not covered yet. We investigate $p-$values based on exchangeable test statistics in order to explore reasons for the validity or failure of the Simes inequality. We provide sufficient conditions for the asymptotic validity of the Simes inequality and its possible strictness. We also show by means of an easy-to-compute counterexample that exchangeability by itself is not sufficient for the validity of the Simes inequality.

EO204 Room CLO 101 BLIND SOURCE SEPARATION MODELS FOR TIME SERIES AND FUNCTIONAL DATA Chair: Hannu Oja

EO0514: **Efficient second order blind identification (SOBI) method for separation of uncorrelated stationary time series**
Presenter: **Sara Taskinen**, University of Jyvaskyla, Finland
Co-authors: Jari Miettinen, Klaus Nordhausen
Second order source separation (SOS) model assumes that the observed p time series are linear combinations of p latent uncorrelated weakly (second-order) stationary time series with different time dependence structures. The aim is to find an estimate for an unmixing matrix, which then transforms the observed time series back to uncorrelated latent time series. The classical approach to estimate the unmixing matrix uses approximate joint diagonalization of autocovariance matrices with different lags and is called SOBI (Second Order Blind Identification). We consider the problem corresponding to the selection of lags for autocovariance matrices to be diagonalized. The choice of lags is known to affect the efficiency of the SOBI estimator, but so far only few ad-hoc guidelines for the lag selection are provided in the literature. Based on recently

derived asymptotical results, we propose an efficient SOBI estimator which uses lags selected from a finite set of candidate sets of lags specified by the user. The theory is illustrated by finite-sample simulation studies as well as a real data example.

EO0536: **Comparison of estimators in performing temporal ICA for fMRI data**
Presenter: **Jari Miettinen**, University of Jyvaskyla, Finland
Co-authors: Roland Boubela, Klaus Nordhausen, Sara Taskinen
Independent component analysis (ICA) has become a standard tool in the analysis of functional magnetic resonance imaging (fMRI) data. So far, spatial ICA has been more frequently applied approach than temporal ICA, mainly due to the large number of voxels (areas of the brain) and small number of time points in fMRI data sets. New high-resolution acquisition techniques have increased the number of time points which has made the more natural approach, temporal ICA, more tempting than before. Most ICA methods, e.g. FastICA, do not take account of the temporal dependence in fMRI data, while some methods, e.g. SOBI (Second Order Blind Identification), use only the temporal dependence. We inspect different ways of comparing the performance of estimators coming from both approaches and find out which estimators are the most suitable for the analysis of the new kind of fMRI data.

EO0590: **On complex valued time series ICA**
Presenter: **Pauliina Ilmonen**, Aalto University School of Science, Finland
Co-authors: Niko Lietzen
In the independent component (IC) model, the elements of a $p-$variate random vector are assumed to be linear combinations of the elements of an unobservable $p-$variate vector with mutually independent components. In independent component analysis (ICA) the aim is to recover the independent components by estimating an unmixing matrix that transforms the observed $p-$variate vector to the independent components. Complex random signals play an increasingly important role in the field of ICA. The complex IC model is used for example in magnetic resonance imaging or antenna array signal processing for wireless communications and radar applications. We consider complex valued time series ICA, in particular, we examine the unmixing matrix estimates that are based on simultaneous use of two complex values autocovariance functionals with different lags. We thus extend the well-known AMUSE algorithm for complex valued variables and we also examine the asymptotic behavior of the obtained complex valued unmixing matrix estimates.

EO1278: **Hilbertian fourth order blind identification**
Presenter: **Germain Van Bever**, The Open University, United Kingdom
Co-authors: Hannu Oja, Frank Critchley, Bing Li, Radka Sabolova
In the classical Independent Component (IC) model, the random vector X is assumed to satisfy $X = \Psi Z$, where Z has independent marginals and Ψ is an invertible mixing matrix. Independent component analysis (ICA) encompasses all methods aiming at unmixing X, that is estimating a (non unique) unmixing matrix Γ such that ΓX has independent components. The celebrated Fourth Order Blind Identification (FOBI) procedure provides such a Γ based on the regular covariance matrix and a scatter matrix based on fourth moments. Nowadays, functional data (FD) are occurring more and more often in practice, and relatively few statistical techniques have been developed to analyze this type of data. Functional PCA is one such technique which focuses on dimension reduction. We propose an extension of the FOBI methodology to the case of Hilbertian data, FD being the go-to example used throughout. When dealing with distributions on Hilbert spaces, two major problems arise: (i) the notion of "marginals" is not naturally defined and (ii) the covariance operator is, in general, non invertible. These limitations are tackled by reformulating the problem in a coordinate-free manner and by imposing natural restrictions on the mixing model. The proposed procedure is shown to be Fisher consistent and affine invariant. A sample estimator is provided and illustrated on simulated and real datasets.

EO082 Room Woburn MULTI-STATE MODELS Chair: Jacobo de Una-Alvarez

EO0594: **Testing for informative observation in multi-state models subject to panel observation**
Presenter: **Andrew Titman**, Lancaster University, United Kingdom
Many observational studies into disease processes do not continually monitor patient status, but instead only observe patients at clinic examination times. In the majority of cases, analysis of the data proceeds by assuming the process which generates the examination times is ignorable. When some or all clinic visits are patient initiated there is a risk that the times will be informative, if for instance patients go to clinic if their condition has deteriorated, and standard estimates may be biased. Building upon previous work which considered a joint parametric model, a joint semi-parametric model is proposed comprising of the underlying parametric multi-state process and an Andersen-Gill counting process model for the observation process. The likelihood for the model can be shown to have a hidden Markov model representation. Moreover, under a null hypothesis of non-informative observation, the two parts of the model can be maximized separately which leads to a convenient construction of a direct score test for informative observation. The test is applied to datasets relating to post-transplantation patients and finite sample properties are investigated by simulation.

EO1366: **Multistate models and causality**
Presenter: **Daniel Commenges**, INSERM, France
Multistate models are more and more used in epidemiology to describe complex evolution of disease status and death. In fact many multi-state models are generated by several events, so that they can also be represented as multivariate counting processes. This representation is useful for several reasons. One important advantage is that causal relationships between events can be represented more clearly this way. Often one of the event is death. This is a special event, however, because once death has occurred, all the other processes that have been attached to the individual lose their meaning. So, in the interpretation of multi-state models, this must be taken into account. Examples will be given as to the interpretation of such models in ageing studies involving disease, death, institutionalization and treatments. The issue of assessing the effect of treatments, which are often adaptative, has been treated in the causal inference literature with most often a counterfactual point of view. It is important to develop some formalism for this issue in a dynamical approach to causality.

EO1352: **Recent advances in nonparametric estimation of a non-Markov transition probability matrix**
Presenter: **Jacobo de Una-Alvarez**, Universidad de Vigo, Spain
Nonparametric estimation of a transition probability matrix will be discussed in a possibly non-Markov progressive multi-state model. Theoretical results, such as consistency and asymptotic normality, and applications to real medical data are included. Several forms of censoring and truncation are considered.

EC1335: **Direct modeling of regression effects for transition probabilities in the non-Markov illness-death model**
Presenter: **Leyla Azarang**, University of Vigo, Spain
We present direct regression analysis for the transition probabilities in the possibly non-Markov progressive illness-death model. The method is based on binomial regression, where the response is the indicator of the occupancy for the given state along time. Randomly weighted score equations which are able to remove the bias due to censoring are introduced. By solving these equations, one can estimate the possibly time-varying regression coefficients, which have an immediate interpretation as covariate effects on the transition probabilities. The performance of the proposed estimator is investigated through simulations. We analyse a real medical data set on colon cancer for illustration purposes.

EO126 Room Montague MODELLING AND COMPUTATION IN STATISTICS OF EXTREMES Chair: Ivette Gomes

EO0667: **Extreme versions of Wang risk measures and their estimation**
Presenter: **Gilles Stupfler**, Aix Marseille Universite, France
Co-authors: Jonathan El Methni
Among the many possible ways to study the right tail of a real-valued random variable, a particularly general one is given by considering the family of its Wang distortion risk measures. This class of risk measures encompasses various interesting indicators such as the widely used Value-at-Risk and Tail Value-at-Risk, which are especially popular in actuarial science, for instance. We start by building simple extreme analogues of Wang distortion risk measures. Special cases of the risk measures of interest include the extreme Value-at-Risk as well as the recently introduced extreme Conditional Tail Moment. Adapted estimators of the resulting extreme Wang distortion risk measures are then introduced when the random variable of interest has a heavy-tailed distribution and their asymptotic normality is shown. The finite sample performance of our estimators is assessed on a simulation study.

EO0993: **Fitting tails truncated at high levels for Pareto or light tails**
Presenter: **Jan Beirlant**, KULeuven, Belgium
Co-authors: Tom Reynkens, Isabel Fraga Alves
Recently, tail fitting has been discussed for truncated and non-truncated Pareto-type distributions. Indeed in some situations ultimately at the largest data, deviations from a Pareto tail behavior can be observed. However truncation can occur also with lighter tails, for instance in the Gumbel domain when studying lifelength of components, or modelling river flows. We generalize the previous methodology to tails with a non-negative extreme value index. We use a pseudo-maximum likelihood approach generalizing the classical Peaks over Threshold approach. Applications to extreme quantile estimation are considered.

EO1012: **Mean of order p reduced bias estimation of the extreme value index: A computational study**
Presenter: **Frederico Caeiro**, Universidade Nova de Lisboa, Portugal
Co-authors: Ivette Gomes
We deal with the estimation of a positive extreme value index, the shape parameter of the extreme value distribution. The classical Hill estimator can be regarded as the logarithm of the geometric mean, or the logarithm of the mean-of-order-0 of a certain set of statistics. Instead of such a geometric mean, the mean-of-order-p (MOP) of those statistics has been previously considered. We work with recent reduced-bias versions of the MOP generalization of the Hill estimator. Apart from the usual integer parameter k, related with the number of top order statistics involved in the estimation, these estimators depend on an the extra real parameter p. Bootstrap and heuristic choices of the tuning parameters p and k are put forward, and an application to simulated and real data is performed.

EO1224: **Extreme episodes and multivariate peaks over thresholds modelling**
Presenter: **Holger Rootzen**, Chalmers, Sweden
Natural catastrophes and fluctuations in financial markets pose very significant societal and economic risks. We review ongoing work on multivariate peaks over thresholds modeling aimed at handling a broad spectrum of environmental and economic risks which develop in time and space. Multivariate Generalized Pareto distribution is used as the main modelling tool, and methods to accommodate time series dependence play an important role. For multivariate distributions the latter is rather unexplored territory. We will discuss relations between max-stable models and multivariate peaks over threshold models, choice of statistical models and methods, and spatial extreme value models which include nuggets.

EO342 Room CLO 102 RECENT ADVANCES IN FUNCTIONAL DATA ANALYSIS AND BAYESIAN STATISTICS Chair: Han Lin Shang

EO0703: **A copula-based imputation model for multilevel data**
Presenter: **Jiali Wang**, Australian National University, Australia
Co-authors: Bronwyn Loong, Anton Westveld, Alan Welsh
The two major approaches to multiple imputation to handle missing data are joint modelling (JM), which assumes a joint distribution for all the variables; and fully conditional specification (FCS), which specifies a conditional distribution for each variable given all the other variables. The former approach has little flexibility in marginal distributions to accommodate different types of variables, whilst the latter approach lacks the theoretical justification to ensure convergence to a proper joint distribution. Combining the best features of both approaches, we propose a semiparametric copula imputation model via the extend rank likelihood function. The proposed method can handle a mixture of variable types and data distributions (for example: continuous and categorical; symmetric and skewed). We can also add random effects to capture cluster effects to handle missing data in a multilevel data set. A Bayesian approach is used to estimate parameters of the copula model and impute the missing values is derived. The method is evaluated through simulations and applied to a dataset from a cluster randomized controlled trial of a multidisciplinary intervention in acute stroke units.

EO0383: **Semiparametric localized bandwidth selection for kernel density estimation**
Presenter: **Tingting Cheng**, Nankai University, China
Co-authors: Xibin Zhang, Jiti Gao
Since conventional cross validation bandwidth selection methods do not work for the case where the data considered are dependent time series, alternative bandwidth selection methods are needed. In recent years, Bayesian based global bandwidth selection methods have been proposed. Our experience shows that the use of a global bandwidth is however less suitable than using a localized bandwidth in kernel density estimation in the case where the data are dependent time series as discussed in an empirical application of this paper. Nonetheless, a difficult issue is how we can consistently estimate a localized bandwidth. We propose a semiparametric estimation method, for which we establish a completely new asymptotic theory for the proposed semiparametric localized bandwidth estimator. Applications of the new bandwidth estimator to the kernel density estimation of Eurodollar deposit rate and the S&P 500 daily return demonstrate the effectiveness and competitiveness of the proposed semiparametric localized bandwidth.

EO0698: **Variable selection in the concurrent functional linear model**
Presenter: **Jeff Goldsmith**, Columbia University, United States
Methods for variable selection are developed when modeling the association between a functional response and functional predictors that are observed on the same domain. This data structure, and the need for such methods, is exemplified by our motivating example: a study in which blood pressure values are observed throughout the day together with measurements of physical activity, heart rate, location, posture, attitude, and other quantities that may influence blood pressure. We estimate the coefficients of the concurrent functional linear model using variational Bayes and jointly model residual correlation using functional principal components analysis. Latent binary indicators partition coefficient functions into included and excluded sets, incorporating variable selection into the estimation framework. The proposed methods are evaluated in simulated- and real-data analyses.

EO1166: **Functional and imaging data in precision medicine**
Presenter: **Todd Ogden**, Columbia University, United States
Co-authors: Adam Ciarleglio, Thaddeus Tarpey, Eva Petkova
A major goal of precision medicine is to use information gathered at the time that a patient presents for treatment to help clinicians determine, separately for each patient the particular treatment that provides the best expected outcome. In psychiatry it is thought that various brain imaging techniques may allow for the discovery of information vital to predicting response to treatment. We will present the general problem of using both scalar and functional data to guide patient-specific treatment decisions and describe some approaches that can be used to perform model fitting and variable selection.

EO336 Room Torrington GENERALIZED LINEAR MODELS AND BEYOND Chair: Christian Kleiber

EO0783: **Beyond Beta regression: Modelling percentages and fractions in the presence of boundary observations**
Presenter: **Ioannis Kosmidis**, University College London, United Kingdom
Co-authors: Anyi Zou
One important limitation of regression models that are based on the Beta distribution is that they are not applicable when at least one of the observed responses is either zero or one - in such scenarios the likelihood function is either $+\infty$ or $-\infty$. The relevant approaches in the literature focus on either the adjustment of the boundary observations by small constants so that the adjusted responses end up in $(0, 1)$, or the use of a discrete-continuous mixture of a Beta distribution and point masses at zero and/or one. The former approach suffers from the arbitrariness of choosing the additive adjustment. On the other hand the latter approach, despite of being natural in some applications, gives a "special" interpretation at the values of zero and/or one relative to observations in $(0, 1)$, and hence it cannot be a general solution to the problem. An extension of the Beta regression model is considered that can naturally accommodate zero and one observations, that avoids the special treatment of such values, and such that it has the usual Beta regression model as a special case. Fitting and inferential procedures for the new model are presented and its usefulness is demonstrated by applications.

EO1127: **Bayesian distributional regression in R**
Presenter: **Jakob Wolfgang Messner**, Universitaet Innsbruck, Austria
Co-authors: Nikolaus Umlauf, Achim Zeileis
Bayesian analysis provides a convenient setting for the estimation of complex generalized additive regression models (GAM). Since computational power has tremendously increased in the past decade it is now possible to tackle complicated inferential problems, e.g., with Markov chain Monte Carlo (MCMC) simulation, on virtually any modern computer. This is one of the reasons why Bayesian methods have become quite popular and it has lead to a number of highly specialized and optimized estimation engines. Because of the very general structure of the additive predictor in GAMs, we propose an unified modeling architecture that can deal with a wide range of types of model terms and can benefit from different algorithms in order to estimate Bayesian distributional regression models. We illustrate the usefulness of the approach using climate data and the new developed R package bamlss (http://BayesR.R-Forge.R-project.org/). A censored normal additive model is employed to derive high resolution spatio-temporal mean and variance fields of precipitation.

EO1108: **Count data regression with excess zeros: A flexible framework using the GLM toolbox**
Presenter: **Achim Zeileis**, Universitaet Innsbruck, Austria
Co-authors: Christian Kleiber
The hurdle model is a two-part model for count data with extra zeros, comprising a binary response part for zeros vs. non-zeros and a zero-truncated count distribution for the positive counts. We show how not only the binary part but also the count component can be analyzed within the GLM framework. This paves the way for flexible extensions of the hurdle model using methods from the extended GLM toolbox such as additive nonlinear terms, boosting, bias reduction, etc. Similarly, it is straightforward to apply visualization techniques such as rootograms, effect displays, or residual plots.

EO1097: **(Quasi-)complete separation in count data regressions with excess zeros**
Presenter: **Christian Kleiber**, Universitaet Basel, Switzerland
Co-authors: Achim Zeileis
We present a computational issue that may occur in count data regressions with excess zeros, notably the widely used hurdle model (aka two-part model). Specifically, the binary response part (for zeros vs. non-zeros) can lead to nonexistence of the MLE under certain conditions. While this is well known for classical binary response models it is somewhat less visible in more flexible frameworks. It is shown how an approach emphasizing the GLM building blocks of the hurdle model helps to identify such situations using a well-known data set from the literature. We also investigate the issue for the related but distinct class of zero-inflation models.

EO196 Room CLO B01 THEORY AND APPLICATIONS OF FUNCTIONAL DATA ANALYSIS Chair: Hans-Georg Mueller

EO0910: **Elliptical distributions and outlier detection for functional data**
Presenter: **Graciela Boente**, Universidad de Buenos Aires and CONICET, Argentina
Co-authors: Matias Salibian-Barrera, David Tyler
As in the multivariate setting, the class of elliptical distributions on separable Hilbert spaces serves as an important vehicle and reference point for the development and evaluation of robust methods in functional data analysis. For example, they allow us to study the Fisher-consistency of robust estimators for functional data when second moments may not exist. We present a simple characterization of elliptical distributions on separable Hilbert spaces that allows us to establish a stochastic optimality property for the principal component subspaces associated with elliptically distributed random elements, without requiring finite second moments. These lower-dimensional approximations can be very useful in identifying potential outliers among functional observations, and we present a robust estimator for these principal subspaces. We illustrate our method on real functional data sets, where the robust estimator is able to discover atypical observations in the data that would have been missed otherwise. Through a simulation study, we also study its performance when used to detect outlying observations.

EO0978: **Additive functional prediction models with multiple predictors**
Presenter: **Byeong Park**, Seoul National University, Korea, South
Co-authors: Hans-Georg Mueller, Kyunghee Han
We propose functional additive models for functional regression with a scalar response and multiple functional predictors that are additive in the functional principal components of the predictor processes. For the case of a single functional predictor, the functional principal components are uncorrelated, so that a simple application of a marginal regression technique can be applied when it is furthermore assumed that the predictor components are independent, as for example in the case of a single Gaussian predictor process. When one has multiple functional predictors this independence assumption cannot be justified and therefore the dependency of the predictor components needs to be addressed. This motivates us to propose a new smooth backfitting technique for the estimation of the additive component functions in functional additive models with multiple

functional predictors. A major difficulty in developing this technique is that the eigenfunctions and therefore the functional principal components of the predictor processes, which are the arguments of the proposed additive model, are unknown and need to be estimated from the data. We investigate how this required estimation of the functional principal components affects the estimation of the additive component functions and develop a complete asymptotic theory. We also study the finite sample properties of the proposed method through a simulation study and a real data example.

EO0456: **A pairwise interaction model for dependent functional data**
Presenter: **Jeng-Min Chiou**, Academia Sinica, Taiwan
Co-authors: Hans-Georg Mueller
Functional data vectors consist of samples of multivariate data in which each component is a random function. We introduce a novel pairwise interaction model that leads to an interpretable decomposition of multivariate functional data and their variation into component-specific processes and pairwise interaction processes. The interaction processes quantify the degree of pairwise interactions between the components of the functional data vectors, while the component-specific processes reflect the functional variation of a particular functional vector component that cannot be explained by the other components. The decomposition of the total variance into component-wise and interaction contributions can be quantified by an R^2-like decomposition. We provide consistency results for the proposed methods and illustrate the model by applying it to sparsely sampled longitudinal data from the Baltimore Longitudinal Study of Ageing, examining the relationships between the body mass index and blood fats.

EC0212: **Multivariate functional principal component analysis for data observed on different (dimensional) domains**
Presenter: **Clara Happ**, Ludwig-Maximilians-University Munich, Germany
Co-authors: Sonja Greven
Existing approaches for multivariate functional principal component analysis are restricted to data on a single interval $\mathcal{T} \subset \mathbb{R}$. The presented approach focuses on multivariate functional data $X = (X^{(1)}, \ldots, X^{(p)})$ observed on different domains $\mathcal{T}_1, \ldots, \mathcal{T}_p$ that may differ in dimension, e.g. functions and images. The theoretical basis for multivariate functional principal component analysis is given in terms of a Karhunen-Loeve theorem. For the practically relevant case of a finite, possibly truncated, Karhunen-Loeve representation, a direct theoretical relationship between univariate and multivariate functional principal component analysis is established. This offers a simple estimation strategy to calculate multivariate functional principal components and scores based on their univariate counterparts. The approach can be extended to univariate components $X^{(j)}$ that have a finite expansion in a general, not necessarily orthonormal basis and is applicable for sparse data or data with measurement error. A flexible software implementation is available. The approach is applied to a neuroimaging study to explore how longitudinal trajectories of a neuropsychological test score covary with FDG-PET brain scans at baseline.

EO174 Room CLO 204 OPTIMAL DESIGNS IN NON-STANDARD SITUATIONS Chair: M Jesus Rivas-Lopez

EO0920: **Optimal designs for nonlinear calibration models**
Presenter: **Jesus Lopez-Fidalgo**, University of Castilla-La Mancha, Spain
Co-authors: Mariano Amo-Salas
The tools provided by optimal design of experiments theory are applied to a nonlinear calibration model. This is motivated by the need of estimating radiation doses using radiochromic films. The calibration model is in this case nonlinear and the explanatory variable cannot be worked out explicitly from the model. The main problem is that a design has to be found on the dependent variable. For that the inverse function theorem will be used to obtain an information matrix to be optimized. We compare in a particular case the proper D-optimal design on the response variable and the transformed design in the explanatory variable using the model function for given nominal values of the parameters.

EO1189: **On convex programming methods for computing optimal designs of experiments under non-standard constraints**
Presenter: **Radoslav Harman**, Comenius University in Bratislava, Slovakia
Co-authors: Guillaume Sagnol
Modern methods of convex optimization, such as the second-order cone programming (SOCP) and the semidefinite programming (SDP), provide powerful alternatives to the traditional, first-order algorithms for computing optimal experimental designs under non-standard experimental constraints. However, the choice of the most efficient mathematical programming method may strongly depend on the specifics of the criterion of optimality as well as on the type of the constraints. We will give a brief survey of the current state of knowledge about the mathematical programming approach to optimal design of experiments under general linear constraints, focusing on the regression models with uncorrelated observations and finite design spaces. We will also compare the mathematical programming and traditional methods with respect to their speed, memory requirements and the scope of application.

EO1292: **Optimal designs for corrected criteria**
Presenter: **Juan M Rodriguez-Diaz**, University of Salamanca, Spain
Many experiments involve random factors, defining a covariance structure in the data, thus Generalized Least Square Estimators (GLSE) of the parameters are used, and their covariance matrix is usually computed using the inverse of the GLSE information matrix. Many optimality criteria are based on this approximation of the covariance matrix. However, this approach underestimates the true covariance matrix of the parameters, and thus the optimality criteria should be corrected in order to pay attention to the actual covariance. The bias in the estimation of the covariance matrix is negligible (or even null) for many models, and for this reason in those cases it makes no sense to deal with the corrected criteria due to the complexity of the calculations involved. But for some models the correction does have importance, and thus the modified criteria should be considered when designing, otherwise the practitioner may risk to deal with poor designs. Some analytical results are presented for simpler models, and optimal designs taking into account the corrected variance will be computed and compared with those using the traditional approach for more complex models, showing that the loss in efficiency may be very important when the correction for the covariance matrix is ignored.

EO1182: **Optimal design for smooth supersaturated models**
Presenter: **Hugo Maruri**, QMUL, United Kingdom
Smooth supersaturated models (SSM) are interpolation models in which the underlying model size, and typically the degree, is higher than would normally be used in statistics, but where the extra degrees of freedom are used to make the model smooth. We will describe the methodology, discuss briefly the role of orthogonal polynomials and then address two design problems. The first is selection of knots and the second a more traditional design problem using SSM to obtain the kernels of interest for D-optimality.

EO208 Room Beveridge Hall HIGH-DIMENSIONAL STATISTICS Chair: Davy Paindaveine

EO0942: **False discovery rate estimation with covariates**
Presenter: **Kun Liang**, University of Waterloo, Canada
Multiple testing becomes an increasingly important topic in high-dimensional statistical analysis. However, most commonly used false discovery rate estimation and control methods do not take covariates into consideration. To better estimate false discovery rate, we propose a novel nonparametric method which efficiently utilizes the covariate information. Our proposed method enjoys some desirable theoretical properties. In addition, we evaluate the performance of our proposed method over existing methods using simulation studies.

EO0768: **Testing the order of a population spectral distribution for high-dimensional data**
Presenter: **Yingli Qin**, University of Waterloo, Canada
Large covariance matrices play a fundamental role in high-dimensional statistics. Investigating the behavior of their eigenvalues can reveal informative structures of large covariance matrices. We propose to test the number of distinct population eigenvalues, i.e. the order of the Population Spectral Distribution (PSD). The proposed statistic is based upon a series of bias-reduced estimators of PSD moments. We develop the limiting distributions of our test statistic and the moment estimators. We also prove the $(n;p)-$strong consistency of these estimators, which are clearly demonstrated in our simulation study.

EO0779: **Statistical and computational trade-offs in estimation of sparse principal components**
Presenter: **Tengyao Wang**, University of Cambridge, United Kingdom
Co-authors: Quentin Berthet, Richard Samworth
In recent years, Sparse Principal Component Analysis has emerged as an extremely popular dimension reduction technique for high-dimensional data. The theoretical challenge, in the simplest case, is to estimate the leading eigenvector of a population covariance matrix under the assumption that this eigenvector is sparse. An impressive range of estimators have been proposed; some of these are fast to compute, while others are known to achieve the minimax optimal rate over certain Gaussian or subgaussian classes. We show that, under a widely-believed assumption from computational complexity theory, there is a fundamental trade-off between statistical and computational performance in this problem. More precisely, working with new, larger classes satisfying a Restricted Covariance Concentration condition, we show that no randomised polynomial time algorithm can achieve the minimax optimal rate. On the other hand, we also study a (polynomial time) variant of the well-known semidefinite relaxation estimator, and show that it attains essentially the optimal rate among all randomised polynomial time algorithms.

EO0664: **High-dimensional tests for principal component analysis**
Presenter: **Thomas Verdebout**, Universite Libre de Bruxelles, Belgium
Co-authors: Davy Paindaveine, Christine Cutting
Principal Component Analysis (PCA) is one of the most important tools in multivariateanalysis. Nowadays, it is getting more and more popular in the statistical communitysince it is a specific high-dimensional problem. Indeed, the main objective of PCA isdimension reduction. In this paper, we tackle the problem of testing that the first principal component can be obtained by projecting the data along a direction which is specified under the null hypothesis. We are interested in the asymptotic behavior of some tests when both the sample size and the dimension become large.

EO142 Room Holden NONPARAMETRIC BAYESIAN METHODS AND INVERSE PROBLEMS Chair: Jan Johannes

EO0991: **Nonregular Bayesian inverse problems**
Presenter: **Natalia Bochkina**, University of Edinburgh, United Kingdom
To overcome ill-posedness of a statistical inverse problem, the likelihood is usually penalised to include a priori information about smoothness, or other regular behaviour of the solution. Such regularised solution to the inverse problem can be interpreted as a maximum a posteriori solution to a Bayesian inverse problem, in which all sources of variation, uncertainty and error are quantified using probability. We consider Bayesian inverse problems where the likelihood can be nonregular, and the unknown function that is observed indirectly is not necessarily bounded away from 0. As a motivating example, we consider SPECT tomography where a stack of images is observed indirectly; from the statistical perspective it represents a linear inverse problem with Poisson errors and the identity link function. A distinct feature of this problem is that at some pixels/ voxels the Poisson mean can degenerate to 0 which affects the rate of contraction of the posterior distribution of the unknown image around its boundary. This is related to nonregular behaviour of the likelihood at the boundary of the parameter space. We present the rate of contraction of the posterior distribution of the intensities of the unknown image for such nonregular models, and illustrate it for the SPECT tomography problem.

EO0937: **Adaptive Bayesian estimation in indirect Gaussian sequence space models**
Presenter: **Jan Johannes**, University Heidelberg, Germany
Co-authors: Anna Simoni, Rudolf Schenk
In an indirect Gaussian sequence space model lower and upper bounds are derived for the concentration rate of the posterior distribution of the parameter of interest shrinking to the parameter value θ_0 that generates the data. While this establishes posterior consistency, however, the concentration rate depends on both θ_0 and a tuning parameter which enters the prior distribution. We first provide an oracle optimal choice of the tuning parameter, i.e. optimized for each θ_0 separately. The optimal choice of the prior distribution allows us to derive an oracle optimal concentration rate of the associated posterior distribution. Moreover, for a given class of parameters and a suitable choice of the tuning parameter, we show that the resulting uniform concentration rate over the given class is optimal in a minimax sense. Finally, we construct a hierarchical prior that is adaptive. This means that, given a parameter θ_0 or a class of parameters, respectively, the posterior distribution contracts at the oracle rate or at the minimax rate over the class. Notably, the hierarchical prior does not depend neither on θ_0 nor on the given class. Moreover, convergence of the fully data-driven Bayes estimator at the oracle or at the minimax rate is established.

EO1814: **Gaussian processes and Bayesian moment estimation**
Presenter: **Anna Simoni**, CREST - ENSAE and CNRS, France
Co-authors: Jean-Pierre Florens
Given a set of moment restrictions that characterize a parameter θ, the purpose is to investigate a semiparametric Bayesian approach for estimation of θ that imposes these moment restrictions in the nonparametric prior for the data distribution. We establish that a Gaussian process (GP) prior for the density function associated with the data distribution F is particularly convenient in order to impose over-identifying restrictions. We circumvent the difficulty of specifying the likelihood function, which is not available, by constructing a linear functional transformation of F such that its empirical counterpart has an asymptotic Gaussian distribution that will be used as the sampling model. We provide a frequentist validation of our procedure by showing: consistency of the maximum a posteriori estimator for θ, consistency and asymptotic normality of the posterior distribution of θ.

EC1307: **Bayesian nonparametric inference for discovery probabilities: Credible intervals and large sample asymptotics**
Presenter: **Stefano Favaro**, University of Torino and Collegio Carlo Alberto, Italy
Co-authors: Julyan Arbel, Bernardo Nipoti, Yee Whye Teh
Given a sample of size n from a population of individual belonging to different species with unknown proportions, a popular problem of practical interest consists in making inference on the probability $D(l;n)$ that the $(n+1)-$th draw coincides with a species with frequency l in the sample, for any $l = 0,1,...,n$. We contribute to the methodology of Bayesian nonparametric inference for $D(l;n)$. Specifically, under the general framework of Gibbs-type priors we show how to derive credible intervals for the Bayesian nonparametric estimator of $D(l;n)$, and we investigate the large n asymptotic behaviour of such an estimator. Of particular interest are special cases of our results obtained under the assumption of the two parameter Poisson–Dirichlet prior and the normalized generalized Gamma prior, which are two of the most commonly used Gibbs-type priors.

EO326 Room Athlone STATISTICS FOR FUZZY- OR SET-VALUED DATA I Chair: Angela Blanco-Fernandez

EO1081: **Statistical tests and confidence intervals for the Aumann mean**
Presenter: **Raffaello Seri**, University of Insubria, Italy
A method was recently proposed to obtain a confidence set for the Aumann mean of a random closed set (RACS) as well as statistical tests for inclusion/exclusion of the Aumann mean with respect to a given set C. This framework has been extended to the case in which the set C depends itself on the Aumann mean of the RACS, to test, say, rotational invariance of the Aumann mean. The proposed techniques used the bootstrap and required a preliminary discretization of the support function of the RACS through the identification of a set of directions on the unit hypersphere. This discretization severely limits the performance of the method. We investigate some alternative computational strategies using classical asymptotic theory coupled with computational techniques for the estimation of the variance matrix. Despite the bootstrap is recognized as a powerful technique, whose accuracy is generally better than asymptotic approximations, in this case it seems to be outperformed.

EO1026: **On the robustness of fuzzy-valued M-estimators**
Presenter: **Beatriz Sinova**, University of Oviedo, Spain
Co-authors: Stefan Van Aelst
Fuzzy-valued data have become a suitable tool to deal with imprecision in real-life experiments. During the last years many statistical techniques and procedures have been adapted to cover this new situation, such as regression analysis, clustering or hypothesis testing. Most of these developments are based on the Aumann-type mean as measure of location, but this fact leads to non-robust statistical conclusions due to the high sensitivity of the mean, inherited from the real settings. To overcome this important drawback, the successful and classical approach of M-estimators of location will be extended to deal with fuzzy-valued data and its main features in terms of robustness will be analyzed.

EO1646: **Robust techniques to estimate linear regression models for interval-valued data**
Presenter: **Angela Blanco-Fernandez**, University of Oviedo, Spain
Co-authors: Gil Gonzalez-Rodriguez
Some robust statistical techniques to estimate linear regression models for interval-valued data are shown. The least-squares (LS) estimation of the models has been previously solved through constrained optimization techniques guaranteeing the coherence of the solutions within the interval framework. However, as happens with classical estimation techniques of regression problems, the LS estimators are not robust againts outliers. Thus, alternative estimation approaches should be developed. One of the best-known procedures in classical regression is the least trimmed estimator. It is based on estimating the regression coefficients by trimming the sample dataset, discarding the observations with greater residuals. The technique is applied to the interval framework. Several alternatives are investigated, and some applications and simulations are shown to illustrate the behaviour of the procedures.

EC1480: **Nonparametric estimation of emergence curves in weed science**
Presenter: **Ricardo Cao**, University of Coruna, Spain
Co-authors: Miguel Reyes, Mario Francisco
The aim is to study nonparametric distribution function estimation when the data are grouped. Grouped data appear whether continuous random variables are measured or used in binned or rounded form or in systems in which the observation time is periodic. The motivational problem comes from a branch of agriculture called weed science. In this area, random variables based on humidity or temperature (or both) are very important for predicting weed emergence. In some weed science experiments the observation time is periodic, so researchers are unable to observe the exact values of those variables; instead, they obtain a data set consisting in counts between variable consecutive monitoring times. Moreover, sometimes, only data expressed as proportions of emerged seedlings are available. The problem of estimating the distribution function, F, with grouped data based on non-equidistant inspection times is addressed. A cumulative distribution function estimator is derived, which by construction, is already adapted for grouped data. Its asymptotic properties are derived, and its performance in different grouping scenarios is analyzed through simulation studies. Also, a brief study on bandwidth selection in this context is included. The proposed method is applied to analyze three seedling emergence datasets. It is also compared with nonlinear regression fits, which are the standard methods to address this problem in the weed science literature.

EO192 Room Jessel APPLICATIONS OF MIXTURE MODELS Chair: Paul McNicholas

EO1090: **Modeling the receiver operating characteristic curve using Gaussian and non-Gaussian mixture models**
Presenter: **Amay Cheam**, McMaster University, Canada
Co-authors: Paul McNicholas
The receiver operating characteristic (ROC) curve remains a topic of discussion and interest after all these years. The curve displays the capacity of a diagnostic test to distinguish between two groups of patients, diseased and non-diseased. In the literature, many approaches have been proposed for modeling the ROC curve whether it is direct or indirect. Because of its tractability, the Gaussian distribution has been extensively used to model both groups. In parallel, the finite mixture models have gained fame as a compelling apparatus in modeling data. We propose to model the ROC curve using Gaussian and non-Gaussian mixture distributions (specifically mixture of t and skew t-distributions), leading to a more flexible model that accounts for heterogeneous data, unlike the classical binormal curve. The Monte Carlo method is used in conjunction to circumvent the absence of a closed-form and to obtain confidence bands for the derived ROC. The proposed method will be illustrated via simulated and real data.

EO1133: **On the interpretation of multiple clusterings**
Presenter: **Ryan Browne**, University of Waterloo, Canada
We explore the idea of identifying different partitions based on mutually exclusive sets of variables with a dataset. In the literature this concept is known as multiple clusterings or multiple cluster structures. We suggest this is a misnomer as the methodology yields a single clustering solution. However, these types of solutions, which we call independent clusters, give rise to different interpretations of clusterings then the usual several distinct components. We relax the assumption that these clusters are composed of variables subsets to directional subsets. Along the way we demonstrate this methodology on simulated and real data.

EO1023: **An efficient SEM algorithm for Gaussian Mixtures with missing data**
Presenter: **Vincent Vandewalle**, Inria, France
Co-authors: Christophe Biernacki
The missing data problem is well-known for statisticians but its frequency increases with the growing size of modern datasets. In Gaussian model-based clustering, the EM algorithm easily takes into account such data by dealing with two kinds of latent levels: the components and the variables. However, the quite familiar degeneracy problem in Gaussian mixtures is aggravated during the EM runs. Indeed, numerical experiments clearly reveal that degeneracy is quite slow and also more frequent than with complete data. In practice, such situations are difficult to detect efficiently. Consequently, degenerated solutions may be confused with valuable solutions and, in addition, computing time may be wasted through wrong runs. A theoretical and practical study of the degeneracy will be presented. Moreover a simple condition on the latent partition to avoid degeneracy will be exhibited. This condition is used in a constrained version of the Stochastic EM (SEM) algorithm. Numerical experiments on real and simulated data illustrate the good behaviour of the proposed algorithm.

EO1080: **Examining service utilisation among the frail elderly using latent class analysis**
Presenter: **Arthur White**, Trinity College Dublin, Ireland
Co-authors: Jason Wyse
Data collected by the Irish Longitudinal Study on Ageing (TILDA) are analysed. This is a prospective cohort study which is representative of the elderly Irish community dwelling population. Our analysis focuses on respondents aged 65 years and over who are deemed to be frail, with variables related to the use of services such as dentistry, informal care, and hospital stay considered. Some of these variables are binary, indicating presence or absence of use, while others are integer valued, and relate to frequency of use. While it is suspected that heterogeneity of service use and need exists among the elderly frail in Ireland, its extent, and the services for which it is most pronounced, are poorly understood. To examine this issue, a latent class analysis model is developed and fitted to the data. This consists of a mixture of binomial and Poisson distributions that correspond to the relevantly valued variables. Methods for model and variable choice are examined, and the broader implications of these choices and the resultant clusterings are discussed.

EO276 Room Bloomsbury GRAPHICAL MARKOV MODELS AND TOTALLY POSITIVE DEPENDENCES Chair: Nanny Wermuth

EO1141: **Latent class analysis for multivariate binary symmetric variables with positive loadings**
Presenter: **Giovanni Marchetti**, University of Florence, Italy
A model for binary variables is discussed satisfying both symmetry and the latent class assumption. The strong relations between the parameters of the model and the correlations are shown and the consequences of positive loadings are discussed.

EO1188: **On MTP_2 distributions**
Presenter: **Caroline Uhler**, IST Austria, Austria
We analyze distributions that are multivariate totally positive of order 2 (MTP_2) and discuss various nice properties of such distributions. We identify necessary and sufficient conditions for a conditional Gaussian distribution to be MTP_2 and show that any independence model generated by an MTP_2 distribution is a singleton-transitive compositional semigraphoid. Finally we prove that any such MTP_2 distribution is faithful to its concentration graph.

EO1167: **Brownian motion tree models**
Presenter: **Piotr Zwiernik**, University of Genoa, Italy
Co-authors: Caroline Uhler
The Brownian motion tree model on a tree T is a special Gaussian graphical model, where the inner nodes are not observed and there is strong positive dependence between neighbours. We will present various structural results on this highly structured model class. We will discuss ways of estimation both when T is known and when it is unknown; one of them will be a version of the structural EM algorithm. Some of these results are already known but we will show how a geometric and algebraic understanding can give additional insight into the estimation process.

EO1190: **Some aspects of dependency in multiple time series**
Presenter: **David Cox**, Oxford University, United Kingdom
Studies of dependencies between time series may take two broad forms. In one, the data consist of one, or a small number, of long sequences. In the other form, which is considered, the data consist of many individually quite short series. For example, a patient may have annual clinic visits at each of which a battery of tests is performed. Observations are thus of three broad types, intrinsic features, such as gender, then a sequence in time of test results, often to be regarded as explanatory, and finally one or more outcome variables. Dependencies in such systems may take various forms and some of the resulting issues of analysis and representation are outlined.

EC038 Room Gordon CONTRIBUTIONS ON SEMI- AND NON-PARAMETRIC STATISTICS Chair: Thomas Kneib

EC0242: **A particular form of non-constant effect in two-stage quantile regression**
Presenter: **Christophe Muller**, Aix-Marseille School of Economics, France
The aim is to study the fitted-value approach to quantile regression in the presence of endogeneity under a weakened form of the IV condition. In this context, we exhibit the possibility of a particular form of non-constant effect models with the fitted-value approach, a situation often believed to be ruled out. However, only the constant effect coefficients of the model can be consistently estimated. Finally, we discuss practical examples where this approach can be useful to avoid misspecification of quantile models.

EC1731: **Testing the difference between two quantiles from independent populations**
Presenter: **Martina Mittlboeck**, Medical University of Vienna, Austria
Co-authors: Harald Heinzl
A test of the difference between two quantiles from independent populations may be based on an empirical distribution function estimator for the difference of two order statistics. Properties of this test are studied by means of a simulation study. Small to moderate sample sizes, tail quantiles and quantiles which do not coincide with the expectation of an order statistic are identified as problematic and can result in both, conservative but also extreme liberal behavior of the test. Consequently, this statistical test should be used with care for small to moderate sample sizes. This is especially true in areas of application where liberal statistical tests are usually not acceptable, like marketing authorization of medicines or medical devices.

EC1749: **Efficient estimation and testing of varying coefficient panel data models**
Presenter: **Alexandra Soberon**, Universidad de Cantabria, Spain
Co-authors: Winfried Stute, Juan Manuel Rodriguez-Poo
Efficient nonparametric estimation and inference of varying coefficient panel data models is addressed. Our aim is twofold. On one hand, we

propose a more accurate nonparametric estimator based on a pairwise differencing transformation. Later, a generalized least squares (GLS) procedure to incorporate the information of the error covariance matrix is presented. Despite the established theory in the literature, it is shown that the resulting feasible GLS estimator is asymptotically more efficient than the above without imposing any closeness property within groups. On the other hand, to estimate the variance coponents we develop a new technique based on the efficient estimation of moments when distributional assumptions such as normality of the error variables cannot be justified. Sufficient conditions for the asymptotic normality of these estimators are given. Furthermore, an alternative test statistic is proposed to test the significance of the individual effects based on the asymptotic properties of the estimator of the variance components. Some simulations are used to examine the finite sample performance of the proposed test statistic and estimators. Also, to see how useful are our results, we illustrate an application about the production efficiency in the European Union's companies.

EC1431: **Semiparametric estimation for model structure discovery**
Presenter: **Takuma Yoshida**, Kagoshima University, Japan
In the parametric regression analysis, it is important to confirm whether the model assumption is correct or not. We propose the new semiparametric estimator to achieve this. First we obtain the parametric estimator such as the least squares method. In the second step, we apply a certain nonparametric smoother to the residual data associated with the parametric estimator in the first step. The final estimator is constructed by summing both the parametric and the nonparametric estimators. In the second step of our method, we use the B-spline method with group lasso penalty. Then if the parametric model assumption is correct, the nonparametric estimator becomes to be zero function and hence the semiparametric estimator is reduced to the parametric estimator. Thus the model structure is discovered. If not so, we obtain the semiparametric estimator and it is known from existing results that the semiparametric estimator has better behavior than the ordinary nonparametric estimator. Consequently, the proposed method discover the model structure and obtain the final estimator simultaneously. The proposed method can also be applied to additive models. For additive models, we use the adaptive group lasso penalty in the second step estimation to assess the different weight to different additive component. We introduce the method, its asymptotic properties and the numerical study.

EG123 Room G21A CONTRIBUTIONS ON NETWORKS Chair: Yi Yu

EC1583: **Modeling measurement error via non parametric Bayesian belief nets**
Presenter: **Paola Vicard**, University Roma Tre, Italy
Co-authors: Daniela Marella
Measurement error is the difference between the value provided by the respondent and the true (but unknown) value. It is sometimes defined as observation error, since it is related to the observation of the variable at the data collection stage. The problem of measurement error in financial assets is studied. The measurement error is modeled by means of non parametric Bayesian belief networks, that are graphical models expressing the dependence structure through bivariate copulas associated to the edges of the graph without introducing any distributional assumption. A new error correction procedure based on non parametric Bayesian belief networks is proposed. Measurement error modeling and microdata correction are illustrated by means of an application to the Banca d'Italia Survey on Household Income and Wealth 2008. The measurement model and its parameters have been estimated via a validation sample. The sensitivity of the conditional distribution of the true value given the observed one to different evidence configurations is analysed.

EC1781: **Dynamics of networks: The mean field approach for probabilistic cellular automata on random and small-world graphs**
Presenter: **Lourens Waldorp**, University of Amsterdam, Netherlands
Co-authors: Jolanda Kossakowski
We describe the dynamics of networks using one-dimensional discrete time dynamical systems theory obtained from a mean field approach to (elementary) probabilistic cellular automata (PCA). Often the mean field approach is used on a regular graph (a grid or torus) where each node has the same number of edges and the same probability of becoming active. We consider elementary PCA where each node has two states (two-letter alphabet): 'active' or 'inactive' (0/1). We then use the mean field approach to describe the dynamics of a random graph and a small-world graph. The mean field can now be viewed as a weighted average of the behaviour of the nodes in the graph, since the behaviour of the nodes is determined by a different number of edges. The mean field predicts (pitchfork) bifurcations and in some settings chaotic behaviour can be seen. The application we have in mind is that of psychopathology. A mental disorder can be viewed as a network of symptoms, each symptom influencing other symptoms. For instance, lack of sleep during the night could lead to poor concentration during the day, which in turn could lead to lack of sleep again by worrying that your job may be on the line. The symptom graph is more likely to be a small-world than a grid. The mean field approach then allows possible explanations of 'jumping' behaviour in depression, for instance.

EC1675: **Estimating exponential random graph models for large networks**
Presenter: **Alberto Caimo**, University of Lugano, Switzerland
The exponential random graph model (ERGM) is a statistical model for analysing social networks. However, estimating ERGM parameters is a computationally intensive procedure that imposes severe limits on the size of networks that can be fitted. Recently, it has been shown that conditional estimation can be used to estimate ERGM parameters by estimating parameters for smaller conditionally independent subsets of the network. Snowball sampling can be used to generate such subsets. A large number of relatively small samples can be estimated in parallel, taking advantage of parallel computing to allow estimation of much larger networks than previously possible.

EC1541: **Corrected network measures**
Presenter: **Vladimir Batagelj**, IMFM, Slovenia
We discuss two well-known network measures: the clustering coefficient of a node and the overlap weight of an edge. For both of them it turns out that they are not very useful for data analytic task to identify important elements of a given network. The reason for this is that they attain the largest values on "complete" subgraphs of relatively small size - they are more probable to appear in a network than that of larger size. We show how their definitions can be corrected in such a way that they give the expected results. We illustrate the proposed corrected measures by applying them on some real-life networks using a Pajek program.

EG099 Room SH349 CONTRIBUTIONS ON INDEPENDENCE AND DISTANCE-BASED METHODS Chair: Xiangrong Yin

EC1600: **Estimating the Zolotarev distance to the class of exponential distributions**
Presenter: **Amparo Baillo**, Universidad Autonoma de Madrid, Spain
Co-authors: Javier Carcamo
The exponential distribution is one of the most important and widely used distributions in statistical applications. Knowing the distance from a positive random variable X to the exponential class $\mathcal{E}$ would assess if it is reasonable or not to assume an exponential model as an approximation to the unknown probability distribution of X. We study the asymptotic behaviour of a plug-in estimator of a normalized version of the Zolotarev distance between X and $\mathcal{E}$. Its performance in practice has been checked via simulations and the analysis of real data sets.

EC0565: **Calibration estimator with new distance functions**
Presenter: **M G M Khan**, The University of the South Pacific, Fiji

Calibration is a method of adjusting the original design weights that increase the precision of the estimates of a characteristic incorporating the known population parameters of auxiliary variables. In order to minimize a given distance measure, the calibration weights are chosen satisfying the constraints related to the available auxiliary information. In survey sampling, many authors have defined calibration estimators by using different constraints and a chi-square type distance function. For stratified random sampling design, we review the calibration approach and propose some new calibration estimators of population mean using new distance functions by varying calibration constraints. A numerical study was carried out to compare the performance of the proposed estimators with the existing calibration estimators. It revealed that the proposed calibration estimators are useful in increasing the precision of the estimates.

EC0700: **A new measure for testing independence**
Presenter: **Qingcong Yuan**, University of Kentucky, United States
Co-authors: Xiangrong Yin

A new measure for testing independence between two random vectors is introduced. Our measure differs from that of distance covariance, by using expected conditional difference of characteristic functions. We propose one empirical version by slicing on one of the random vectors. We show that this particular version is equivalent to DISCO. This empirical measure is based on certain Euclidean distance. Its properties, asymptotics and applications in testing independence are discussed. Implementation and Monte Carlo results are also presented.

EC1747: **A general approach to orthonormal coordinate representation of compositional tables**
Presenter: **Kamila Facevicova**, Palacky University Olomouc, Czech Republic
Co-authors: Karel Hron

Compositional tables can be considered as a continuous counterpart to the well-known contingency tables. Accordingly, their cells, containing in general positive real numbers rather than just counts, carry relative information about relationships between two factors. As a consequence, compositional tables can be considered as a special case of (vector) compositional data. Since the standard analytical methods are not suitable for this kind of data, compositions are popularly expressed in the orthonormal coordinates using sequential binary partition. Even though the resultant coordinates (balances) are well interpretable in sense of logratio between two groups of parts, they do not respect the two-dimensional nature of compositional tables and the information about relationship between factors is thus not preserved. The aim is to present an alternative system of orthonormal coordinates with respect to the Aitchison geometry, which enables analysis of independence of factors. Moreover, it has an intuitive interpretation due to consisting partly of balances between whole rows (columns) and partly of logarithms of odds ratios, which are popularly applied also for description of relations within contingency tables. Finally, the new coordinate system enables to process a sample of compositional tables and thus it represents a possible alternative to loglinear models.

EP653 Room Macmillan Hall and Crush Hall POSTER SESSION I Chair: Francisco Torres-Ruiz

EP1370: **Cross validating extensions of kernel, sparse or regular partial least squares regression models to censored data**
Presenter: **Frederic Bertrand**, Universite de Strasbourg, France
Co-authors: Philippe Bastien, Myriam Maumy-Bertrand

When cross-validating standard or extended Cox models, the commonly used criterion is the cross-validated partial loglikelihood using a naive or a van Houwelingen scheme. Quite astonishingly, we will show, using a strong simulation study involving three different data simulation algorithms, that these two methods fail with the extensions, either straightforward or more involved ones, of partial least squares regression to the Cox model. This is quite an interesting result for at least two reasons. Firstly, several nice features of PLS based models, including regularization, interpretability of the components, missing data support, biplots of individuals and variables -and even parsimony for SPLS based models-, account for a common use of these extensions by statisticians who usually select their hyperparameters using cross-validation. Secondly, they are almost always featured in benchmarking studies to assess the performance of a new estimation technique used in a high dimensional context and often show poor statistical properties. We carried out a vast simulation study to evaluate more than a dozen of potential cross-validation criteria, either AUC or prediction error based. Several of them lead to the selection of a reasonable number of components. Using these newly found cross-validation criteria to extensions of partial least squares regression to the Cox model, we performed a benchmark reanalysis that showed enhanced performances of these techniques.

EP1393: **Certainty bands for the conditional cumulative distribution function and applications**
Presenter: **Myriam Maumy-Bertrand**, Universite de Strasbourg, France
Co-authors: Sandie Ferrigno, Bernard Foliguet, Aurelie Muller-Guedin

We establish uniform asymptotic certainty bands for the conditional cumulative distribution function. To this aim, we give exact rate of strong uniform consistency for the local linear estimator of this function. The corollaries of this result are the asymptotic certainty bands for the quantiles and the regression function. We illustrate our results with simulations and an application on fetopathologic data.

EP1445: **Stopped sum model variants**
Presenter: **Wan Jing Low**, University of Wolverhampton, United Kingdom
Co-authors: Paul Wilson, Mike Thelwall

Stopped sum models have been used to model damage processes, risk assessments, and branching processes in ecology. They use a two stage process with the second generation depending upon the numbers in the first generation, where the generations may follow the same or different distributions. This could apply to biological populations, for example, where the second generation is the offspring of the first generation. We propose two stopped sum model variants. Unlike the standard stopped sum model, which only models the second generation of a population process, these variants consider the sum of both generations. We consider the cases where both generations are Poisson distributed, or one is Poisson and the other is negative binomial. These stopped sum variant models are fitted to citation count and other data and compared with more standard models. The results show that the new variants have a superior fit in some cases compared to more standard models, for example, negative binomial, discretised lognormal and Neyman type-A.

EP1523: **Optimal filtering algorithm with stochastic nonlinearity, random parameter matrices, sensor delays and correlated noises**
Presenter: **Josefa Linares-Perez**, Universidad de Granada, Spain
Co-authors: Raquel Caballero-Aguila, Aurora Hermoso-Carazo

The estimation problem in sensor network stochastic systems has become an important research topic in the last years due to their wide applications. Usually in practice, these systems are influenced by additive nonlinear disturbances and the communication capacity limitations and unreliable network characteristics may yield transmission uncertainties. Such nonlinear disturbances and uncertainties may depend on the network conditions which often change randomly. The sensor network system involves simultaneously random parameter matrices, stochastic nonlinear disturbances, and correlated noises in the state-space model, as well as randomly delayed measurements in the transmission. The stochastic nonlinearity is described by statistical means; these stochastic nonlinear functions encompass different types of nonlinearities that have been considered in stochastic

systems, such as state-multiplicative noises or vectors dependent on the norm of the state. The process noise and all the sensor measurement noises are one-step autocorrelated, and the process noise and different sensor noises are two-step cross-correlated. The random delay phenomenon is modelled by a different sequence of Bernoulli variables in each sensor. Using an innovation approach, the centralized least-squares linear filter is designed by a recursive algorithm. A simulation example is given to illustrate the feasibility of the proposed filtering algorithm.

EP1525: **A non-homogeneous Gompertz process with jumps as model of tumor dynamics**
Presenter: **Francisco Torres-Ruiz**, Granada, Spain
Co-authors: Virginia Giorno, Patricia Roman-Roman, Serena Spina
We consider a stochastic model of cancer growth subject to an intermittent treatment leading to a reduction in tumor size but producing an increase in the growth rate. The considered model is based on a non-homogeneous Gompertz stochastic diffusion process with jumps, representing an application of a therapy that shifts the cancer mass to a return state. Nevertheless, the successive application of the therapy produces an increase in the growth rate. This increase is represented in the model by the inclusion of time functions depending on the cycle of application of the therapy. The resulting process consists in a combination of different non-homogeneous Gompertz diffusion processes for which the estimation of the parameters is provided when the time functions modeling the growth rate is known, as well as the estimation of these functions in the unknown case. Some simulations are performed to illustrate the validity of the proposed procedure.

EP1622: **Fusion fixed-point smoothing in systems with random losses and correlated noises**
Presenter: **Maria Jesus Garcia-Ligero**, Universidad de Granada, Spain
Co-authors: Aurora Hermoso-Carazo, Josefa Linares-Perez
In the transmission of measurement data in sensor networks, errors frequently occur, due, for example, to intermittent failures in the observation mechanism or limited battery energy. These errors can lead to the loss of data, or packet dropouts, which are inherently random. An appropriate way to address this problem is to use the last measurement received, if the measurement in the current time is not available, which is usually modeled using Bernoulli random variables sequences. The aim is to study the least-squares linear fixed-point smoothing problem from measurements coming from different sensors with random losses modeled by independent Bernoulli random variables with different characteristics for each sensor. We also assume that the noises of the observations are correlated; this assumption, that generalizes the usual independence hypothesis, appears in many real problems and, hence, this consideration is receiving great interest in recent years. Under these assumptions, and considering that the state-space model of the signal is not fully available, a fusion smoothing algorithm is derived from the information provided by the first and second-order moments of the processes involved in the observation model.

EP1522: **Multi-sensor fusion filtering problems with random measurement matrices, cross-correlated noises and packet dropouts**
Presenter: **Aurora Hermoso-Carazo**, Universidad de Granada, Spain
Co-authors: Josefa Linares-Perez, Raquel Caballero-Aguila
In recent years, the use of random matrices in research on fusion estimation problems of networked systems has gained great interest since they arise in many situations involving stochastic sensor gain degradation, measurement multiplicative noises or missing measurements. Furthermore, sensor networks usually produce communication random packet losses which could degrade the network performance. We address the centralized and distributed fusion filtering problems of discrete-time signals using measurements perturbed by random parameter matrices, which are transmitted by different sensors subject to random packet dropouts. Different white sequences of Bernoulli random variables with known probabilities are used to model the potential packet dropouts. Moreover, the fairly conservative assumption that the measurement noises are uncorrelated is weakened and it is assumed that all the sensor noises are one-step autocorrelated and different sensor noises are one-step cross-correlated. Using covariance information, a recursive algorithm for the centralized least-squares linear filter is derived by an innovation approach. Also, local least-squares linear filters based on the measured data of each sensor are obtained and the distributed fusion method is then used to obtain a fusion filter as the matrix-weighted sum of such local estimators that minimizes the mean squared estimation error.

EP1515: **Distributed fusion signal filter from uncertain observations subject to random transmission delays and packet dropouts**
Presenter: **Raquel Caballero-Aguila**, University of Jaen, Spain
Co-authors: Aurora Hermoso-Carazo, Josefa Linares-Perez
Fusion estimation problems in networked systems have become an interesting research area due to the increasing use of sensor networks, where the measured data from the different sensors must be combined to address the estimation problem. The unreliable network characteristics can yield only-noise measured outputs (uncertain observations) and, also, due to the communication channel imperfections, delays and/or packet dropouts can randomly occur during the data transmission to the fusion center. The distributed fusion signal filtering problem is investigated from uncertain observations with random transmission delays and packet dropouts. These mixed random uncertainties are modeled by different sequences of Bernoulli random variables in each sensor. First, using an innovation approach, local least-squares filtering estimators, together with their error covariance matrices, are obtained by recursive algorithms without requiring the state-space model generating the signal, but only the mean and covariance functions of the signal and observation noises as well as the uncertainty probabilities. Second, the cross-covariance matrices of the filtering errors between any two local estimators are determined. These covariances, together with the local filters and their error covariance matrices are then merged to determine the optimal distributed matrix-weighted fusion estimator in the linear minimum variance sense.

EP1640: **Bayesian estimation of a GARCH(1,1) model with asymmetric Student$-t$ innovations**
Presenter: **Jacinto Martin Jimenez**, Universidad de Extremadura, Spain
Co-authors: Lizbeth Naranjo Albarran, Carlos Javier Perez, Amparo Rubio Leal
The Autoregressive Conditional Heteroscedastic (ARCH) model is a time series model that deals with changing volatility, exhibiting a non constant variance structure. The ARCH model often requires many parameters to adequately describe the volatility process of financial data. This extension is named the Generalized ARCH (GARCH) model. An extension of the GARCH(1,1) model under Bayesian methodology is proposed. In order to handle asymmetry and heavy tails properties in the volatility process, an asymmetric Student$-t$ (AST) distribution is used. The AST family can successfully handle both symmetric/asymmetric and heavy tails simultaneously. Even more, the distributions can fit each tails separately. The idea of using a scale mixture of normal representation of the AST distribution is aimed to derive efficient Markov chain Monte Carlo methods, which allow to explore the joint posterior distribution of the model parameters.

EP1642: **Asymptotic theory of elliptical moment based extreme value index estimators**
Presenter: **Yves Dominicy**, Universite libre de Bruxelles, Belgium
Co-authors: Matias Heikkila, Pauliina Ilmonen
We extend two extreme value index estimators, namely the moment estimator and the mixed moment estimator, to an elliptical setting. We will make use of the results previously developed regarding the so-called 'Multivariate Hill estimators', in order to define the new multivariate extreme value index estimators and to prove their asymptotic theory. These new multivariate extreme value index estimators for a regular varying elliptical random vector are based on the distance between a tail probability contour and the observations outside this contour. We also discuss the advantages and disadvantages of those estimators in comparison with the multivariate Hill estimators.

EP1684: **New developments of sparse PLS regressions**
Presenter: **Jeremy Magnanensi**, University of Strasbourg, France
Co-authors: Myriam Maumy-Bertrand, Nicolas Meyer, Frederic Bertrand
Processes based on the so-called Partial Least Squares (PLS) regression, which recently gained much attention in the analysis of high-dimensional genomic datasets, were recently developed to perform variables selection. Most of these processes rely on some tuning parameters that are usually determined by Cross-Validation (CV), which raises important stability issues. We developed a new dynamic bootstrap based PLS process for significant predictors selection, suitable for both PLS regression and its extension to Generalized Linear (GPLS) regression frameworks. It depends on a single tuning parameter, the number of components, which is determined by bootstrap, avoiding the use of CV. We also developed an adapted version of the Sparse PLS (SPLS) and SGPLS regression processes, using bootstrap for the determination of the numbers of components. We benchmarked their variable selection accuracy, as well as their stability concerning the tuning parameters and their predictive ability. This benchmarking is performed on simulations for PLS framework and on real microarray gene expression datasets for PLS-logistic classification.

EP1725: **Using tree-based sampling algorithms in hidden Markov models**
Presenter: **Dong Ding**, Imperial College London, United Kingdom
Co-authors: Axel Gandy
A recently proposed algorithm called divide-and-conquer sequential Monte Carlo splits a probabilistic model into small parts, samples the particles independently in these parts and subsequently merges the particles in a tree structure. A hidden Markov model is a stochastic process where the hidden process is a Markov process and the observations are independent given the hidden states. We investigate using the divide-and-conquer sequential Monte Carlo algorithm in such models. We investigate how it can be used for standard particle filtering, particle smoothing and sampling from the joint distribution of the hidden states given all the observations. We use two algorithms to generate and merge the samples. In both algorithms, we first run a standard bootstrap particle filter. The first approach uses the samples from the particle filter directly and the second approach constructs a parametric approximation. We study adaptive methods for improving the initial and the intermediate target distributions in the tree. We illustrate the performance in a simulation study.

EP1748: **Inference for censored mixture data applied to wood fibre length estimation**
Presenter: **Konrad Abramowicz**, Umea University, Sweden
Co-authors: Sara Sjostedt de Luna
Wood fibre (tracheid) length is an important trait targeted for genetic and silvicultural improvement. Such studies require large-scale non-destructive sampling, and accurate length determination. We consider data from cylindric wood samples (increment cores) of 5 mm diameter, considered to be a non-destructive sampling method. The sample contains both cut and uncut fibres and other cells, the so-called fines. The lengths are measured by an automatic fibre-analyser, which is not able to distinguish fines from fibres and cannot tell if a cell has been cut. The data thus come from a censored version of a mixture of the fine and fibre length distributions in the tree. Taking into account the censoring induced by the increment core, the fibre and fine length distributions in the tree may be estimated under lognormal assumptions via a recently proposed consistent EM algorithm. We build on these results to handle the more flexible Generalized Gamma distributional assumptions. Standard inferential procedures use weighted averages of all observed cell lengths in the core to obtain point estimates of the mean fibre length. Based on real data from Norway spruce trees, we compare the performance of these methods with true length distributions inferred from microscopy measurements.

EP1776: **Experimental design to discriminate between models**
Presenter: **Elvira Delgado-Marquez**, University of Castilla - la Mancha, Spain
Co-authors: Mariano Amo-Salas, Jesus Lopez-Fidalgo, Lenka Filova
During the discharge of a two-dimensional silo, the flow of grains through an opening is arrested if the size of the outlet is not large enough. In the outpouring of grains, jamming occurs due to the formation of an arch at the outlet. After breaking the arch, the grains fall until a new arch is formed. Several models have been proposed to explain this process. D-optimal designs have been computed for four models, two with 2 unknown parameters and two with 3 unknown parameters. In addition, using the methodology of KL-optimality, optimal designs for discriminating between models have been computed.

EP1730: **Optimal estimation and sampling allocation in survey sampling under a general correlated superpopulation model**
Presenter: **Ioulia Papageorgiou**, Athens University of Economics and Business, Greece
The aim is to contribute on the problem of sampling from autocorrelated populations. Assuming an autocorrelation among the population units we focus on deriving an efficient sample allocation and its resulting statistical inference about the population parameters. Both stages of sampling and inference incorporate the existing correlation. The optimal sample allocation is closely related with the type of correlation and therefore is a problem with no unique answer. The proposed methodology can cover any type of correlation function among the units. It is based on a continuous approximation of finite sums, is practically feasible and not computationally expensive. The result includes the sampling allocation that ensures optimal efficiency of the population parameters estimates and the expressions of the estimates and their mean square error. The gain in efficiency if the correlation type is taken into account is significant, and it is shown by a number of experiments with simulated data sets. An application in quality control with correlated measurements is also presented, aiming to exhibit the advantage of the proposed methodology over standard techniques that do not take into account the present correlation.

EP1742: **Analysis of kinematic data using functional-on-scalar linear models**
Presenter: **Lina Schelin**, Umea University, Sweden
Co-authors: Charlotte Hager
We present a distribution-free procedure for testing a functional-on-scalar linear model with fixed effects, based on the interval testing procedure. It does not only test the global hypothesis on the entire domain, but also selects the intervals where statistically significant effects are detected. The proposed tests are provided with an asymptotic interval-wise control of the family-wise error rate, i.e. the probability of falsely rejecting any interval of true null hypotheses. The methodology is applied to kinematic data from a long-term follow-up study after injury of the anterior cruciate ligament. This injury to the knee is common worldwide, and frequently occurring in sports. There is still no evidence which treatment (surgery in combination with physiotherapy or physiotherapy alone) that works best, especially in the long term. 3D-motion systems are used to collect information during functional assessments (e.g., jumps), in order to evaluate long-term treatment effects. The presented methodology is used to compare kinematic data from three groups of individuals, with the possibility to take individual-specific covariates into account. Using the methodology we can detect intervals during the jump where we have an effect of groups and/or covariates, adding valuable information to the commonly used traditional univariate analyses.

EP1495: **Sparse predictive modeling for bank telemarketing success using smooth-threshold estimating equations**
Presenter: **Yoshinori Kawasaki**, The Institute of Statistical Mathematics, Japan
Co-authors: Masao Ueki
We attempt to build and evaluate several predictive models to predict success of telemarketing calls for selling bank long-term deposits using a publicly available data from a Portuguese retail bank collected from 2008 and 2013. The data include multiple predictor variables, either numeric

or categorical, related with bank client, product and social-economic attributes. Dealing with a categorical predictor variable as multiple dummy variables increases model dimensionality and redundancy in model parameterization must be of practical concern. This motivates us to assess prediction performance with more parsimonious modeling. We apply contemporary variable selection methods with penalization including lasso, elastic net, smoothly-clipped absolute deviation, minimum concave penalty as well as the smooth-threshold estimating equation. In addition to variable selection, the smooth-threshold estimating equation can achieve automatic grouping of predictor variables, which is an alternative sparse modeling to variable selection and could be suited to a certain problem, e.g., dummy variables created from categorical predictor variable. Predictive power of each modeling is assessed by repeating cross-validation experiments or sample splitting, one for training and another for testing.

EP1542: **Functional data analysis approach of Mandel h and k statistics in interlaboratory tests**
Presenter: **Javier Tarrio-Saavedra**, Universidade da Coruna, Spain
Co-authors: Miguel Flores, Salvador Naya, Ruben Fernandez Casal

The Interlaboratory Tests (ILT) are defined as the statistical quality control process used to evaluate the precision and consistency of the test results obtained using a well-defined experimental procedure. They are performed using the same specific controlled material and tested by different laboratories. To identify the laboratories that provide results significantly different from the others, the use of h and k Mandel statistics is proposed by the ASTM E691. The h statistic measures the inter-laboratories consistency by comparing the replicate averages with the overall average: higher h absolute value involves less consistency. Moreover, the k statistic provides information about the intra-laboratory consistency by comparing the replicate standard deviations with respect to the repeatability standard deviation. But many types of experimental results in analytical chemistry are functional and using univariate tools some information could be obviated. Thus, Functional Data Analysis (FDA) approaches of h and k statistics are presented to deal with functional experimental data. Both functional statistics are estimated using bootstrap resampling and characteristic graphics are obtained. Results are compared with those obtained by the univariate techniques. Real and simulated data related to thermal analysis are obtained in order to know the performance of the proposed FDA h and k statistics.

EC1540: **Reliability of materials through the TTS R library**
Presenter: **Salvador Naya**, University of A Coruna, Spain
Co-authors: Antonio Meneses, Javier Tarrio-Saavedra

The TTS R library has been developed to estimate the behavior of materials under thermal and mechanical loads. It provides the prediction of viscoelastic properties such as modulus at short and long observation times through time temperature superposition (TTS) principle. TTS is used to estimate mechanical properties of linear viscoelastic materials from known properties at a reference temperature. The hypothesis that all curves present the same characteristic time variation and shape with temperature is assumed. Previous models are implemented to modulus curves obtained at different temperatures to estimate the shift factors. The master curve is obtained from shift factors and it provides viscoelastic properties estimates at times different from the experimental. In addition, a newer method for obtaining the shift factors based on the study of the derivatives of viscoelastic properties is presented and implemented in the TTS package. The resulting shift factors provide smooth master curves applying Bsplines.

EC1645: **Model selection problem in wavelet shrinkage and wavelet nonparametric estimation of correlated functional data**
Presenter: **Maria Pilar Frias Bustamante**, University of Jaen, Spain
Co-authors: Maria Dolores Ruiz-Medina

Wavelet-based estimation methodologies have been traditionally applied in the nonparametric setting, e.g., probability density estimation, in the resolution of filtering problems, and in the parametric setting, for definition of log-wavelet regression estimators, among others. However, the associated model selection problem has not been studied in depth yet. Particularly, special attention must be paid to the wavelet basis selection problem, and the resolution level selection problem (related with the so-called bandwidth parameter selection problem). The aim is to address these problems attending to the local regularity of the functional dataset analyzed, their spatial or temporal correlation, and the density in space or time of the functional observations. We illustrate the results obtained in terms of two functional datasets displaying very different local regularity properties and correlation structure, as well as with different spatial and temporal distributions, affecting in a decisive way the kind of interpolation methods to be applied. These two datasets are related to two fields of application: on the one hand, we address the wavelet-based shrinkage, and wavelet non-parametric estimation of brain cancer maps, and, on the other hand, we perform the wavelet shrinkage, and wavelet nonparametric estimation of ocean surface temperature maps. Final comparative conclusions are then conducted.

EP1625: **Evolution of HIV diagnoses in AIDS patients**
Presenter: **Ana Lara-Porras**, Granada, Spain

To analyze the evolution of HIV / AIDS is one of the most important aspects to consider in health policies to prevent this disease. The indicative systems of HIV infection are not easy to determine. The infected patient is introduced into a process in which cellular immunity is deteriorated gradually until appearing diseases indicative of AIDS as the final phase of the infectious process. It exists a large number of drugs that has managed to increase the life expectancy of HIV infected patients. Our interest is to know the influence of different features in the variable survival time with AIDS. This analysis will let determine variables that are influential to prolong the lives of AIDS patients. Variables that are registered by the Andalusian Register of AIDS cases are described. 5875 records between 1992-2013 are considered and the relations between them have been analyzed.

EP1850: **Data-adaptive estimation of time-varying spectral densities**
Presenter: **Anne van Delft**, Maastricht University, Netherlands
Co-authors: Michael Eichler

A data-adaptive approach for spectral density estimation of nonstationary processes. Estimation of time-dependent spectra commonly proceeds by means of local kernel smoothing. The performance of these nonparametric estimators depends however crucially on the smoothing bandwidths that need to be specified in both time and frequency direction. The objective is to construct local spectral density estimates where the respective smoothing kernels are iteratively adapted to the data at hand. The main idea, inspired by the concept of propagation-separation, is to describe the largest local vicinity of every design point in the time-frequency plane over which smoothing is justified by the data. Our method circumvents the problem of optimal bandwidth selection in the strict sense without imposing additional assumptions. The procedure permits full flexibility for the degree of smoothing and automatically adjusts for structural breaks in the time-dependent spectrum.

Saturday 12.12.2015 16:35 - 18:40 Parallel Session F – CFE-CMStatistics

CI018 Room Senate SPECIAL SESSION ON BOOTSTRAP INFERENCE Chair: Jean-Pierre Urbain

CI0665: **A discrete model for bootstrap iteration**
Presenter: **Russell Davidson**, McGill University, Canada

In an attempt to free bootstrap theory from the shackles of asymptotic considerations, the aim is to study the possibility of justifying, or validating, the bootstrap, not by letting the sample size tend to infinity, but by considering the sequence of bootstrap $p-$values obtained by iterating the bootstrap. The main idea is that, if this sequence converges to a random variable that follows the uniform $U(0,1)$ distribution, then the bootstrap is valid. The idea is studied by making the model under test discrete and finite, so that it is characterised by a finite three-dimensional array of probabilities. This device, when available, renders bootstrap iteration to any desired order feasible. It is used for studying a unit-root test for a process driven by a stationary $MA(1)$ process, where it is known that the unit-root test, even when bootstrapped, becomes quite unreliable when the $MA(1)$ parameter is in the vicinity of -1. Iteration of the bootstrap $p-$value to convergence achieves reliable inference except for a parameter value very close to -1. We then endeavour to see these specific results in a wider context, and try to cast new light on where bootstrap theory may be going.

CI0802: **Bootstrap inference for VAR models under rank uncertainty**
Presenter: **Stephan Smeekes**, Maastricht University, Netherlands
Co-authors: Lenard Lieb

In impulse response analysis using VAR models it is common to construct confidence intervals using bootstrap techniques. However, in many practical applications, uncertainty regarding the true (unknown) cointegration rank is typically ignored, and impulse responses and their confidence intervals are constructed as if the cointegration rank were known, whether the rank has been estimated or simply assumed to be equal to a certain rank. Recently many methods have been proposed to robustify impulse responses to an unknown rank or order of integration. These studies however do not consider how the bootstrap is affected when constructing confidence intervals. We therefore investigate how bootstrap inference for VAR models, such as used for impulse response analysis, is affected by a misspecified cointegration rank. We derive theoretical results on the asymptotic validity of the bootstrap in this setting, and analyze finite sample effects through Monte Carlo simulation. This allows us to quantify how serious the problems are for empirical work if the uncertainty regarding the rank is ignored. We also consider modifications of the bootstrap that provide better guard against misspecification of the cointegration rank. Finally our results are illustrated with an empirical example.

CI1602: **Dependent wild bootstrap for the empirical process and von Mises-statistics**
Presenter: **Michael H Neumann**, Friedrich Schiller University, Germany

Many important quantities in statistics can be written as a functional of the empirical process. Von Mises (V-) statistics appear as approximations of test statistics of Cramer-von Mises-type. It will be shown how modifications of the dependent wild bootstrap, which was originally introduced for smooth functionals of the sample mean, can be used for bootstrapping the empirical process and degenerate V-statistics of dependent random variables. Consistency of the bootstrap approximations is proved under minimal conditions.

CO438 Room G21A LARGE DIMENSIONAL PANEL MODELS Chair: Xun Lu

CO0188: **Determining the number of groups in latent panel structures with an application to income and democracy**
Presenter: **Xun Lu**, Hong Kong University of Science and Technology, China

A latent group panel structure where the number of groups is unknown and has to be determined empirically is considered. We propose a testing procedure to determine the number of groups. Our test is a residual-based LM-type test. We show that after being appropriately standardized, our test is asymptotically normally distributed under the null hypothesis of a given number of groups and has power to detect deviations from the null. Monte Carlo simulations show that our test performs remarkably well in finite samples. We apply our method to study the effect of income on democracy and find strong evidence of heterogeneity in the slope coefficients. Our testing procedure determines three latent groups among eighty two countries.

CO0658: **Estimation of principal functional coefficient models for longitudinal data**
Presenter: **Degui Li**, University of York, United Kingdom

The estimation of the functional coefficient longitudinal data models is studied. In order to achieve dimension reduction for the nonparametric functional coefficients and improve the estimation efficiency, we introduce a novel semiparametric estimation procedure which combines a principal component analysis of the functional coefficients and a Cholesky decomposition of the within-subject covariance matrices. Under some regularity conditions, we derive the asymptotic distribution theory for the proposed semiparametric estimators and show that the efficiency of the estimation of the (principal) functional coefficients can be improved when the within-subject covariance structure is correctly specified. Furthermore, we apply two approaches to consistently estimate the autoregressive coefficients in the Cholesky decomposition, which help avoid a possible misspecification of the within-subject covariance structure and ensure the efficiency improvement for the estimation of the (principal) functional coefficients. Some numerical studies including Monte Carlo experiments and an empirical application show that the developed semiparametric method works reasonably well in finite samples.

CC1093: **Semiparametric trending regression for unbalanced panel data with application to realized volatility**
Presenter: **Alev Atak**, City University London, United Kingdom

A methodology is outlined for developing a semiparametric panel data model to describe the realized volatility and the trend in monthly dataset of US equity returns by using the Center for Research in Security Prices (CRSP) while relinquishing the assumption of global stationarity. We allow the trend to evolve in a nonparametric way, with an unknown smooth function. While we first provide idiosyncratic trends for each individual i, we aim to test for the common trends assumption based on a measure of nonparametric goodness-of-fit test before imposing it. We propose a semiparametric profile likelihood approach to estimate the model. We assume an asymptotic framework in which T is large; but not necessarily N.

CC1680: **Generalized least squares estimation of panel with common shocks**
Presenter: **Marco Avarucci**, University of Glasgow, United Kingdom
Co-authors: Paolo Zaffaroni

The estimation of linear regression such as $Y_i = X_i\beta_{i0} + u_i$ is considered, where $Y_i = (y_{i1}, ...y_{iT})'$ is a $Tx1$ vector of dependent variables, X_i is a $T \times K$ matrix of regressor and β_{i0} are individual-specific parameters. The innovation $u_i = (u_{i1}, ..., u_{iT})'$ has a factor structure $u_i = Fb_i + e_i$, for a TxM matrix of latent factors $F = (f_1, ..., f_T)'$ with loadings b_i and $e_i = (e_{i1}, ..., e_{iT})'$ is a vector of idiosyncratic innovations. A factor structure in both the innovation u_i and the regressors X_i can make the ordinary least squares estimator inconsistent for the true regression coefficients. To overcome this problem, we propose a GLS-type estimator. The procedure can be summarized as follows: (i) Obtain the $T \times 1$ vector of residuals $\hat{u}_i$ by OLS. (ii) Construct the $T \times T$ variance covariance matrix $W = N^1 \sum_{i=1}^{N} \hat{u}_i\hat{u}_i'$. (iii) Compute the GLS estimator using the matrix W. We show that,

if T^2/N approaches zero for T,N diverging to infinity, the GLS estimator is consistent and asymptotically normal. This result is due to an important insight, namely the existence of a form of asymptotic orthogonality between the latent factor F and inverse of W. This result holds despite the inconsistency of the OLS and does not require a preliminary estimate of the factor or a priory knowledge of their number.

CC1784: **Canonical correlation analysis of panel VEC models**
Presenter: **Piotr Keblowski**, University of Lodz, Poland

The focus is on performance of the fully parametric system estimators of long-run relationships in the panel vector error correction framework that are based on eigenvalue problem. Therefore, we compare performance of Box-Tiao levels canonical correlation analysis with the classical approach proposed by Johansen in the panel data setting. The panel VEC model is considered with different sets of restrictions on the system's structure. Therefore, it is consecutively allowed that (i) cross-sectional dependence in the error terms occurs, (ii) there is interaction of the short-run dynamics between cross-sections, (iii) there is interaction of the error-correction terms between cross-sections. The results for the individual time-series analyses, where cross-sections are assumed to be independent, are related to the results of the panel analysis. It is showed that there is a trade-off between the dimensionality effect which is well known from the standard time-series analysis and efficiency gains, which are due to cross-sectional dependencies. However, if there is a common cointegration rank and significant cross-sectional relationships, then the results of the MLE of long-run parameters in panels usually outperform the results for the standard time-series analyses, where cross-sections are assumed to be independent. Moreover the performance of the MLE of long-run parameters in the panel VEC is enhanced if the cross-sections share the same long-run structure.

CO456 Room Woburn INFLATION ANALYSIS AND FORECASTING Chair: Till Strohsal

CO0226: **Analysis of aggregated inflation expectations based on the ECB SPF survey**
Presenter: **Maritta Paloviita**, Bank of Finland, Finland

The aim is to examine aggregated inflation expectations based on the ECB Survey of Professional Forecasters (ECB SPF). The focus of the analysis is on possible impacts of changing panel composition on aggregated short and long term point forecasts. We also investigate corresponding forecast uncertainties, which are based on subjective probability distributions. We compare changes in aggregated forecasts in the original unbalanced panel data with aggregated forecast changes based on a set of sub-panels of fixed composition. We also construct lower and upper bounds around aggregated forecast revisions. Our results indicate that the unbalanced panel data do not cause systematic distortions to aggregated survey information, but there are some minor differences between alternative survey aggregates, which are not necessarily non-negligible from the monetary policy point of view. We provide evidence that both micro and macro level analysis of the ECB SPF survey information is needed, especially in times of wide disagreement across forecasters and high levels of inflation uncertainty.

CO0216: **How oil price forecast errors impact inflation forecast errors**
Presenter: **Frederique Bec**, THEMA University of Cergy-Pontoise and CREST, France
Co-authors: Annabelle De Gaye

The aim is to propose an empirical investigation of the impact of oil price forecast errors on inflation forecast errors for two different sets of recent forecasts data: the median of SPF inflation forecasts for the U.S. and the Central Bank inflation forecasts for France. Mainly two salient points emerge from our results. First, there is a significant and dominant contribution of oil price forecast errors to the explanation of inflation forecast errors, whatever the country or the period considered. Second, the pass-through of oil price forecast errors to inflation forecast errors is multiplied by around 2 when the oil price volatility is large.

CO0320: **The time-varying degree of inflation expectations anchoring**
Presenter: **Till Strohsal**, Freie Universitaet Berlin, Germany
Co-authors: Rafi Melnick, Dieter Nautz

Well-anchored inflation expectations have become a key indicator for the credibility of a central bank's inflation target. Since the outbreak of the recent financial crisis, the existence and the degree of de-anchoring of U.S. inflation expectations have been under debate. An encompassing time-varying parameter model is introduced to analyze the changing degree of U.S. inflation expectations anchoring. Our model nests the two most common existing approaches as special cases. We confirm that inflation expectations have been partially de-anchored during the financial crisis. Yet, our results suggest that inflation expectations have been successfully re-anchored ever since.

CO1095: **A credit-based indicator for the risk of low inflation**
Presenter: **Roberta Colavecchio**, Hamburg University, Germany

We employ a credit-based early warning model in order to analyse the risk of a low inflation regime in the four major Euro area countries. The model specification allows for three different inflation regimes: Low, Medium and High inflation, with time-varying transition probabilities depending on a national credit aggregate. Using Bayesian techniques, we estimate the model with quarterly data from the early 1970s up to the end of 2014. Our analysis uncovers several country-specific features and suggests that, from 2011 on, the risks of a Low inflation regime have been increasing in Italy, Spain and, to a lesser extent in Germany while in France they have started to alleviate in the course of the last seven quarters of the sample. Moreover, credit growth appears to play a role in the assessment of the risk of entering a low inflation state: the inclusion of a credit indicator variable signals an increase in such risk, especially for Italy and Spain.

CO422 Room Bedford ECONOMETRICS OF ART MARKETS Chair: Douglas Hodgson

CO0313: **Efficiency of Italian opera houses: A stochastic frontier production function approach**
Presenter: **Sabrina Auci**, University of Palermo, Italy
Co-authors: Antonio Cognata

The empirical literature on the production of performing arts has mainly focused on cost functions. Studies have explored the cost structure of symphony orchestras, theatres and museums, mostly with the aim of finding evidence of scale economies. Until now no work has studied opera houses production or cost functions. Only recently there have been attempts in the use of a more suitable methodology and in finding evidence of the efficiency of performing arts institutions. The aim is to investigate efficiency of Italian opera houses using a stochastic frontier approach (SFA). The empirical analysis based on the concept of output maximization is performed on firm level unique database of 14 major Italian opera houses in the period 2001-2012. Dividing the error component into two aspects - the systematic and the noise components - the SFA allows to consider separately inputs of the production function, such as physical, labour and human capital from factors of the inefficiency model influencing the behaviour of opera houses. These latter factors represent the opera houses heterogeneity and show the influence on technical efficiency scores. Finally, we rank opera companies on the basis of the estimated technical inefficiency.

CO0326: **The relationship between artistic movements and artist careers: Evidence from individual-level hedonic regression**
Presenter: **Douglas Hodgson**, UQAM, Canada
Co-authors: John Galbraith, Christiane Hellmanzik

The literature on age-valuation profiles of artists has paid limited attention to the effects of membership in artistic movements. There are many reasons why membership in a movement can be important for the career dynamic of an artist. The relation between careers and movement membership has been previously studied by considering data on numbers of reproductions in art history books. The hedonic analysis of auction data in this area is limited, with results of regressions of pooled groups of artists being reported. Ideally, one would like to estimate individual-artist profiles relating valuation to date of production, and compare these with pooled profiles estimated for groups or movements to which the artists belong, to assess the relation between individual- and group-level price dynamics. Until recently, such an endeavour was rendered difficult by the small number of observations, compared to a large number of hedonic covariates, often available at the individual-artist level. But the successful application to this problem of recent dimensionality-reduction and model-averaging methods in the context of estimating individual age-valuation profiles suggests the utility of applying the same approach to estimating individual profiles in the context of movement membership. We thus apply these methods to a large data set on auction prices for major modern painters.

CO0583: **A first approach to pricing on the painting secondary market in the Argentine Republic**
Presenter: **Carolina Boufflet**, University of Salvador and National Ministry of Industry, Argentina
Co-authors: Marcos Leonardo Chamorro

The ongoing discussion of methods for predicting prices on the painting secondary market in the Argentine Republic is advanced. The research is bounded to 3542 sales transactions (not repeated) conducted by auction houses located in the city of Buenos Aires in the period from January to December 2014. Given the specificity for works of arts demand, one approach to the demand from the Theory of characteristics was proposed. The estimation of a hedonic price index was performed by translog regressions. According to preliminary results, Argentines market of art shows the existence of "Masterpiece effect", the "Main auctions effect" and a "batch effect" in relation to the sales strategy conducted by auction houses. The model threw an $R2$ of 65% and this pose the challenge of running a model with 2SLS instrumental variable. Artist index was taken as instrumental variable and elements of his biography and career as dependent variables that were not explicitly included in the previous model. Also a model of latent or hidden variables was developed to find out that the artwork has Its own life beyond the artist. The goodness of fit was tested to analyze the strengths and weaknesses (or shortcomings) of each model.

CO0669: **From quality to utility, an empirical study of artist reputations impact on contemporary art market price**
Presenter: **Simeng Chang**, Erasmus University Rotterdam, Netherlands

The aim is to analyze how an artists reputation determines the market price of his or her artworks. Reputation is a proxy of quality as well as art price. In other words, reputation signals the artistic value of the artwork and generates the economic value. In the art field, an artistic reputation is built through the process where the artistic value is recognized and enhanced by getting constantly attention from the art experts. In the art market, consumers aesthetic utility is confirmed by assessing the quality from artistic reputation. The aesthetic utility is further enhanced by the added value derived from market reputation. Therefore, the price of an artwork with higher reputation is higher since the quality uncertainty is reduced and the consumer utility is raised. Using the reputation quantified tool, Artfacts artist ranking, and auction data, we are intended to empirically identify the reputation effect on the art market price.

CC0570: **Historic art exhibitions and modern day auction results**
Presenter: **Christiane Hellmanzik**, University of Hamburg, Germany

Historic art exhibitions are used in order to investigate the impact of the contemporary success of artistic careers on today's auction prices of modern paintings. That is, if an artist's work was displayed in a historic art exhibition in a given year, paintings dated from this year fetch higher prices at auction today. This can be attributed to two effects: artists who participated in such shows were already acknowledged as superstars contemporaneously and participants in art exhibitions benefited from a longer-lasting career boost as reflected by positive mark-ups on paintings made in the years following a show. For both channels participation in historic art exhibitions is a strong quality signal for today's art buyers. The study is based on a global sample of 273 'superstars' of modern art born between 1800 and 1945, 34,141 auction results of paintings and participation in important historic art exhibitions.

CO394 Room Torrington COMMODITY MARKETS: PRICING AND TRADING Chair: Ana-Maria Fuertes

CO0318: **The determinants of convenience yields**
Presenter: **Marcel Prokopczuk**, Leibniz University Hannover, Germany
Co-authors: Yingying Wu

The determinants of convenience yields across a broad range of commodities are investigated. We find that the convenience yields of commodities are exposed to both commodity-specific and systematic factors, but to a different extent. The difference in explanatory power of these factors for each commodity sheds light on the heterogeneity of commodity markets. One main difference between commodity sectors lies in their different sensitivities towards the state of the economy.

CO1001: **The skewness of commodity futures returns**
Presenter: **Joelle Miffre**, EDHEC Business School, France
Co-authors: Ana-Maria Fuertes, Adrian Fernandez-Perez, Bart Frijns

The relationship between skewness of the distribution of past returns and expected returns in commodity futures markets is explored. Both time-series tests and cross-sectional tests indicate that more positively skewed commodities accrue significantly lower mean excess returns. Sorting a cross-section of commodities by their past skewness, we demonstrate that a fully collateralized portfolio that buys commodities with the most negative skewness and shorts commodities with the most positive skewness earns an excess return of 8.01% a year. A commodity pricing model that utilizes as risk factors the excess returns of a long-only equally-weighted portfolio of all commodities, alongside term structure, momentum, and hedging pressure portfolios yields a significant alpha of 6.58% obtained which indicates that the profitability of skewness portfolios is not a mere manifestation of backwardation and contango risk. Skewness risk may uniquely relate to the preferences of investors for lottery-like commodity futures. The findings are robust to transaction costs, liquidity considerations and sample periods.

CO0975: **The earnings-price ratio and predictability of earnings in the dry bulk shipping industry**
Presenter: **Nikos Nomikos**, Cass Business School, United Kingdom
Co-authors: Ioannis Moutzouris

We examine second-hand vessel prices, net earnings, and returns on capital in the dry bulk shipping industry. We demonstrate that the bulk of variation in net earnings-price ratios reflects varying expected net earnings growth. Furthermore, we contribute to the literature by examining a forward-looking definition of the net earnings-price ratio, and extending the variance decomposition framework to assets with limited economic lives. Our results strongly indicate that shipping net earnings-price ratios negatively forecast future net earnings growth. In addition, there is no statistical evidence that the net earnings-price ratio is negatively related to future returns. These results are in contrast to the recent empirical asset pricing literature in the U.S. equity and housing markets. Importantly, however, our findings agree with recent results obtained from global equity

markets. In line with this international evidence, we believe that predictability of earnings growth is driven by the extreme volatility of shipping earnings.

CC0363: **Comovement and financialization in the commodity market**
Presenter: **Matteo Bonato**, University of Johannesburg, Switzerland
Co-authors: Luca Taschini
The boom-and-bust that characterized the commodity market and resulted in the 2008 crash highlighted the effect of the financialization as one of the drivers of increased comovement of commodities returns. We build on the recent literature on comovement and borrow the concept of 'index inclusion' as describing the financialization process. This allows us to distinguish between two views of return comovement: the traditional view, in which this is explained by comovements in news about fundamental value, and the alternative view, in which excess comovement is attributed to frictions or traders' sentiment. As a result of the financialization we find that for index commodities the beta with the commodity index increases. This result holds also for some non-index commodities, although the increase in beta is less prominent, but not for Energy commodities. This indicates that the financialization affected all (non Energy) commodities while Oil and the other Energy commodities as a group remained unaffected. This may signal that for this class the financialization process started earlier. We extend our analysis to account for high-frequency returns dynamics by means of the so called realized betas and reach similar results. Our finding cannot be explained by the fundamental-based view which considers the commodity price bubble and crash as solely explained as driven by fundamentals. We therefore provide new evidences supporting the friction or sentiment based view explanations.

CC1650: **On hedgers' speculation: Evidence from commodities futures markets using mixed-frequency data**
Presenter: **Yannick Le Pen**, Universite Paris Dauphine, France
Co-authors: Marie Bessec, Benoit Sevi
In financial markets, it is common to distinguish between hedgers, who take positions in futures contracts to reduce their risk, and speculators, who engage in futures markets to benefit from a risk premium. The most standard view commonly assumes that hedgers do not speculate, i.e. their positions are not influenced by market prices. However, early equilibrium models of commodidity markets support the view that hedgers respond to price changes. From these two competing theories, it appears that the issue of hedgers sensitivity to commodity prices ultimately resembles an empirical question. To answer this question, we investigate the response of commercial and non-commercial categories of traders to changes in futures prices for a set of commodities quoted on the NYMEX-CME. We use daily commodity returns and weekly trader's positions published by the CFTC. We resort to mixed-data sampling (MIDAS) regressions to use higher frequency variables (prices) as explanatory variables for lower frequency variables (positions of traders). The number and coefficients of significant lagged returns are endogenous. Our empirical analysis show that hedgers do react to price changes, with the exception of several energy commodities. Our results are robust to changes in volatility.

CO548 Room Jessel BANKS AND THE MACROECONOMY IN THE CONTEXT OF STRESS TESTING Chair: Ching-Wai Jeremy Chiu

CO0332: **The rate elasticity of retail deposits in the United Kingdom: A macroeconomic investigation**
Presenter: **Ching-Wai Jeremy Chiu**, Bank of England, United Kingdom
The behaviour of major banks' household deposit funding in the United Kingdom is quantitatively studied. We estimate a panel of Bayesian vector autoregressive models on a unique dataset compiled by the Bank of England, and identify deposit demand and supply shocks, both to individual banks and in aggregate, using micro-founded sign restrictions. Based on the impulse responses, we estimate how much banks would be required to increase their deposit rates by to cover a deposit gap caused by funding shocks. Banks generally find it costly to bid-up for deposits to cover a funding gap in the short-run. The elasticity of household deposits with respect to the interest rate paid are typically of the order of 0.3, indicating that retail deposits are rate-inelastic. But this varies across banks and the types of shock conditioned on. We also show evidence that banks are more vulnerable to deposit supply shocks than deposit demand shocks. These results carry important implications to policy makers who conduct stress-testing analyses.

CO0306: **A top-down approach to stress-testing banks**
Presenter: **Pavel Kapinos**, FDIC, United States
Co-authors: Oscar Mitnik
The aim is to propose a simple, parsimonious, and easily implementable method for stress-testing banks using a top-down approach that captures the heterogeneous impact of shocks to macroeconomic variables on banks' capitalization. Our approach relies on a variable selection method to identify the macroeconomic drivers of banking variables as well as the balance sheet and income statement factors that are key in explaining bank heterogeneity in response to macroeconomic shocks. We perform a principal component analysis on the selected variables and show how the principal component factors can be used to make projections, conditional on exogenous paths of macroeconomic variables. We apply our approach, using alternative estimation strategies and assumptions, to the 2013 and 2014 stress tests of medium- and large-size U.S. banks mandated by the Dodd-Frank Act, and obtain stress projections for capitalization measures at the bank-by-bank and industry-wide levels. Our results suggest that bank responses to shocks are indeed heterogeneous, and that while capitalization of the U.S. banking industry has improved in recent years, under reasonable assumptions regarding growth in assets and loans, the stress scenarios can imply sizable deterioration in banks' capital positions.

CO0909: **Choosing stress scenarios for systemic risk through dimension reduction**
Presenter: **Matt Pritsker**, Federal Reserve Bank of Boston, United States
Current regulatory stress-tests ensure the banking system is well capitalized against the scenarios considered in the test, but it is unclear how well capitalized the banking system will be against other plausible scenarios. A methodology is proposed for choosing a regulatory stress-scenario based on measures of systemic risk. Under certain regularity conditions, when the banking system is well capitalized against the chosen scenario, then systemic risk is low, i.e. the banking system will be well capitalized against the other plausible scenarios that could affect it with high probability. The stress-scenario is chosen by using dimension reduction techniques that select variables and then create factors based on the variables' and factors' ability to explain systemic risk. The main result shows that under appropriate regularity conditions stress-scenarios can be chosen based on movements in the systemic risk factors, and that doing so can approximately achieve systemic risk objectives. Under some conditions the methodology also shows that stress-tests and capital injections based on a small number of scenarios cannot alone attain the systemic risk objective, indicating other steps may be needed. The methodology should be especially valuable if regulatory stress-testing continues to rely on a small number of stress scenarios.

CO1003: **A stress testing framework with interactions between solvency and liquidity risks and macro-financial linkages**
Presenter: **Tak Chuen Wong**, Hong Kong Monetary Authority, China
Co-authors: Cho-hoi Hui, Kelvin Ho, Edward Tan
A macro-stress testing framework is developed that incorporates interactions between solvency risk, funding and market liquidity risks, and macro-financial linkages. Specifically, the framework simulates macroeconomic shocks that increase solvency risk of banks, which in turn determines endogenously the timing and extent of deposit runs. Downward spirals between funding and market liquidity risks are incorporated in the framework, as in response to deposit outflows, banks are assumed to buffer the liquidity risk by selling financial assets. The resulting fall in prices of

common assets held by banks (e.g. corporate debts) deteriorates the net worth of banks and thus exacerbating deposit outflows. For macro-financial linkages, the framework empirically relates the severity of macroeconomic shocks to the aggregate supply of bank credit, which in turn is jointly determined by the simulated solvency and liquidity conditions of banks. The framework is applied to the Hong Kong banking sector as an example. A counter-factual analysis is conducted to analyze how Basel III would improve banks resilience to such interactions of risks.

CO0789: **Systemwide commonalities in market liquidity**
Presenter: **Mark Flood**, Office of Financial Research, United States
Co-authors: John Liechty, Tom Piontek
We calculate daily invariant price impacts to assemble a panel of liquidity measures for equity, corporate bond, and futures markets. We estimate hidden Markov chains for three latent regimes - high, medium, and low price-impact - and use Markov chain Monte Carlo to identify commonalities liquidity at the systemwide level. For the equities subpanel, we test whether a collection of daily summary time series recovers the liquidity dynamics. This allows an economically meaningful attribution of the liquidity states and meaningful predictions of liquidity disruptions as far as 15 trading days before the 2008 crisis.

CO388 Room Chancellor's Hall STRUCTURE IN MULTIVARIATE AND HIGH DIMENSIONAL TIME SERIES Chair: Manfred Deistler

CO0366: **Priors for the long run**
Presenter: **Michele Lenza**, European Central Bank, Germany
Co-authors: Domenico Giannone, Giorgio Primiceri
We propose a class of prior distributions that discipline the long-run behavior of Vector Autoregressions (VARs). These priors can be naturally elicited using economic theory, which provides guidance on the joint dynamics of macroeconomic time series in the long run. Our priors for the long run are conjugate, and can thus be trivially implementedusing dummy observations and combined with other popular priors. In VARs with standard macroeconomic variables, a prior based on the long-run predictions of a wide class of dynamic stochastic general equilibrium models yields substantial improvements in the forecasting performance.

CO1068: **Estimation of VAR systems from mixed-frequency data: The stock and the flow case**
Presenter: **Lukas Koelbl**, Vienna University of Technology, Austria
Co-authors: Alexander Braumann, Elisabeth Felsenstein, Manfred Deistler
Estimation and properties of estimators of VAR systems are considered in the case of observations with different sampling rates, the so-called mixed-frequency (MF) observations. Only the case where the output variable can be separated into a fast (high-frequency) and a slow (low-frequency) component will be given attention. It is assumed that the underlying system generates the output at each time point, the so-called high grid, however, the output of the slow component is only observed at an integer multiple of the high grid. As mentioned above, the asymptotic behavior of estimators of autoregressive systems is a central theme. The main focus is on estimators that are based on the extended Yule-Walker (XYW) equations as well as on (Gaussian) maximum likelihood type estimators based on the EM algorithm. Two cases for the slow component are considered: the stock and the flow case. In addition, the XYW estimator and the generalized method of moments estimator (GMM) are discussed and it is shown that they are asymptotically normal under certain assumptions. Therefore, a generalization of Bartlett's formula for the mixed-frequency case is required. As shown by examples, the GMM estimator is, in general, not efficient. Finally, the loss of information due to mixed-frequency data when compared to the high-frequency situation as well as the gain of information when using mixed-frequency data relative to low-frequency data is discussed.

CO0897: **A Bayesian approach to sparse plus low rank network identification**
Presenter: **Alessandro Chiuso**, University of Padova, Italy
Co-authors: Mattia Zorzi
We consider modeling an (high-dimensional) time series in terms of a structured model encopassing a latent (low-rank) and a sparse component. This corresponds to describing the observed time series through a Bayesian network where nodes are time series and edges encode Granger causality conditions. This network is composed of two layers: a "visible" layer with observed variables and an hidden layer (where the number of hidden nodes is to be estimated) encoding latent variables. The visible layer has sparse connections. We propose a Bayesian approach, based on Gaussian Regression and Sparse Bayesian Learning ideas, to estimate this model while guaranteeing a small number of hidden nodes as well as a sparse connection pattern between visible nodes.

CO1253: **Measuring global financial connectedness**
Presenter: **Michael Binder**, Goethe University Frankfurt, Germany
Co-authors: Soroosh Soofi Siavash
The global financial crisis has made clear how critical it is to properly measure and model the interconnectedness of financial institutions. To date, limited work has been done on financial connectedness in a global setting. This may be due to the challenges of estimating a network model with a very large number of nodes. We build on recent methodological advances in the Global VAR literature, and proposes a new global network model in which cross-country dependencies are not only due to global shocks, but are also due to structural dependencies within blocks of countries. Our model reflects how an increase in the number of large institutions within some of the sectors coupled with the rise of financial globalization led to some institutions effectively becoming common lenders and borrowers in the global financial system. From a network perspective, some of the nodes became more central, and this modified network structure changed the stability and resilience properties of the global financial system. Evaluation of our model using weekly asset returns of the major financial sectors in 43 industrial and emerging-market countries for the time period from July 1998 to June 2014 shows that inclusion of the commonalities within blocks of countries significantly improves performance of the large-dimensional VAR model. Our model suggests adopting a mixture of regional and global policies towards regulation.

CO1252: **Fitting latent variable models to multivariate time series**
Presenter: **Michael Eichler**, Maastricht University, Netherlands
In systems that are affected by latent variables conditional independences are often insufficient for inference about the structure of the underlying system. One common example is a system in which four observed variables X_1, X_2, X_3, and X_4 are conditionally independent given a fifth unobserved variable Y. While there are no conditional independences among the observed variables, they must satisfy the so-called tetrad constraints. In the time series case, these can be expressed in terms of the spectral matrix of the observed variables. We discuss how these constraints can be used to fit time series models that involve latent variables. In particular, we consider so-called latent variable models in which the observed variables are independent given the latent variables while the latent variables are interconnected.

CO484 Room Athlone FISCAL POLICY Chair: Evi Pappa

CO0611: **Sovereign risks and fiscal fundamentals in the Eurozone**
Presenter: **Nora Traum**, North Carolina State University, United States
Co-authors: Huixin Bi
Using Bayesian methods, a nonlinear general equilibrium model is estimated for various Eurozone nations. The model incorporates a fiscal limit, representing the maximum debt-to-GDP ratio that the government is economically able to service, and quantifies the various transmission channels of sovereign risk across nations. We find sizeable differences in the fiscal limit across countries due to country-specific features.

CO1248: **State capacity and pro-cyclical fiscal policy**
Presenter: **Francesco Pappada**, Banque de France, France
Co-authors: Yanos Zylberberg
Using data for 70 countries spanning 30 years, we show that, in countries with imperfect tax enforcement, the capacity to collect tax revenues is pro-cyclical and responds markedly to the level of taxes. We then incorporate this tax compliance channel into a standard model of sovereign debt in which a benevolent government can use fiscal policy or debt as consumption-smoothing instruments. While the first-best fiscal and debt policies should be counter-cyclical - e.g. deficit should increase in recessions through tax cuts - we show that, in the presence of a "debt-ceiling" constraint, they may be pro-cyclical. Consistent with the data, these pro-cyclical fiscal policies are much more likely to be observed in countries with low and very responsive tax compliance.

CO1260: **A model of market and political power interactions for Southern Europe**
Presenter: **Vanghelis Vassilatos**, Athens University of Economics and Business, Greece
Co-authors: Tryphon Kollintzas, Dimitris Papageorgiou
In recent years the growth pattern of most Southern European countries has been disturbed, as those countries are suffering from economic crises that go beyond the usual business cycle. We develop a dynamic general equilibrium model of market and political power interactions that explains this growth reversal. The model is a synthesis of the insiders-outsiders labor market structure and the concept of an elite government. Outsiders form a group of workers that supply labor to a competitive private sector. And insiders form a group of workers that enjoy market power in supplying labor to the public sector and influence the policy decisions of government, including those that affect the development and maintenance of public sector infrastructures.

CO1318: **Fiscal consolidation in a disinflationary environment: Price-based versus quantity-based measures**
Presenter: **Evi Pappa**, EUI and UAB, Italy
Co-authors: Eugenia Vella, Rana Sajedi
The aim is to understand the specificities of how inflation developments are relevant to the achievement of fiscal policy targets. We examine alternative consolidation strategies by focusing on the public wage bill under different inflation environments. We propose a theoretical model through which we can study the differential effects of quantity-based and price-based consolidation measures. In particular, we consider a New-Keynesian model with nominal rigidities in the form of monopolistic retailers facing price-stickiness. In order to build a complete model of the labor market, we incorporate both search and matching frictions, leading to involuntary unemployment, and an endogenous labor force participation decision, leading to voluntary unemployment. Finally, to study the effects of the public wage bill, we incorporate a public sector in the form of a public firm that hires public employees to produce a public good that is used by private firms.

CO1268: **Structural tranformation and the obesity epidemic: Growth and taxation**
Presenter: **Evangelos Dioikitopoulos**, Kings College London, United Kingdom
We unify existing theories and empirical evidence on the origins of obesity and examine the effects of fiscal policy on the dynamic evolution of weight. We build a dynamic general equilibrium growth model, with two sectors, one producing food and the other producing a composite consumption good. Weight is a function of rational choice as well as labor allocation between the two sectors. By estimating utility from weight and calibrating the US economy we show that (i) technological advances in agriculture decrease food prices and increase weight but not necessarily through higher food consumption but through increases weight levels through lower food prices; (ii) steady state food consumption increases due to a price substitution effect but weight increases due to higher calorie expenditure; (iii) increasing taxation on food decreases food consumption and weight levels in equilibrium. Labor reallocation towards the less sedentary sector on one hand and higher income on the other function as contradictory forces.

CO508 Room Court REGIME CHANGE MODELING IN ECONOMICS AND FINANCE I Chair: Willi Semmler

CO0637: **A solution technique for regime switching DSGE models with occasionally binding constraints**
Presenter: **Gary Anderson**, Federal Reserve Board, United States
It has been recently shown how to reliably compute global approximations for time-invariant solutions for a wide class of nonlinear rational expectations models. It has also been shown how to apply these techniques to compute rational expectations solutions for models subject to both regime changes and occasionally binding constraints. The technique recursively computes solutions that honor the constraints for successively longer horizons. The technique also provides a metric for determining apriori how long the horizon must be for accurate determination of the current state vector. The solutions computed by the technique accommodate the possibility that model trajectories can depart from and re-engage the constraints as well as transition between various regimes.

CO1328: **Destabilising effects of bank overleveraging on economic activity**
Presenter: **Willi Semmler**, New School for Social Research, United States
Co-authors: Marco Gross
We investigate the consequences of overleveraging and the potential for destabilizing effects from financial and real-sector interactions. In a theoretical framework we model overleveraging and demonstrate that - in the presence of regime-dependent macro feedback relations - a highly leveraged banking system can lead to unstable dynamics and downward spirals. A Threshold Mixed-Cross-SectionGlobal Vector Autoregressive (T-MCS-GVAR) model will be the basis for the empirical analysis. The threshold-switching component of the model aims to make the relationship between credit and real activity dependent on the extent to which observed leverage deviatesfrom optimal leverage in the banking systems. Our regime-constellation-dependent impulse response simulations suggest that the relation between credit supply and real activity is characterised by significant nonlinearities. The farther observed leverage in the banking system exceeds optimal leverage, the more detrimental is the effect of a deleveraging shock on credit supply and economicactivity.

CO1107: **An empirical investigation of sovereign CDS market**
Presenter: **Rosella Castellano**, University of Macerata, Italy
Co-authors: Luisa Scaccia
Since 2007, credit markets have witnessed a repricing of credit risk that has affected all sectors. This turbulence reached its peak with the collapse of Lehman Brothers in September 2008, causing large state interventions to control systemic risk and its negative consequences. The Lehman Brothers event led also to a severe repricing of credit risk of developed countries sovereigns. Especially in the Euro area, sovereign debt markets came under large stress in 2010, causing large sell offs of risky assets due to flight to safety episodes, and upward jumps in CDS quotes. The 2010 sovereign debt crisis has focused scholars and policy makers attention to the role of financial investors activities in CDS market. In particular, some studies have suggested that speculative attitudes among investors could destabilize market quotes and create excessive volatility. The main goal is to contribute to the discussion on this issue by exploring via Markov Switching Models when the data generating process entered into a high volatility state. These evidences are useful to determine whether the severe repricing of credit risk of developed countries occurred at the same time and regardless of macroeconomic fundamentals dynamics.

CO0814: **Identifying the stance of monetary policy at the zero lower bound: A Markov-switching estimation**
Presenter: **Manuel Gonzalez-Astudillo**, Board of Governors of the Federal Reserve System, United States
An econometric technique is proposed to estimate a Markov-Switching Taylor rule subject to the zero lower bound of interest rates. We show that incorporating a Tobit-like specification allows to obtain consistent estimators. More importantly, we show that linking the switching of the Taylor rule coefficients to the switching of the coefficients of an auxiliary uncensored Markov-switching regression improves the identification of an otherwise unidentifiable prevalent monetary regime. To illustrate the proposed estimation technique, we use U.S. quarterly data spanning 1960:1-2013:4. The chosen auxiliary Markov-switching regression is a fiscal policy rule where federal revenues react to debt and the output gap. Results show that there is evidence of policy co- movements with debt-stabilizing fiscal policy more likely accompanying active monetary policy, and vice versa.

CO0596: **Catching-up, leapfrogging, and falling-back in economic growth: A nonparametric approach**
Presenter: **Joachim Schnurbus**, University of Passau, Germany
Co-authors: Harald Haupt, Joachim Schnurbus, Willi Semmler
Classical growth convergence regressions fail to account for various sources of heterogeneity and nonlinearity. Recent contributions advocating nonlinear dynamic factor models remedy those problems by allowing for club-specific convergence paths. Unfortunately and similarly to statistical clustering methods, those results are sensitive to choices made in the clustering mechanism. We improve existing clubbing algorithms while providing an economic rationale for dependence and heterogeneity of number, size, and composition of convergence clubs over time. We propose a nonparametric strategy for tackling neglected heterogeneity and nonlinearity jointly while alleviating the problem of underspecification of growth convergence regressions. Furthermore, based on the club-specific transition paths derived from the factor model, we are able to estimate convergence effects not only on club- but on country-level. The proposed approach is illustrated using a current Penn World Table data set. We find empirical evidence for leapfrogging and falling-back of countries over time. Guise and degree of nonlinearities in convergence regressions also differ substantially over time. Furthermore, some countries on club-based convergence paths exhibit, in contrast to their fellow club members, insignificant convergence effects.

CO532 Room Bloomsbury FINANCIAL RISK Chair: Toshiaki Watanabe

CO0678: **Combining classification tree analysis with network grouping of Japanese stock returns**
Presenter: **Takashi Isogai**, Bank of Japan, Japan
A set of classification trees is built that provide sorting rules to reproduce the stock groups identified by correlation clustering of the Japanese stock returns. The clustering is achieved by hierarchical network division by the modularity maximization algorithm of complex networks theory. We try to link the clustering results that are based on the stock price data with non-price external data in order to explore how the hierarchical division process can be explained by other categorical and numerical variables. Various non-price data including price performance data and sector classification are examined as effective variables to explain the splits of the stock groups. Variables with a high level of relative importance scores are identified; specifically, the market capitalization and price book-value ratio are included as significantly important variables. The selected variables seem to be consistent with variables included in the standard stock price model such as Fama-French factor model. Some other variables are also identified as ones that clarify the properties of the Japanese stock market. Further, variables that represent local features of the Japanese stock market are also detected. The classification tree analysis method can also be applied to find the closest group of stocks even for stocks that have limited price data due to low liquidity.

CO0844: **Option implied volatility of JGB using American option prices**
Presenter: **Kosuke Oya**, Osaka University, Japan
The option implied volatility is useful financial device for risk management, economic forecasting, and also used as an alert for the market turmoil. The standard volatility index such as the CBOE volatility index (VIX) requires European option prices written on the underlying asset. Japanese volatility index (VXJ) and Nikkei VI are published for Japanese stock market. These indexes are based on Nikkei 225 option prices. Despite attracting much attention, the option implied volatility index for Japanese government bond market is not available since the option traded on the bond market is not European exercise style, but American style. Several approaches to obtain the implied volatility based on American option prices have been proposed. One simple approach is to use the extracted European option prices from American option prices by binomial tree. Once European option prices are extracted, the option implied volatility can be obtained by usual manner. The extracted European option prices approach is applied for the option implied volatility index for Japanese government bond and then examined its property.

CO0837: **Corporate bond spreads and investor risk appetite**
Presenter: **Teruko Takada**, Osaka City University, Japan
Co-authors: Yasutomo Tsukioka
A measure of investor risk appetite based on corporate bond spreads is proposed. Global risks in financial markets include liquidity risks, credit risks, and volatility risks, which are reflected in our proposed measure. Based on daily data of US financial markets from 1995 to 2013, the level of investor risk averseness is measured and the differences with other candidate measures are analysed. Moreover, the effect of monetary policy on stock markets via investor risk appetite and corporate bond markets is investigated. Dependence structures among key factors are nonparametrically visualised using pointwise mutual information estimation.

CO0840: **Asymptotic inference for common factor models in the presence of jumps**
Presenter: **Yohei Yamamoto**, Hitotsubashi University, Japan
We investigate the effects of infrequent and large jumps on asymptotic inference for large-dimensional common factor models. We first derive the upper bound of jump magnitudes with which the standard asymptotic inference goes through. Second, we propose a jump-correction method based

on a series-by-series outlier detection algorithm without accounting for the factor structure. Finally, we propose a test to investigate whether the jumps at a common date are independent outliers or are of factors.

CO0838: **Empirical comparison of several skew-*t* copulas**
Presenter: **Toshinao Yoshiba**, Bank of Japan, Japan

In financial portfolio risk management, Student-*t* copula is frequently used to capture the tail dependence of risk factors. Azzalini-Capitanio (AC) and Generalized Hyperbolic (GH) skew-*t* copulas are considered to incorporate asymmetric tail dependence of risk factors. Two problems are indicated in estimating the parameters of the skew-*t* copulas by maximizing the log-likelihood for pseudo observations. The first problem is the calculation speed of the log-likelihood function including univariate skew-*t* quantile functions. The second problem is the positive semi-definiteness of the correlation matrix in the parameters. Having solved the two problems, the estimated parameters of the AC skew-*t*, GH skew-*t*, Student-*t*, skew-Normal, Normal copulas are compared for the daily returns of three major stock indices: the Nikkei225, S&P500, and DAX. The asymmetric tail dependence is examined both for the unfiltered returns and for the filtered returns by GARCH and EGARCH models for the observation period of recent five-year and ten-year data. It shows that AC skew-*t* copula with the same skewness is selected by AIC for both unfiltered and filtered returns. Regarding ten-year data, the AC skew-*t* copula is selected even by BIC for both unfiltered and filtered returns.

CO402 Room Holden QUANTIFYING SOCIAL MEDIA IMPACT ON FINANCIAL DYNAMICS Chair: Giacomo Livan

CO0686: **Theory and methods in using text data for economic prediction: The UCL relative sentiment shift approach**
Presenter: **David Tuckett**, UCL, United Kingdom

Directed Algorithmic Text Analysis is introduced. Its usefulness is illustrated for nowcasting and forecasting the economy and for assessing financial fragility. The method relies on Conviction Narrative Theory, a social-psychological approach to decision-making under uncertainty which focuses the analysis of text documents on specific emotions to detect relative sentiment shifts - RSS. Using Reuters News articles for 1996-2015 RSS out-performs current forecasting methods for the US economy.

CO0952: **Coupling news sentiment with web browsing data predicts intra-day stock prices**
Presenter: **Giacomo Bormetti**, Scuola Normale Superiore Pisa, Italy
Co-authors: Gabriele Ranco, Ilaria Bordino, Guido Caldarelli, Fabrizio Lillo, Michele Treccani

The new digital revolution of big data is deeply changing our capability of understanding society and forecasting the outcome of many social and economic systems. Unfortunately, information can be very heterogeneous in the importance, relevance, and surprise it conveys, affecting severely the predictive power of semantic and statistical methods. We show that the aggregation of web users' behavior can be elicited to overcome this problem in a hard to predict complex system, namely the financial market. Specifically, we show that the combined use of sentiment analysis of news and browsing activity of users of Yahoo! Finance allows to forecast intra-day and daily price changes of a set of 100 highly capitalized US stocks traded in the period 2012-2013. Sentiment analysis or browsing activity when taken alone have very small or no predictive power. Conversely, when considering a news signal where in a given time interval we compute the average sentiment of the clicked news, weighted by the number of clicks, we show that for more the 50% of the companies such signal Granger-causes price returns. Our result indicates a "wisdom-of-the-crowd" effect that allows to exploit users' activity to identify and weigh properly the relevant and surprising news, enhancing considerably the forecasting power of the news sentiment.

CO0885: **From social media to endogenous activity: Their effects on Bitcoin price bubbles and user adoption**
Presenter: **Claudio Tessone**, University of Zurich, Switzerland
Co-authors: David Garcia, Pavlin Mavrodiev, Nicolas Perony

What is the role of social interactions in the creation of price bubbles? Answering this question requires obtaining collective behavioural traces generated by the activity of a large number of actors. Digital currencies offer a unique possibility to measure socio-economic signals from such digital traces. We focus on Bitcoin, the most popular cryptocurrency. Bitcoin has experienced periods of rapid increase in exchange rates (price) followed by sharp decline; we hypothesise that these fluctuations are largely driven by the interplay between different social phenomena. We thus quantify four socio-economic signals about Bitcoin from large datasets: price on online exchanges, volume of word-of-mouth communication in online social media, volume of information search and user base growth. By using vector autoregression, we identify two positive feedback loops that lead to price bubbles in the absence of exogenous stimuli: one driven by word of mouth, and the other by new Bitcoin adopters. We also observe that spikes in information search, presumably linked to external events, precede drastic price declines. We further identify users within the Bitcoin transaction network, and show how the user adoption and endogenous economic activity (signalled by both: the economic transactions between them, and capital accumulation) makes this system a rather different one with respect to how it was originally envisioned.

CO1049: **On the nonlinear dependency between social media and the stock market: Twitter and news**
Presenter: **Tharsis Souza**, University College London, United Kingdom
Co-authors: Tomaso Aste

The importance of news in the financial markets is now commonplace. On the other hand, the true value of social media analytics is still under debate. Traditional research in this area assumes an underlying Gaussian probability density function and a linear relationship between market sentiment and financial variables. We use a information-theoretic framework for the nonparametric estimation of the dependency between market sentiment and stock returns. We conclude that: (i) sentiment analytics have a non-linear dependency with stock returns; (ii) compared to the traditional Gaussian assumption, there is up to 10 times more information between the market sentiment and stocks returns if nonlinearity is assumed; (iii) Twitter analytics can provide more information than news under a nonparametric framework. Our results suggest that, when it comes to sentiment analysis of news and social media applied to finance, a nonlinear model is highly preferable over the common Gaussian (linear) assumption. Also, the evidence presented support the use of Twitter as a significant datasource in the context of financial markets even when compared to traditional newswires.

CO1112: **In quest of significance: Identifying types of Twitter sentiment spikes that predict events in sales**
Presenter: **Olga Kolchyna**, University College London, United Kingdom
Co-authors: Tharsis Souza, Tomaso Aste

We introduce a sentiment classification approach that combines the traditional lexicon-based approach with support vector machines algorithm. Using this model we analyse over 150 million tweets related to 75 companies from the retail sector and study the power of Twitter sentiment to predict sales events for the selected companies. To perform the analysis we developed a robust method for identifying and clustering bursts in sales and Twitter series based on their shape. The result of events clustering suggests that Twitter time series can be separated into six clusters with unique signatures. We demonstrate that some types of Twitter events have predictive power that is more significant than the predictive power of the aggregated Twitter sentiment signal. The methodology can be easily extended to the field of finance.

CO482 Room Montague FINANCIAL FORECASTING Chair: Fotis Papailias

CO1101: **Volatility discovery**
Presenter: **Cristina Scherrer**, Aarhus University and CREATES, Denmark
Co-authors: Gustavo Fruet Dias, Fotis Papailias

There is a large literature investigating how homogenous securities traded on different markets incorporate new information (price discovery analysis). We extend this concept to the stochastic volatility process and investigate how markets contribute to the efficient stochastic volatility process that is attached to the common efficient price (volatility discovery analysis). We use daily measures of realized variance as estimates of the latent market integrated variance and adopt the fractionally cointegrated vector autoregressive (FVAR) framework. We extract the common fractionally stochastic trend associated with the efficient stochastic volatility, which is common to all markets. We evaluate volatility discovery by the adjustment coefficients of the FVAR. We work with 30 of the most actively traded stocks in the US, that span from January 2007 to December 2014. We document that volatility discovery does not necessarily takes place at the same venue as price discovery. The results hint that market efficiency should be analysed by broader measures, that take into consideration the stochastic volatility.

CO1020: **Quasi-qualitative methods for assessing the expected accuracy of volatility forecasts of equity prices**
Presenter: **Nikolaos Kourogenis**, University of Piraeus - Research Center, Greece
Co-authors: Antonios Antypas, Marianna Birmpa

We provide a set of criteria which allow us to classify the equities within a market with respect to the expected accuracy of their volatility forecasts. Because these criteria depend on specific stylized facts of equity prices, and not on any backtesting, we refer to our approach as "quasi-qualitative". The advantage of this approach is that it does not require very long time series data. Moreover, the resulting classification of equities does not restrict us to the use of specific volatility forecasting models. We apply our method to construct portfolios whose equities, along with a set of standard equity selection criteria, satisfy the criterion of high expected accuracy of their volatility forecasts. We compare the performance of these portfolios with the performance of similar portfolios, which are constructed without the "accuracy criterion", over different horizons.

CO1151: **A horse for (almost) every course: Forecasting financial time series with the Theta Method**
Presenter: **Konstantinos Nikolopoulos**, Bangor University, United Kingdom
Co-authors: Dimitrios Thomakos

Building on earlier work on the properties and performance of the univariate and bivariate Theta method we: (a) derive new theoretical formulations for the application of the method on multivariate time series; (b) investigate the conditions for which the multivariate Theta method is expected to forecast better than the univariate one; (c) evaluate through simulations the univariate, bivariate and multivariate form of the method; and (d) evaluate these latter models in real macroeconomic and financial time series. The study provides sufficient empirical evidence to illustrate the suitability of the method for financial forecasting; furthermore it provides the motivation for further investigation of the Theta method in other application areas.

CO1191: **Interval-based trading strategies**
Presenter: **Dimitrios Thomakos**, University of Peloponnese, Greece
Co-authors: Fotis Papailias

We consider an alternative approach in constructing interval-based trading strategies using smoothing on financial returns and not on prices directly. We suggest different smoothers based on the properties of the returns and approximations for the corresponding variances of the smoothed series, from which we can obtain confidence intervals of various widths. Based on these intervals we then examine the theoretical and empirical properties of some easily implementable strategies. We show that the theoretical properties of these strategies relate to the distribution of the standardized returns or functions of the returns thereoff and that their variances can be less than the buy-and-hold benchmark depending on a threshold value on the distribution of the standardized returns. Our results relate to three different strands of the literature, the trading strategies that are based on some form of smoothing, the trading strategies that exploit mean reversion and the sign predictability of financial returns. Our empirical results show that all suggested strategies are meaningful investment devices as their performance has higher average returns and lower risk, across a variety of different parametrizations.

CO1061: **Improved yield curve forecasting**
Presenter: **Fotis Papailias**, Queens University Belfast, United Kingdom

An improved methodology for yield curve forecasting is introduced. The forecasts of the Nelson-Siegel time-varying parameters are estimated using a two-stage long memory forecasting algorithm. These values are then plugged into the Nelson-Siegel model producing forecast estimates for the entire yield curve. Empirical results illustrate the applicability of the method.

CO362 Room SH349 RECENT DEVELOPMENTS IN TIME VARYING MODELLING Chair: Natalia Bailey

CO1701: **Time-varying LASSO**
Presenter: **Filip Zikes**, Federal Reserve Board, United States
Co-authors: George Kapetanios

A LASSO-type estimator is introduced for large linear models with time-varying parameters. The estimator of the parameter vector at a particular point in time is obtained by applying standard LASSO to kernel weighted observations, where the kernel is centered on the time-point of interest and decays for more distant observations. The estimator is easy to implement in practice and standard algorithms developed for LASSO with fixed parameters can be readily used. We derive the theoretical properties of the estimator, allowing for deterministic or stochastic smoothly varying parameter processes. We discuss ways in which tuning parameters, such as the one underlying the LASSO penalty and the kernel bandwidth, can be data dependent. We present a Monte Carlo simulation to corroborate the theoretical results and to shed some light on the properties of the estimator in small samples and in particular on the type of situations where our time-varying estimation is preferable to full-sample LASSO. Finally, we present an application to forecasting macroeconomic and financial variables.

CO1283: **Time varying price discovery and market microstructure noise**
Presenter: **Gustavo Fruet Dias**, Aarhus University and CREATES, Denmark
Co-authors: Cristina Scherrer, Marcelo Fernandes

We construct a continuous-time market microstructure model in which assets prices co-vary in a stochastic manner. We derive the exact discrete daily price discovery measures and show explicitly how sampling frequency and continuous parameters affect them. With regard to estimation, we address the issue of estimating daily measures of price discovery either in the presence of market microstructure noise (ultra high frequency data) or on its absence (lower frequency data). When there is no market microstructure noise, we show that a kernel based OLS estimator delivers consistent estimates of the price discovery measures and it also compares favourable in finite sample to the daily VEC estimation. When prices are observed with market microstructure noise, we adopt an instrumental variable estimator together with the robust estimator of the realized variance to obtain consistent estimates of the price discover measures. We illustrate our theoretical findings by analysing price discovery for Bank of America (BAC).

CO0195: **Bayesian local likelihood method**
Presenter: **Katerina Petrova**, Queen Mary University London, United Kingdom
Co-authors: George Kapetanios, Ana Galvao, Liudas Giraitis
A Bayesian Local Likelihood (BLL) method for estimating time-varying parameter models is developed. First, we show that the standard Normal Inverse Gamma conjugate prior results carry over to our estimator, which deliver closed-form posterior distributions. Second, we present Monte Carlo evidence in order to further exploit the strengths and weaknesses of our approach and compare it to existing methods for introducing parameter time-variation. Third, we exploit the idea of informed priors varying over time and by augmenting our estimator with different information, we can assess its performance. Finally, we apply our method to a non-linear setting and we find evidence that it remains valid.

CO1082: **Dynamic neighbor effects in public debt ratios: The case of uncertainty**
Presenter: **Konstantinos Baltas**, London School of Economics and Political Science, United Kingdom
Co-authors: Athanasios Andrikopoulos
We model and test the idea that countries public debt-GDP ratios depend on neighbouring countries characteristics. More specifically, we propose that there is an important determinant of public debt ratios: the economic uncertainty of neighbouring economies. We further argue that, there is a pecking order among countries in the transmission through which countries specific characteristics affect the debt-GDP ratios of other countries.

CO1030: **Trend-cycle analysis using robust global VAR models**
Presenter: **Silvia Sze Wai Lui**, University of Groningen, Netherlands
Co-authors: James Mitchell
How structural breaks affect the trend-cycle estimates derived from Global Vector-Autoregressive (GVAR) models is considered. GVAR models provide an attractive means of decomposing a time-series into trend and cyclical components, with trends defined as long-run forecasts. In contrast to univariate Beveridge-Nelson decompositions, GVARs condition on intra- and inter-country information rather than just information contained in the history of the specific series of interest. Restrictions can be placed on the GVAR to aid structural interpretation and facilitate more economically-motivated trend-cycle decompositions. However, the recent financial crisis has exposed the potential sensitivity of both trend-cycle estimates and forecasts to structural instabilities, to breaks in trend and/or cycle stemming from parameter change and to shocks which violate Gaussianity assumptions. In an empirical application, focusing on the Euro area but within a global context, we explore this sensitivity. Both common and country-specific instabilities are considered. We suggest various ways in which the GVAR might be rendered more robust to instabilities such as models/forecast combination, and Impulse Indicator Saturation (IIS). We also look at the evidence for synchronisation of European and global business cycles business cycles and consider the degree to which the global financial crisis has affected trend growth prospects in Europe.

CO348 Room Gordon TIME-FREQUENCY ANALYSIS OF ECONOMIC AND FINANCIAL DATA Chair: Luis Aguiar-Conraria

CO0553: **Estimating the Phillips curve in the time-frequency domain**
Presenter: **Manuel Martins**, Universidade do Porto, Portugal
The U.S. Phillips curve in the time-frequency domain is estimated. We use a set of continuous wavelet tools, comprising the coherency, phase-difference and gain. In particular, using their multivariate counterparts, including a multivariate generalization of the wavelet gain, we estimate the Phillips Curve coefficients in the time-frequency domain. Our framework provides estimates of the Phillips curve coefficients allowing for changes simultanesouly along time and across frequencies. Our results provide new answers to three questions that have been prevalent in recent analyses of the Phillips Curve and are of key relevance for theoretical and policy considerations, namely: Has the Phillips Curve slope changed? Has the persistence of inflation changed? Is there a long-run Phillips tradeoff?

CO0545: **The impact of the labour share on growth in the 19th century**
Presenter: **Matthieu Charpe**, International Labour Organization, Switzerland
The impact of the labour share on growth is discussed by using historical national accounts for three countries: the United Kingdom from 1856-2010, France from 1896-2010 and the United States from 1898-2010. The added value is the use of data over a longer timespan than the usual system of national account series and the performance of single country estimations in contrast with existing panel data analysis. Another contribution is to perform a time-frequency analysis and a time-varying analysis of the relationship between (functional) income distribution and growth. We find evidence of common information between growth and income distribution at low frequency, with the labour share leading growth. We also show that the sign of the coefficient associated with the labour share is negative at high frequencies and turns positive at low frequencies. Lastly, the coefficient associated with the labour share increases over time at low frequencies.

CO0735: **Bidirectional relationship between investor sentiment and excess returns: New evidence from the wavelet perspective**
Presenter: **Martyna Marczak**, University of Hohenheim, Germany
Co-authors: Thomas Beissinger
New light is shed on the mutual relationship between investor sentiment and excess returns corresponding to the bubble component of stock prices. We propose to use the wavelet concept of the phase angle to determine the lead-lag relation between these variables. The wavelet phase angle allows for decoupling short- and long-run relations and is additionally capable of identifying time-varying comovement patterns. By applying this concept to excess returns of the monthly S&P500 index and two alternative monthly US sentiment indicators we find that in the short run (until 3 months) sentiment is leading returns whereas for periods above 3 months the opposite can be observed.

CO0738: **Evaluating exchange rate forecasts along time and frequency**
Presenter: **Petre Caraiani**, Institute for Economic Forecasting, Romania
A key puzzle in international macroeconomics is the proposition that no model can beat the random walk in predicting the exchange rate. Nevertheless, recent models are able to outperform the random walk for certain specifications and horizons. The aim is to contribute to this literature by performing an evaluation of exchange rate forecasts of reference models relative to the exchange rate in time and frequency. While the literature has usually addressed the performance of exchange rate models relative to the random walk in time only, whether this relative performance is uniform along different frequencies, or whether it is driven by certain frequencies, is studied.

CO0803: **The Q theory: New evidence from wavelet analysis**
Presenter: **Fabio Verona**, Bank of Finland, Finland
We test the Q theory of investment using wavelets analysis. The motivation is given by 1) the conjecture that both Q and cash flow may not be related with investment in the same way at all frequencies and 2) the recent progresses made in the wavelet literature that allow now to estimate the coefficients in the investment equations in the time-frequency domain. The usual findings in the empirical investment literature are that 1) Q has very little explanatory power, 2) the implied structural parameter yields implausibly slow adjustment of the stock of capital, and 3) other variables (such as cash flow) have larger effects on investment. Using wavelet tools, we find that 1) the fit of the investment equation is not always bad, and not at all frequencies, 2) the estimated elasticity of investment to Q is on average 4 times higher than in the time-domain (it can even be more than 12 times larger in some time-frequency regions), and 3) cash flow does not drive out Q, in fact Q is as important as cash flow.

EI008 Room CLO B01 SPECIAL SESSION ON STATISTICS FOR FUNCTIONAL DATA Chair: Byeong Park

EI0204: **Optimal Bayes classifiers for functional data**
Presenter: **Hans-Georg Mueller**, University of California Davis, United States
For functional data, probability density functions do not exist, due to the behavior of the small ball probabilities in function spaces. Therefore the classical construction of Bayes classifiers that uses density quotients needs to be modified when the predictors are functional data. We develop a well-defined approach that is based on density ratios of projections on a sequence of eigenfunctions that are common to the groups to be classified. It can be shown that in the large sample limit the proposed classifiers achieve perfect classification under certain conditions. The proposed functional Bayes classifiers exhibit favorable performance compared to previously studied functional classifiers in simulations and various data applications, including classification of gene expression data, spectral data or functional magnetic resonance imaging (fMRI) data for attention deficit hyperactivity disorder (ADHD) patients.

EI1383: **Nonparametric registration to low-dimensional function spaces**
Presenter: **Alois Kneip**, University of Bonn, Germany
Co-authors: Heiko Wagner
Registration aims to decompose amplitude and phase variation of samples of curves. Phase variation is captured by warping functions which monotonically transform the domains of sample curves. Resulting registered curves should then only exhibit amplitude variation. Most existing registration method rely on aligning typical shape features like peaks or valleys to be found in each sample function. It is shown that this is not necessarily an optimal strategy for subsequent statistical data exploration and inference. In this context a major goal is to identify low dimensional linear subspaces of functions that are able to provide accurate approximations of the observed functional data. We present a registration method where warping functions are defined in such a way that the resulting registered curves span a low dimensional linear function space. Problems of identifiability are discussed in detail, and connections to established registration procedures are analyzed. The method is applied to real and simulated data.

EI1750: **Regression on functional data: Methodological approach with application to near-infrared spectrometry**
Presenter: **Frederic Ferraty**, Mathematics Institute of Toulouse, France
Nowadays, observing a scalar response and a functional variable as predictor is a common situation. For instance, in our petroleum industry problem, the response is the octane number of a gasoline sample and the functional predictor is a curve representing its near-infrared spectrum (wavelengths vs absorbances). The statistician community developed numerous models for handling such datasets and we focus on four regression models: two standard ones as the functional linear model and the functional nonparametric regression, and two more recently developed: the functional projection pursuit regression and a parsimonious model involving a nonparametric variable selection method. Each of these models are implemented with two datasets containing near-infrared spectrometric curves. In a first stage, a brief comparative study of these models is carried out in order to emphasize their possible advantages and drawbacks. At a second stage, one proposes to gather most relevant informations obtained from these analyses to boost most recent regression models in terms of prediction quality and interpretability.

EI1137: **Functional data analysis for positron emission tomography**
Presenter: **John Aston**, University of Cambridge, United Kingdom
Co-authors: Ci-Ren Jiang, Jane-Ling Wang
Positron Emission Tomography (PET) is an imaging technique which can be used to investigate chemical changes in human biological processes such as cancer development or neurochemical reactions. Most dynamic PET scans are currently analyzed based on the assumption that linear first order kinetics can be used to adequately describe the system under observation. However, there has recently been strong evidence that this is not the case. In order to provide an analysis of PET data which is free from this compartmental assumption, we propose a nonparametric deconvolution and analysis model for dynamic PET data based on functional principal component analysis. This yields flexibility in the possible deconvolved functions while still performing well when a linear compartmental model setup is the true data generating mechanism. As the deconvolution needs to be performed on only a relative small number of basis functions rather than voxel by voxel in the entire 3-D volume, the methodology is both robust to typical brain imaging noise levels while also being computationally efficient. The new methodology is investigated through simulations and also applied to a neuroimaging study whose goal is the quantification of opioid receptor concentration in the brain.

EO046 Room CLO 101 RESAMPLING PROCEDURES FOR DEPENDENT DATA Chair: Andres M Alonso

EO0155: **Resampling techniques for dependent functional data**
Presenter: **Han Lin Shang**, Australian National University, Australia
The bootstrap is a useful method in functional data analysis for estimating the distribution of an estimator or test statistics by resampling data or a model estimated from data. However, the bootstrap validity heavily depends on whether the functional data are generated from an independent random variable or a time series. The work is concerned with the application of the bootstrap to functional time series. We extend some commonly used methods that have been proposed in the univariate time series literature to functional time series context, including Markovian bootstrap, nonparametric residual bootstrap, and the residual bootstrap for parametric models. Illustrated by a series of simulation studies and real-world applications, we examine the estimation accuracy of an estimator obtained from different resampling techniques using the notion of interval score. We argue that methods for implementing the bootstrap with functional time series are not as well understood as methods for independent functional random variables.

EO0502: **Model uncertainty and the forecast accuracy of arma models: A survey**
Presenter: **Joao Henrique Goncalves Mazzeu**, Universidad Carlos III de Madrid, Spain
Co-authors: Esther Ruiz, Helena Veiga
The objective is to survey the literature on the effects of model uncertainty on the forecast accuracy of linear univariate ARMA models. We consider three specific uncertainties: parameter estimation, error distribution and lag order. We also survey the procedures proposed to deal with each of these sources of uncertainty. The results are illustrated with simulated data.

EO0509: **Bootstrap VAR forecasts: The effect of model uncertainties**
Presenter: **Diego Fresoli**, Vienna University of Technology, Austria
Co-authors: Esther Ruiz
VAR models are popular to forecast macroeconomic time series. However, the model, the parameters and the error distribution are rarely known without uncertainty. Bootstrap methods are successfully designed to deal with all these sources of uncertainties. The performance of the popular forecast Bonferroni cubes constructed using the traditional Gaussian methodology and several variants of the bootstrap procedure that incorporate error distribution, parameter uncertainty, bias correction and lag order uncertainty are compared. As a result, the relative importance, for different forecast horizons, number of parameters and persistence of each of these sources of uncertainty is measured. The Monte Carlo results suggest that the best performance of bootstrap cubes are obtained when taking into account the parameter uncertainty and that bias and model uncertainty

only become important when constructing long-run forecast regions of persistent VAR models. Similar conclusions are obtained in an empirical application to a four variate system containing US monthly percent changes of crude oil prices, inflation, industrial production growth and federal funds rates.

EO0683: **Bootstrap-based bias correction for dynamic factor models**
Presenter: **Carolina Garcia-Martos**, Universidad Politecnica de Madrid, Spain
Co-authors: Andres M Alonso, Guadalupe Bastos
The aim is to consider forecasts of a multivariate time series that follows a dynamic factor model. We propose to obtain interval forecasts for the common factors as well as the original time series by using bootstrap techniques. In particular, we consider the case when the factors are dominated by highly persistent AR processes. The factors' AR coefficients are estimated using small sample bias correction techniques. A Monte Carlo study points out that bias-correcting the AR coefficients of the factors allows us to obtain better results in terms of interval coverage both for the common factors and the time series. As expected, the simulation reveals that bias-correction is more successful for smaller samples. Results are gathered assuming the AR order and number of factors known as well as unknown.

EO1076: **Bootstrap prediction intervals**
Presenter: **Soumendra Lahiri**, North Carolina State University, United States
We consider prediction intervals for a general stationary time series based on block bootstrap. Theoretical properties of the bootstrap prediction intervals are studied. Finite sample properties are illustrated through a small simulation study.

EO176 Room MAL B33 HIGH-PERFORMANCE COMPUTING FOR STATISTICS, ML AND BIG DATA Chair: Alireza Mahani

EO0158: **Yin-Yang sampling: Merging parallel MCMC output**
Presenter: **Alexandra Posekany**, WU Vienna University of Economics and Business, Austria
Co-authors: Sylvia Fruehwirth-Schnatter
Due to a multitude of data available, applied Bayesian researchers and machine learners have recently turned their interest towards approaches for splitting big data into subsets, performing inference independently in parallel and then merging these outputs. Fields like social networks' and search engines' data analysis, as well as biometrics and econometrics, have increasing need for such methods. Today's challenge is that data become too large for a single analysis due to the computational burden. Thus, researchers' attention turned towards finding reasonable approaches for combining independently obtained results, often samples from posterior distributions, and subsequently obtaining a common result which recovers the joint posterior distribution and resulting posterior estimators or decisions. We face this challenge through the yin-yang sampler, a mathematically sound approach to merging two samples from posterior distributions based on different data partitions. We apply the notion of correcting for applying the same prior for each subset instead of only once for the full data set. Sequential or treewise usage of yin-yang sampling steps recovers the full sample's posterior from subsamples' posteriors for any given number of reasonably large subsets which have to contain enough information for sound inference results. To demonstrate our approach, we provide several simulation studies, including Bayesian linear regression and a real economic data example.

EO0160: **Statistical model selection with Big Data**
Presenter: **David Hendry**, Oxford, United Kingdom
Co-authors: Jurgen Doornik
Big Data offer potential benefits for statistical modelling, but confront problems including an excess of false positives, mistaking correlations for causes, selecting by inappropriate methods and tackling vast computations. Paramount considerations when searching for a data-based relationship using Big Data include the formulation problem of embedding underlying relationships in general initial models, possibly restricting the number of variables to be selected over by non-statistical criteria; the selection problem of using good quality data on all variables, analyzed at tight significance levels by a powerful selection procedure, while retaining available theory insights; the evaluation problem of testing for relationships being well specified and invariant to shifts in explanatory variables, and the computational problem of using a viable approach to handling immense numbers of possible models. The last is especially important for the extended general-to-specific approach in Autometrics, but a feasible solution using mixing multiple block path searches while retaining theory insights is described.

EO0274: **SIMD parallel MCMC sampling with applications for big-data Bayesian analytics**
Presenter: **Alireza Mahani**, Sentrana Inc, United States
Co-authors: Mansour Sharabiani
Two opportunities for Single-Instruction Multiple-Data (SIMD) parallelization of MCMC sampling for probabilistic graphical models are presented. In exchangeable models with many observations such as Bayesian Generalized Linear Models (GLMs), child-node contributions to the conditional posterior of each node can be calculated concurrently. In undirected graphs with discrete-value nodes, concurrent sampling of conditionally-independent nodes can be transformed into a SIMD form. High-performance libraries with multi-threading and vectorization capabilities can be readily applied to such SIMD opportunities to gain decent speedup, while a series of high-level source-code and runtime modifications provide further performance boost by reducing parallelization overhead and increasing data locality for Non-Uniform Memory Access architectures. For big-data Bayesian GLM graphs, the end-result is a routine for evaluating the conditional posterior and its gradient vector that is 5 times faster than a naive implementation using (built-in) multi-threaded Intel MKL BLAS, and reaches within the striking distance of the memory-bandwidth-induced hardware limit.

EO0389: **Distributed particle filters: Alpha SMC and forest resampling**
Presenter: **Nick Whiteley**, University of Bristol, United Kingdom
Co-authors: Anthony Lee, Kari Heine
One challenge in inference for large, or otherwise complex, datasets is the design of algorithms that scale on parallel and distributed architectures. In the context of Sequential Monte Carlo (SMC) for hidden Markov models, longer data records necessitate an increased number of particles, N, in order to provide accurate estimates: when N is large one is naturally drawn to distributed implementations of the algorithm with particles simulated on different computers on a network. We will discuss alpha SMC, a generalization of Sequential Monte Carlo in which interaction between particles is modulated by a sequence of "alpha" matrices. Theoretical results then motivate the design of adaptive choices of alpha matrices that satisfy a generalized effective sample size criterion. Combined with a useful lower bound on this effective sample size that does not require the transmission of all particle weights to a central processor, this motivates forest resampling - a specific implementation of alpha SMC amenable to implementation in a distributed environment.

EO0694: **Variable selection in high-dimensional data sets using GPU**
Presenter: **Witold Rudnicki**, University of Bialystok, Poland
Co-authors: Szymon Migacz, Krzysztof Mnich, Antoni Rosciszewski, Andrzej Sulecki, Pawel Tabaszewski
The aim is to describe new algorithms for knowledge discovery that are possible due to high computational power of GPUs. We focus on algorithmic

side and provide illustrations with biomedical applications. The algorithms are general and can be applied to any problem described with millions of features. The ideas originate from problems in modern molecular biology, which generates datasets with thousands and even millions of features. Finding relevant features is crucial for building explanatory models. Their huge number combined with relatively small number of objects and weak signal hidden behind inevitable noise constitutes really hard problem. We have developed the feature selection scheme in the information system based on exhaustive search in two and more dimensions. We generate all possible pairs, triplets etc. of variables and compute information gain for decision variable due to each $n-$tuple of variables in reference to the a priori distribution. The variables of $n-$tuples that are statistically significant are deemed relevant and selected for further analysis. The GPU implementation of the algorithm is capable of analysing billions of $n-$tuples per second, therefore enabling the exhaustive search of synergistic effects in large multi-dimensional datasets.

EO098 Room MAL 421 STATISTICAL MODELLING Chair: Jochen Einbeck

EO0202: **Continuous-time multi-state survival models for cross-sectional data**
Presenter: **Ardo van den Hout**, University College London, United Kingdom
Co-authors: Luz Sanchez-Romero
A multi-state survival model describes change of status over time. The archetype example is the three-state illness-death model, where individuals are in state 1 when they are healthy, in state 2 when they are ill, and in state 3 when they are deceased. Typically, longitudinal data are available with repeated observations within individuals. When transition times between the states are interval censored, the data are called panel data. There is range of methods in the literature for fitting continuous-time multi-state survival models to panel data. With longitudinal data there is individual-specific information on the transition process during the follow-up. With cross-sectional data, this information is not available. In case the time scale of the process is age in years, cross-sectional data typically consist of a table with sample frequencies pertaining to different individuals for each age/state cross-classification. We consider the statistical modelling of a continuous-time multi-state survival process using cross-sectional data. Included topics are identifiability problems, hazard models, estimation, and goodness of fit. The modelling will be illustrated with a data analysis.

EO0428: **A goodness-of-fit test for the stratified proportional hazards model for survival data**
Presenter: **Jean-Francois Dupuy**, INSA de Rennes, France
Co-authors: Rim Ben Elouefi
Goodness-of-fit testing is addressed in the stratified proportional hazards model for survival data. Stratified proportional hazards regression generalizes usual Cox regression by allowing different groups (called strata) of the population under study to have distinct baseline hazard functions. The stratified proportional hazards model has recently been applied in various fields, such as economy, marketing, medicine and public health. Goodness-of-fit testing has thus become a crucial issue in this model. A test statistic based on within-strata cumulative sums of martingale residuals over covariates is proposed. Its asymptotic distribution is derived under the null hypothesis of model adequacy. A Monte Carlo procedure is proposed to approximate the critical value of the test. Simulation studies are conducted to examine finite-sample performance of the proposed statistic for various sample sizes, censoring fractions and numbers of strata. It is shown that the proposed test performs well under a wide range of conditions usually encountered in practice.

EO1015: **A unified multivariate survival model in presence of a cure fraction**
Presenter: **Francisco Louzada**, University of Sao Paulo, Brazil
We propose a new lifetime model for multivariate survival data with a surviving fraction. We develop this model assuming that there are m types of unobservable competing risks, where each risk is related to a time of the occurrence of an event of interest. We explore the use of Markov chain Monte Carlo methods to develop a Bayesian analysis for the proposed model. We also perform a simulation study in order to analyse the frequentist coverage probabilities of credible interval derived from posteriors. Our modelling is illustrated through a real data set on medical area.

EC0633: **The gradient test for generalised linear models with random effects**
Presenter: **Antonio Hermes Marques da Silva Junior**, Durham University, United Kingdom
Co-authors: Jochen Einbeck, Peter Craig
A common interest in regression analysis is explanatory variable selection. This problem is often addressed via hypothesis testing. In the context of generalised linear models with random effects, we discuss the problem of hypothesis testing for the regression parameters. The focus is on a relatively new test, named gradient test, which has the advantage of not relying on matrix operations, and not requiring estimation of the Fisher Information matrix. The reference distribution of this new statistic is chi-square, similarly as for the well known likelihood ratio, Wald and Rao tests. The performance of the gradient test is evaluated in terms of nominal and power levels and compared to the three other tests. We provide numerical results and discussions on the performance of the four tests under analysis.

EO0806: **A new test for number inflation or number deflation**
Presenter: **Paul Wilson**, University Of Wolverhampton, United Kingdom
Co-authors: Jochen Einbeck
We present a test of zero-modification which checks if the observed number of zeros is consistent with the hypothesized count distribution. This test has three major advantages over currently available tests, e.g. likelihood ratio, score and Wald tests: (i) it is highly intuitive and easily understood by non-statisticians; (ii) both the sample size and probability of a zero observation are incorporated into the determination of the distribution of the test statistic, this is not the case for current tests which are reliant on the asymptotic distribution of the test statistic which may be a poor approximation to the true distribution; (iii) for low values of the parameter of the count distribution it is considerably more powerful than existing tests. This test is easily extended to test for inflation or deflation of any non-negative values, and, by performing tests of inflation/deflation of the counts present in observed data relative to any given model, it is possible to assess the suitability of that model, thus the test has applications far beyond its original remit as a test of zero-inflation. Such testing may be represented diagrammatically.

EO092 Room MAL B35 SPATIAL EXTREMES AND MAX-STABLE PROCESSES Chair: Clement Dombry

EO0271: **Estimation and assessment of anisotropic Brown-Resnick space-time models**
Presenter: **Sven Buhl**, Technische Universitaet Muenchen, Germany
Co-authors: Claudia Klueppelberg
Max-stable processes can be viewed as the natural infinite-dimensional generalisation of multivariate extreme value distributions. We focus on the Brown-Resnick space-time process, a prominent max-stable model. We extend existing spatially isotropic models to more general anisotropic versions and use pairwise likelihood to estimate the model parameters. For regular grid observations we prove strong consistency and asymptotic normality of the estimators for fixed and increasing spatial domain, when the number of observations in time tends to infinity. We also present a statistical test for spatial isotropy versus anisotropy, which is based on asymptotic confidence intervals of the pairwise likelihood estimators and carried out using a subsampling procedure. We fit the spatially anisotropic Brown-Resnick model and apply the proposed test to precipitation measurements in Florida. In addition, we present some recent diagnostic tools for model assessment.

EO0278: **Asymptotic properties of the empirical spatial extremogram**
Presenter: **Yong Bum Cho**, Columbia University, United States
Co-authors: Richard Davis, Souvik Ghosh
The extremogram is a useful tool for measuring extremal dependence and checking model adequacy in a time series. We define the extremogram in the spatial domain when the data is observed on a lattice or at locations distributed as a Poisson point process in d-dimensional space. We establish a central limit theorem for the empirical spatial extremogram. We show these conditions are applicable for max-moving average processes and Brown-Resnick processes and illustrate the empirical extremogram's performance via simulation. We also demonstrate its practical use with a data set related to rainfall in a region in Florida.

EO0447: **Spatial extremes for fire risk assessment**
Presenter: **Benjamin Shaby**, Penn State University, United States
Wildfires have the potential to inflict huge losses of life, infrastructure, and habitat. High-resolution climate models can simulate weather variables that influence ignition probability and spread potential, but these models produce distributions for these variables that have unrealistic tails. We use spatial max-stable models to extract meaningful information about extreme fire weather in California from high-resolution weather model output by constructing a spatial downscaling model within the max-stable framework. This allows us to treat climate model output as covariates, rather than as future weather variables, in a spatial extreme value regression model fit to observed data.

EO1160: **Spatial modeling of drought events using max-stable processes**
Presenter: **Marco Oesting**, University of Siegen, Germany
Co-authors: Alfred Stein
Having severe environmental and socioeconomic impact, drought events belong to the most far-reaching natural disasters in Africa. We analyzed and modeled the spatio-temporal statistical behaviour of the Normalized Difference Vegetation Index (NDVI) as an indicator for drought, reflecting its effects on vegetation. The study used a data set for Rwanda obtained from multitemporal satellite remote sensor measurements during a 14 year period and divided into season-specific spatial random fields. Extreme deviations from average conditions were modeled with max-stable Brown-Resnick processes taking methodological and computational challenges into account. Those are caused by the large spatial extent and the relatively short time span covered by the data. Extensive simulations enabled us to go beyond the observations and, thus, to estimate several important characteristics of extreme drought events, such as their expected return period.

EO0890: **On functional records and champions**
Presenter: **Maximilian Zott**, University of Wuerzburg, Germany
Co-authors: Clement Dombry, Michael Falk
Records among a sequence of iid random variables $X_1, X_2, \ldots$ on the real line have been investigated extensively over the past decades. A record is defined as a random variable X_n such that $X_n > \max(X_1, \ldots, X_{n-1})$. Trying to generalize this concept to the case of random vectors, or even stochastic processes with continuous sample paths, the question arises how to define records in higher dimensions. We introduce two different concepts: A *record* is meant to be a stochastic process (or a random vector) X_n that is larger than $X_1, \ldots, X_{n-1}$ in at least one component, whereas a *champion* has to be larger than its predecessors in all components. The behavior of records and champions is investigated. In particular, the probability that a stochastic process X_n is a record (or champion) as n tends to infinity is studied, assuming that the processes are in the max-domain of attraction of a max-stable process. Furthermore, the distribution of X_n, given that X_n is a record (or champion) is derived.

EO076 Room MAL B36 MACHINE LEARNING, APPROXIMATION AND ROBUSTNESS Chair: Andreas Christmann

EO0279: **Approximation theory viewpoints of some learning algorithms**
Presenter: **Ding-Xuan Zhou**, City University of Hong Kong, China
Methods and ideas from approximation theory and function spaces play an important role in learning theory. We provide approximation theory viewpoints of some learning algorithms: randomized Kaczmarz algorithm, online learning and gradient descent algorithms, learning with functional data, and some other related methods.

EO0374: **On robustness properties of kernel based methods**
Presenter: **Andreas Christmann**, University of Bayreuth, Germany
The focus is on statistical robustness properties of kernel based methods. Computational issues for the kernel methods and their bootstrap approximations will also be discussed.

EO0606: **Learning theory for vector-valued distribution regression**
Presenter: **Zoltan Szabo**, University College London, United Kingdom
Co-authors: Bharath Sriperumbudur, Barnabas Poczos, Arthur Gretton
The focus is on the distribution regression problem (DRP): we regress from probability measures to Hilbert-space valued outputs, where the input distributions are only available through samples (this is the 'two-stage sampled' setting). Several important statistical and machine learning problems can be phrased within this framework including point estimation tasks without analytical solution (such as entropy estimation), or multi-instance learning. However, due to the two-stage sampled nature of the problem, the theoretical analysis becomes quite challenging: to the best of our knowledge the only existing method with performance guarantees to solve the DRP task requires density estimation (which often performs poorly) and the distributions to be defined on a compact Euclidean domain. We present a simple, analytically tractable alternative to solve the DRP task: we embed the distributions to a reproducing kernel Hilbert space and perform ridge regression from the embedded distributions to the outputs. We prove that this scheme is consistent under mild conditions, and construct explicit finite sample bounds on its excess risk as a function of the sample numbers and the problem difficulty. Specifically, we establish the consistency of set kernels in regression, which was a 15-year-old-open question, and also present new kernels on embedded distributions. The practical efficiency of the studied technique is illustrated in supervised entropy learning and aerosol prediction.

EC0468: **Optimal rates for regularization of statistical inverse problems**
Presenter: **Nicole Muecke**, University of Potsdam, Germany
Co-authors: Gilles Blanchard
An ill-posed inverse problem $Af = g$ with $A : \mathcal{H}_1 \longrightarrow \mathcal{H}_2$ a known bounded linear operator between Hilbert spaces is considered. The aim is to recover the element $f \in \mathcal{H}_2$ from a partial and noisy observation of the element g. In a statistical learning setting, $\mathcal{H}_2$ is a space of real-valued functions over some set $\mathcal{X}$ and observations are the noisy values $Y_i := g(X_i) + \varepsilon_i$ at iid data points $X_1, ..., X_n$, drawn according to a probability distribution ν on $\mathcal{X}$ with $\mathbb{E}[Y_i|X_i] = g(X_i)$. The ε_i are independent noise variables, satisfying $\mathbb{E}[\varepsilon_i|X_i] = 0$. An estimator of f is obtained via some spectral regularization method. In order to obtain a non-trivial rate of convergence it is assumed, that f satisfies some source condition, which quantifies the regularity of f and determines the possible rates of reconstruction. If $\mathbf{z} = (\mathbf{x}, \mathbf{y})$ are observed data and $(\lambda(n))_n$ is a sequence of regularization parameters, there is an optimal a-priori choice $\lambda_0(n)$ to get best minimax rates for the estimator $\hat{f}_{\mathbf{z}}^{\lambda_0(n)}$ in probability, for the scale of norms $||B^s(f - \hat{f}_{\mathbf{z}}^{\lambda_0(n)})||_{\mathcal{H}_1}$, $0 \leq s \leq 1/2$, where $B = S^\star S$, $S = \iota \circ A$ and $\iota : \mathcal{H}_2 \longrightarrow L^2(\mathcal{X}, \nu)$ an embedding.

EO070 Room MAL 414 MODEL ASESSMENT Chair: Maria Dolores Jimenez-Gamero

EO0348: **Minimum distance estimators for count data based on the probability generating function with applications**
Presenter: **Apostolos Batsidis**, University of Ioannina, Greece
Co-authors: Maria Dolores Jimenez-Gamero, Francisco Novoa-Munoz

Parameter estimation is a crucial aspect of statistical data analysis. Maximum likelihood (ML) estimation is the most popular method of estimating the unknown parameters of a model. Although efficient, it is well-known that ML estimates are rather sensitive to outlying observations or/and computationally difficult if the corresponding probability density function or probability mass function is complicated. For these reasons, alternative procedures have been proposed and in this frame the use of transforms such as the characteristic function, the moment and the probability generating function in point estimation has been investigated. Specifically, when dealing with count data, inferential methods based on the empirical probability generating function have been proposed. Properties of parameter estimators obtained my minimizing a distance between the empirical probability generating function and the probability generating function of a model for count data are studied. Specifically, it is proved that, under certain not restrictive conditions, the resulting estimators are consistent and, suitably normalized, asymptotically normal, even if the model is misspecified. Applications of the obtained results to the goodness-of-fit problem, model selection problem and to the problem of testing for separate families of distributions are considered.

EO0718: **Robust testing for superiority between two regression curves**
Presenter: **Juan-Carlos Pardo-Fernandez**, Universidade de Vigo, Spain
Co-authors: Graciela Boente

The focus is on the problem of testing the null hypothesis that the regression functions of two populations are equal versus one-sided alternatives under a general nonparametric homoscedastic regression model. To protect against atypical observations, the test statistic is based on the residuals obtained by using a robust estimate for the regression function under the null hypothesis. The asymptotic distribution of the test statistic is studied under the null hypothesis and under root-n contiguous alternatives. A Monte Carlo study is performed to compare the finite sample behaviour of the proposed tests with the classical one obtained using local averages. A sensitivity analysis is carried on a real data set.

EO1102: **New models for underdispersed count data based on M/M/1 queues**
Presenter: **Amanda Fernandez-Fontelo**, Autonomous university of Barcelona, Spain
Co-authors: Pedro Puig

In count data the phenomenon of underdispersion is less common than overdispersion. Accordingly, the researchers have been developed a wide range of statistical models which are able to deal with overdispersion, but few one are able to treat with underdispersion. One of these few models is the COMPoisson distribution which is based on the theory of M/M/1 queues. It was introduced assuming that the service rate could depend on the length of the queue following a power function $\mu_n = \mu n^c$, in which c is a constant indicating the degree to which the service rate is affected by the system state, and describing the (over) underdispersion of the data. Similarly, we propose other models for the service rate, like an exponential function $\mu_n = \mu e^{\beta n}$, obtaining in this case an equilibrium distribution expressed as $p_n = \frac{\rho^n}{e^{\beta n(\frac{n+1}{2})}} p_0$ in which the parameter β is also able to model the underdispersion of the data. We also propose a graphical procedure to distinguish in which cases are better our distribution than the COMPoisson. This methodology is applied to several examples, one of them is the number of chromosomal aberrations in irradiated cells, useful in Dosimetry.

EO1086: **On how to distinguish between tail distribution with right-truncated or with exponential decay**
Presenter: **Isabel Serra**, Universitat Autonoma de Barcelona, Spain
Co-authors: Paul Rochet

In extreme value theory, the performances of most estimators of the tail index highly rely on the family of the limit distribution (e.g. Weibull, Gumbel or Frechet). We consider the test on separate families of hypothesis for exponential versus uniform distributions that can be used to determine the family of the limit distribution. We derive the most powerful test for such hypothesis under mild restrictions.

EO1062: **On the estimation of the characteristic function in finite populations with applications**
Presenter: **Maria Dolores Jimenez-Gamero**, University of Sevilla, Spain
Co-authors: Juan Luis Moreno-Rebollo, Jose Antonio Mayor-Gallego

The estimation of the characteristic function of a finite population is studied. Specifically, the weak convergence of the finite population empirical characteristic process is studied. Under suitable assumptions, it has the same limit as the empirical characteristic process for independent, identically distributed data from a random variable, up to a multiplicative constant depending on the sampling design. Applications of the obtained results for the two-sample problem, testing for symmetry and testing for independence are given.

EO234 Room MAL B34 R PACKAGES FOR MULTIVARIATE OUTLIER DETECTION Chair: Anne Ruiz-Gazen

EO0349: **The restlos package: Robust estimation of location and scatter**
Presenter: **Steffen Liebscher**, Martin-Luther-University Halle-Wittenberg, Germany
Co-authors: Thomas Kirschstein

The restlos package is an R package available on CRAN which implements several methods for robust estimation of multivariate location (mean and mode) and scatter. Inherently, all methods can also be used for outlier detection. The methods are based on self-organizing maps (flood algorithm), minimum spanning trees (pMST), Delaunay triangulations (RDELA), and minimum volume convex sets (MVCH). A brief overview is presented on how to efficiently use the package. A small example is provided for each method in order to demonstrate how the corresponding package functions should be used and how the resulting numerical and graphical outputs can be interpreted. Current limitations of the package and possible directions for future improvements are outlined as well.

EO0460: **Using projection pursuit for outlier detection in R: The package REPPlab**
Presenter: **Klaus Nordhausen**, University of Turku, Finland
Co-authors: Daniel Fischer, Alain Berro, Anne Ruiz-Gazen

The R-package REPPlab is designed to explore multivariate data sets using one-dimensional unsupervised projection pursuit. The implemented indices are either adapted to cluster or to outlier detection. The strategy of exploration in REPPlab is divided into two steps. In a first step the software calculates many potentially interesting projections using several starting values for a given projection index. The second step is dedicated to the exploration of the obtained projections. The indices and tools useful for outlier detection will be demonstrated using simulated and real examples.

EO0646: **Intepretation of outliers in compositional data using the R-package mvoutlier**
Presenter: **Karel Hron**, Palacky University, Czech Republic
Co-authors: Peter Filzmoser
Multivariate outliers are often the most interesting data points because they show atypical phenomena. Several classical and robust methods have been proposed for their identification also in the context of compositional data - multivariate observations carrying relative information (typically proportions, percentages, mg/kg, etc.). The proposed tools help to better understand the reason, why these samples are defined as atypical by representing them in maps, in a compositional biplot, in univariate scatter plots, and in parallel coordinate plots. In all plots, the same special colors and symbols can be selected, referring to the relative position of the outliers in the multivariate data cloud and thus supporting an interpretation of these observations. Since compositional data are considered, a relation to the single compositional parts can be established only by special isometric logratio coordinates. The developed tools are freely available in the R package mvoutlier. The function mvoutlier.CoDa() computes the multivariate outliers, and it prepares the information for the symbols and colors. The resulting object can then be used for the plot function. All arguments are consistent for the presentations, which makes it possible to see the same symbol and color choices in different views, revealing the structure of the multivariate outliers.

EO0766: **Comparison of statistical methods for multivariate outlier detection**
Presenter: **Aurore Archimbaud**, Toulouse School of Economics, France
Co-authors: Anne Ruiz-Gazen, Klaus Nordhausen
Detection of multivariate outliers is a relevant topic in many fields such as fraud detection or manufacturing defects detection. Several non-supervised multivariate methods exist and some are based on robust and non-robust covariance matrices estimators such as the Mahalanobis distance (MD) and its robust version (RD), the robust Principal Component Analysis (PCA) with its diagnostic plot and the Invariant Coordinate Selection (ICS). The objective is to compare these different methods. Note that all these methods lead to one or several scores associated to each observation and high scores are associated with potential outliers. For robust PCA and ICS, some components are selected and outliers are identified by using some test procedure. This last step is not trivial: relevant cut-offs have to be determined and the simultaneity of tests has to be taken into account. The comparison is performed on simulated data sets with mixtures of Gaussian distributions in the context of a small proportion of outliers and when the number of observations is at least five times the number of variables. The Minimum Covariant Determinant (MCD) estimator is considered. The implementation is based on functions from the R packages: robustbase, rrcov, ICS and CerioliOutlierDetection.

EO1033: **Practical outlier detection in business surveys using robust estimation**
Presenter: **Valentin Todorov**, UNIDO, Austria
It is common that data collected through business surveys may contain outliers and be incomplete (missing values and non-response), usually skewed, and semi-continuous distributions occur often. The multivariate aspect of the data collected in business surveys makes the task of outlier identification particularly challenging. The outliers can be completely hidden in one or two dimensional views of the data. Not many studies of outlier detection methods for business survey data based on robust methods are found in the literature, and the two studies that we can mention present two R packages (rrcovNA and modi) implementing several reliable methods for outlier detection. New methods, added to the R package rrcovNA will be described and their performance will be investigated using the real data set included in the modi package as well as a data set based on a survey carried by UNIDO.

EO150 Room MAL B29 CONVEX OPTIMIZATION IN STATISTICS Chair: Keith Knight

EO0395: **Convex optimization for models of unobserved heterogeneity**
Presenter: **Roger Koenker**, University of Illinois, United States
Recent developments in convex optimization have greatly expanded the potential scope of semiparametric methods in statistics. Some new applications to quantile regression methods will be described focusing on first order gradient descent methods for large problems, as well as Kiefer Wolfowitz non-parametric maximum likelihood estimator for general mixture models.

EO1134: **Early false discoveries along Lasso path**
Presenter: **Malgorzata Bogdan**, Wroclaw University, Poland
Co-authors: Weijie Su, Emmanuel Candes
It is widely believed that as long as the true signals (nonzero regression coefficients) are sufficiently strong, Lasso shall first select true signals along its path. We will show that this common belief holds only if the number of true regressors is sufficiently small. In other case false discoveries are bound to occur relatively early on the LASSO path. We will thoroughly discuss this phenomenon using the classical regression model with the Gaussian design matrix, where we provide a tight lower bound for the false discovery proportion to be obtained if one wants to discover a certain proportion of true signals. We also provide an upper bound on the waiting time for the first false discovery.

EO1060: **Accelerated proximal point methods for solving penalized regression problems**
Presenter: **Sangkyun Lee**, TU Dortmund University, Germany
Efficient optimization methods to obtain solutions of penalized regression problems, especially in high dimensions, have been studied quite extensively in recent years, with their successful applications in machine learning, image processing, compressed sensing, and bioinformatics, just to name a few. Amongst them, proximal point methods and their accelerated variants have been quite competitive in many cases. These algorithms make use of special structures of problems, e.g. smoothness and separability, endowed by the choices of loss functions and regularizers. We will discuss two types of first-order proximal point algorithms, namely accelerated proximal gradient descent and accelerated hybrid proximal extra gradient methods, focusing on the latter, in the context of lasso and Dantzig selector.

EO0799: **A simple selection of threshold for estimators defined as a solution to convex optimization**
Presenter: **Sylvain Sardy**, University of Geneva, Switzerland
Many thresholding estimators are defined as solution to a convex optimization problem, e.g., lasso. We present two such estimators: GLM-lasso and a new block total variation estimator, both motivated by telescope image processing in Cosmology. We discuss how one can solve the corresponding convex optimization, and importantly, how to select the threshold parameter(s) in a simple and computer efficient way.

EO1084: **Minimum integrated L_p norm estimation**
Presenter: **Keith Knight**, University of Toronto, Canada
For the general "signal plus noise" model $Y_i = \theta_i + \varepsilon_i$ $(i = 1, \cdots, n)$, we consider estimators of $\{\theta_i\}$ minimizing $\int_2^\infty \ell_p(Y_1 - \theta_1, \cdots, Y_n - \theta_n)\, \omega(dp)$ subject to some constraints on $\{\theta_i\}$, where ω is some measure on $[2, \infty)$ and $\ell_p(u_1, \cdots, u_n) = \left\{ n^{-1} \sum_{i=1}^n |u_i|^p \right\}^{2/p}$. For light-tailed noise, this type of estimation can be very useful relative to least squares. At first glance, minimization of the objective function seems quite difficult; however, we will show that it can be minimized using a simple iteratively reweighted least squares algorithm.

EO210 Room MAL B20 MODELING AND STATISTICAL INFERENCE OF RECURRENT EVENTS Chair: Maria del Carmen Pardo

EO0415: **Tests for monotonic and nonmonotonic trend in time censored recurrent event data**
Presenter: **Jan Terje Kvaloy**, University of Stavanger, Norway
Co-authors: Bo Henry Lindqvist
In recurrent event data it is often of interest to detect possible systematic changes in the pattern of events. An example is a repairable system for which it is important to detect changes in the pattern of failures. Another example is to test for changes in the return of a recurrent disease. We say that there is a trend in the pattern of events if the inter-event times tend to alter in some systematic way, which means that the inter-event times are not identically distributed. By using statistical trend tests it is possible to decide whether such an alteration is statistically significant or not. In general a trend in the pattern of events can be either monotonic or nonmonotonic and it is thus useful to have tests with power against both monontonic and nonmonotonic trend. We present a class of statistical tests for trend in the event times in time censored recurrent event data based on the general null hypothesis of a renewal process. This class does in particular include a test which is attractive for general use by having good power properties against both monotonic and nonmonotonic trends.

EO0559: **A large class of models for recurrent events for which profile likelihood inference fails**
Presenter: **Eric Beutner**, Maastricht University, Netherlands
Co-authors: Laurent Bordes, Laurent Doyen
Recurrent event data arise from the study of processes that generate events repeatedly over time. Such processes occur in many settings such as biomedicine, clinical trials and engineering to mention a few. We consider a semi-parametric model for recurrent events that consists of an unknown hazard rate function, the infinite-dimensional parameter of the model, and a parametrically specified effective age function. We will present a condition on the family of effective age functions under which the profile likelihood function evaluated at the parameter vector e, say, exceeds the profile likelihood function evaluated at the parameter vector e', say, with probability p. From this we derive a condition under which profile likelihood inference for the finite-dimensional parameter of the model leads to inconsistent estimates. Examples will be presented. In particular, we provide an example where the profile likelihood function is monotone with probability one regardless of the true data generating process. Furthermore, we present statistical models containing the renewal process and the nonhomogeneous Poisson process as special cases for which it turn out that the profile likelihood function at the parameter corresponding to the nonhomogeneous Poisson process is never less than at the parameter corresponding to the renewal process regardless of whether the data come from a renewal process or an nonhomogeneous Poisson process.

EO0679: **On the use of Fleming and Harrington's test to detect late effects in clinical trials**
Presenter: **Nicolas Savy**, Toulouse Institute of Mathematics, France
Co-authors: Valerie Gares, Jean-Francois Dupuy, Sandrine Andrieu
The question of detection of late effects in the setting of clinical trials has been previously investigated. The most natural test for detecting this kind of effects depends on a parameter that, in the context of clinical trials, must be chosen a priori. We examine the reasons why this test is adapted to the detection of late effects by studying its optimality in terms of Pitman Asymptotic Relative Efficiency. We give an explicit form of the function describing alternatives for which the test is optimal. Moreover, we observe by means of a simulation study that this test is not very sensitive to the value of the parameter, which is very reassuring for its use in clinical trials. The main conclusion is that the logrank test is optimal to test the proportional risks assumption, while Fleming Harrington test is optimal to test the late effect assumption. However, it is not always easy to choose between these alternatives. A test, based on a so-called expert a priori, has been introduced to override this difficulty.

EO0751: **Estimation for counting processes with high-dimensional covariates**
Presenter: **Sarah Sarah Lemler**, Ecole CentraleSupelec, France
Recurrent event data arise in such fields as medicine, insurance, economics, and reliability. Such events include, for example, relapse from a disease in biomedical research, monetization in marketing or blogging in social network study. In this context, proportional hazards models have been largely studied in the literature to model the rate functions of recurrent event data, that represent the instantaneous probability of experiencing a recurrent event at a given time. We propose a novel kernel estimator of the baseline function in a general high-dimensional Cox model, for which we derive non-asymptotic rates of convergence. To construct our estimator, we first estimate the regression parameter in the Cox model via a Lasso procedure. We then plug this estimator into the classical kernel estimator of the baseline function, obtained by smoothing the so-called Breslow estimator of the cumulative baseline function. We propose and study an adaptive procedure for selecting the bandwidth. We state non-asymptotic oracle inequalities for the final estimator, which reveal the reduction of the rates of convergence when the dimension of the covariates grows. Lastly, we conduct a study to measure the practical performances of the resulting estimator on simulated data and we apply the implemented procedure to a real dataset.

EO1231: **Joint scale-change models for recurrent events and failure time**
Presenter: **Chiung-Yu Huang**, Johns Hopkins University, United States
Co-authors: Gongjun Xu, Sy Han Chiou, Mei-Cheng Wang, Jun Yan
Recurrent event data arise frequently in various fields such as biomedical sciences, public health, engineering, and social sciences. In many instances, the observation of the recurrent event process can be stopped by the occurrence of a correlated failure event, and thus violates the independent censoring assumption required by most conventional statistical methods. We propose a joint scale-change model for the recurrent event process and the failure time that allows the censoring time to be informative about the recurrent event process. In particular, a shared frailty variable is used to model the association between the two types of outcomes. In contrast to the popular Cox-type joint modeling approaches, the regression parameters in the proposed joint scale-change model have marginal interpretations. Moreover, the proposed approach is robust in the sense that no parametric assumption is imposed on the distribution of the unobserved frailty and that we do not need the strong Poisson-type assumption for the recurrent event process. We establish consistency and asymptotic normality of the proposed semiparametric estimators under suitable regularity conditions. To estimate the corresponding variances of the estimators, we develop a computationally efficient resampling-based procedure. Simulation studies and an analysis of hospitalization data from the Danish Psychiatric Central Register illustrate the performance of the proposed method.

EO116 Room CLO 203 ASYMPTOTIC PROPERTIES IN NONPARAMETRIC PROBLEMS Chair: Sophie Dabo

EO0437: **Strong approximation of multidimensional P-P plots processes by Gaussian processes with statistical applications**
Presenter: **Salim Bouzebda**, Universite de Technologie de Compiegne, France
The aim mainly concerns with the strong approximation of P-P plot processes in R^d by sequences of Gaussian processes. In order to evaluate the limiting laws, a general notion of bootstrapped multidimensional P-P plots processes, constructed by exchangeably weighting sample, is presented. Finally, we discuss some applications.

EO0540: **Concentration rates in semi-Markov models**
Presenter: **Ghislaine Gayraud**, University of Technology of Compiegne, France
Co-authors: Nikolaos Limnios, Irene Votsi
The asymptotic behavior of the posterior distribution in semi-Markov models is investigated. Taking into consideration the advantages of the semi-Markov process versus Markov processes, we extend previous results to the later processes. Since one of ingredients for our purposes is the construction of relevant test procedures, we construct suitable statistical tests between Hellinger balls based on observations from a semi-Markov process.

EO0725: **Semiparametric estimation with spatially correlated recurrent events**
Presenter: **Akim Adekpedjou**, Missouri University of Science and Technology, United States
Co-authors: Sophie Dabo
Consider n geographical regions that are monitored for the occurrence of a recurrent event. Further assume that various environmental factors, common to the regions trigger the recurrence of the event of interest leading to spatially correlated recurrent event data. We first propose a model for the spatial correlation structure. Techniques for estimating theregression coefficients in a Cox-type model as well as the parameters of the correlation structure are presented. The estimators obtained have population interpretation and help identify risk factors that contribute to recurrence. The estimation procedures are facilitated by transforming the original gap-times yielding a multivariate Gaussian random field where the marginal Cox-type models for the original gap-times is preserved. Asymptotic properties are discussed. A simulation study ascertaining the theoretical results and an illustration with a real spatially correlated recurrent event data are presented.

EO0715: **Asymptotic properties of quasi-empirical likelihood with functional constraints**
Presenter: **Emmanuelle Gautherat**, CREST and Reims University, France
The purpose is to review some recent extensions of the so-called generalized empirical likelihood method, when the Kullback distance is replaced by some general convex divergence. We propose to use, instead of empirical likelihood, some regularized form or quasi-empirical likelihood method, corresponding to a convex combination of Kullback and chi-2 discrepancies. We show that for some adequate choice of the weight in this combination, the corresponding quasi-empirical likelihood is Bartlett-correctable. We also establish some bounds for the confidence regions obtained by using this method. We show that this kind of results may be extended to process valued infinite dimensional parameters.

EO1170: **Adaptive variable selection in nonparametric sparse models**
Presenter: **Cristina Butucea**, University Paris-Est Marne, France
Co-authors: Natalia A Stepanova
We consider the problem of almost full recovery of the s-sparse additive components of an unknown multivariate signal f observed in a d-dimensional Gaussian white noise model. We assume that f belongs to a class of smooth functions F. For two different choices of F, we establish conditions under which almost full variable selection is possible, and provide a procedure that gives almost full variable selection. The procedure does the best (in the asymptotically minimax sense) in selecting most non-zero components of f. Moreover, it is adaptive in the parameter s.

EO144 Room MAL G15 SPATIAL AND SPATIO-TEMPORAL PROCESSES AND THEIR APPLICATIONS Chair: Jooyoung Jeon

EO0512: **A constructive spatio-temporal approach to modeling spatial covariance**
Presenter: **Ephraim Hanks**, Penn State, United States
An approach for modeling areal spatial covariance is presented by considering the stationary distribution of a spatio-temporal Markov random walk. This stationary distribution corresponds to an intrinsic simultaneous autoregressive (SAR) model for spatial correlation, and provides a principled approach to specifying areal spatial models when a spatio-temporal generating process can be assumed. The approach is applied to a study of spatial genetic variation of trout in a stream network in Connecticut, USA, and a study of crime rates in neighborhoods of Columbus, OH, USA.

EO0660: **Quadratic estimation of random field non-stationarity**
Presenter: **Ethan Anderes**, University of California at Davis, United States
More than a decade ago two physicists, Wayne Hu and Takemi Okamoto, invented a new estimator for measuring the dark matter distortion imprinted on the our observations of the cosmic microwave background (which is a relic signal of the big bang). Their estimator, called the quadratic estimator, quickly became the state-of-the-art tool for the detecting, measuring and mapping dark matter. From a spatial statistics perspective this estimator has some remarkable properties. We present an analysis of the quadratic estimator in the context of both cosmology and in a more generalized setting of estimating random field non-stationarity. We also present some extensions to the quadratic estimator from both the frequentest and Bayesian perspective.

EO1198: **Cauchy cross-covariance functions for multivariate random fields**
Presenter: **Tatiyana Apanasovich**, George Washington University, United States
The class which is refereed to as the Cauchy family allows for the simultaneous modeling of the long memory dependence and correlation at short and intermediate lags. We introduce a valid parametric family of cross-covariance functions for multivariate spatial random fields where each component has a covariance function from a Cauchy family. We present the conditions on the parameter space that result in valid models with varying degrees of complexity. We discuss practical implementations, including reparameterizations to reflect the conditions on the parameter space. We show results of various Monte Carlo simulation experiments to explore the performances of our approach in terms of estimation and cokriging. The application of the proposed multivariate Cauchy model is illustrated on a dataset from the field of Satellite Oceanography.

EO0998: **Semiparametric modeling of intraday temperature data**
Presenter: **Iryna Okhrin**, Dresden University for Technology, Germany
Co-authors: Yarema Okhrin, Wolfgang Schmid
The modeling of weather data differs from the modeling of other type of data like financial, biological, medical data. Temperature have property of daily and annual seasonality. Parametric and semiparametric procedures are derived for modeling this type of data. The parametric approach is based on the truncated Fourier series. The extraction of both daily and annual seasonal factors results stationary time series. This approach provides a pretty but not perfect forecast for temperature at any time of a year and of a day. The curve of temperature behaves similarly for all days. This property motivates the Shape Invariant Modeling. The parameters of the model, which are estimated for each day separately, are by theirself a multivariate time series. This procedure is quite flexible and allows a forecasting. The application of Dynamic Semiparametric Factor Models provides good results, too. This method is parametric in time and nonparametric in space and can be adapted to our type of data. The application of these three approaches to real data shows an advantage of semiparametric modeling.

EO452 Room MAL B30 COMPUTATIONAL APPROACHES IN FINANCIAL ECONOMICS Chair: Marcel Gorenflo

EO0576: **Portfolio diversification in financial banking networks**
Presenter: **Philipp Johannes Kremer**, EBS Universitaet fuer Wirtschaft und Recht, Germany
Co-authors: Sandra Paterlini, Marcel Gorenflo, Ben Craig
The events following the financial crisis of 2007-09, showed that single monetary financial institutions are often too exposed to some institutions, while lacking an adequate level of diversification. Using data on portfolio holdings at micro-level, we evaluate the risk and diversification profile of each monetary financial institution in the German banking network. We consider common macroeconomic diversification measures - i.e. Herfindahl Index and Gini Coefficient - and complement the empirical analysis with network measures derived from graph theory. Moreover, we build a diversification index to provide further insight on the effect of crises on financial portfolio holdings. By considering the period from 2007-2014, we are able to evaluate their evolution over time, to develop tailored policy recommendations and to derive early warning indicators that should help to better deal with systemic shocks and crises.

EO1006: **Network structure and systemic risk analysis in Eurozone**
Presenter: **Wenwei Li**, EBS Universitaet fuer Wirtschaft und Recht, Germany
Co-authors: Ulrich Hommel, Sandra Paterlini
Recent financial crises highlight the importance of the interconnectedness in an economic or financial system as the source of systemic risk and channel for shock transmission. We aim to analyze the network structure and systemic risk of Eurozone corporate sector. In particular, we apply Graphical Lasso approach to estimate the sparse dependence structure in the Eurozone corporate sector. Through the analysis of the network topological structure and statistical metrics, we investigate the network configuration and its evolution in time. The empirical results provide meaningful implications for both regulators and practitioners.

EO0472: **Adaptive regression selection and estimation under sparse ℓ_q-constraints**
Presenter: **Sandra Paterlini**, European Business School Germany, Germany
Considering fixed design linear regression settings, the aim is to construct an estimator that target the best performance among all the linear combinations of the predictors under a sparse ℓ_q-pseudonorm ($0 \leq q \leq 1$) constraint on the linear coefficients. Our strategy is to choose a model using the model selection criterion ABC' for the case of unknown variance, and the resulting least squares estimators are used for the beta coefficients. Although our estimator does not directly consider the ℓ_q-constraint, we show that it automatically adapts to the sparsity of the regression function in terms of ℓ_q-representation and achieve the minimax rate. By relying on heuristics, we introduce a new algorithm for optimal ABC' model selection and estimation. Comparisons on simulated data with state-of- art methods, including Lasso and non-convex penalties, show remarkable properties in terms of model selection, estimation error and mean squared error. Finally, by applying it to financial data, we point out its practical use in developing effective investment strategies.

EO0581: **Risk diversification in portfolio selection**
Presenter: **Margherita Giuzio**, EBS Universitaet fur Wirtschaft und Recht, Germany
Co-authors: Sandra Paterlini
The aim is to investigate the trade-off between diversification, sparsity and risk minimization in portfolio selection. Our target is to find a sparse vector of asset weights able to optimize the out-of-sample level of risk while diversifying the investments. Considering the mean-variance framework, we introduce non-convex diversification penalties in the optimization problem in order to control for sparsity and risk diversification simultaneously and derive the relationships between the proposed penalized approached and the minimum variance and risk parity portfolios. Finally, we show that state-of-art optimization methods for non-convex penalties identify suboptimal portfolios when compared to heuristics.

EO1470: **Low-latency liquidity inefficiency strategies**
Presenter: **Christian Oesch**, University of Basel, Switzerland
Co-authors: Dietmar Maringer
The vast amount of high-frequency data heralds the use of new methods in financial data analysis and quantitative trading. We deliver a proof-of-concept for a high-frequency based trading system based on an evolutionary computation method. Motivated by a theoretical liquidity asymmetry theorem from the market microstructure literature, grammatical evolution (GE) is used to exploit volume inefficiencies at the bid-ask spread. By using NASDAQ Historical TotalView-ITCH level two limit order book data execution volumes can be tracked which allows for testing of the strategies with minimal assumptions. The system evolves profitable and robust strategies with high returns and low risk.

EO266 Room MAL 415 STATISTICAL METHODS FOR ACTUARIAL SCIENCES AND FINANCE Chair: Tim Verdonck

EO0662: **Extreme value theory for (re-)insurance applications: Truncation, interval censoring and splicing**
Presenter: **Tom Reynkens**, KU Leuven, Belgium
Co-authors: Jan Beirlant, Katrien Antonio, Roel Verbelen
Extreme Value Theory has many useful applications in insurance, especially in reinsurance where only claims exceeding a certain loss level are considered. Due to the nature of insurance claims, specific problems such as truncation and interval censoring arise. Analysis of a real life insurance portfolio shows that it can happen that the maximal claim loss is not infinite but is bounded from above, i.e. right truncation arises. For some types of policies it can take a long time before claims are completely developed. Interval censoring then arises since the amount paid at the final year where information is available is an underestimation of the true total amount that will be paid, whereas an upper bound is given by incurred values provided by experts. In risk analysis, one needs a fit for the body and the tail of the distribution. Splicing of mixed Erlang distributions for the body with Pareto distributions for the tail, possibly adapted for censoring and truncation, is proposed and compared with earlier modelling approaches. We illustrate the flexibility of this approach using practical examples.

EO1138: **Some new insights into large commercial risks**
Presenter: **Davide Benedetti**, Imperial College Business School, United Kingdom
Co-authors: Enrico Biffis
Large commercial risks, such as commercial property and liability, present modelling challenges due to the paucity of data available for model estimation/validation, and the complex relation between hazard events and realized losses. There is therefore a tendency for insurance practitioners to apply a considerable degree of judgement in pricing, reserving, and capital modelling. Some new evidence on large commercial risks will be presented based on unique datasets on large commercial risks based on contributions from two leading Lloyd's of London syndicates, and a global reinsurer. The datasets contain granular information on more than 3500 losses and exposures, which are used to shed some light on the risk profile of medium to high layers of exposure as a function of different rating factor configurations, geographic region and period. In particular, differences between tail risk profiles in the North American, European, and Asia-Pacific region are explored, and structural breaks around major events are tested. Small sample issues and tail risk heterogeneity are addressed with weighted Hill estimators, tail index regressions, and Generalized Pareto

Distribution with covariates. Finally, it is carried out a benchmarking exercise in which it will be quantified the risk premiums embedded in market rates as proxied by suitably constructed indices.

EO0908: **The minimum shrinkage covariance determinant estimator**
Presenter: **Yukai Yang**, Uppsala University, Sweden
Co-authors: Kris Boudt, Steven Vanduffel, Tim Verdonck
The Minimum Covariance Determinant (MCD) estimator is often used for the high breakdown point calculation of robust Mahalanobis distances requiring the covariance matrix to be invertible, which is problematic in high dimensions. We propose the Minimum Shrinkage Covariance Determinant (MSCD) estimator, which preserves the high explosion breakdown point properties of the MCD estimator, and has the additional advantages that it is well-conditioned and has a higher implosion breakdown point. A simulation study confirms the good properties of the estimator. Finally, we illustrate the advantages of the MSCD estimator for outlier detection in high dimensions and for portfolio selection on a large investment universe.

EO0821: **Copula-based joint distributions of extreme bank losses: Single country versus European Union**
Presenter: **Simone Russo**, Joint Research Centre, Italy
Co-authors: Andrea Pagano, Jessica Cariboni
We deal with the problem of modelling extreme bank losses joint distributions with respect to an individual country versus those coming from the whole European Union (EU) banking sector. Losses are simulated via a micro simulation model (SYMBOL-Systemic Model of Bank Originated Losses), and the joint distributions, single-country versus EU, are obtained by fitting data with Archimedean Copulas. For a selection of EU countries, we estimate the Conditional Probability Density function (CPDF) of losses associated to an extreme event experienced at EU level, and viceversa. Results show that the meaning of extreme event differs country by country. In particular, we found that in one specific case individual country conditional default risk exceeds the one of EU by an order of magnitude, whereas all the other countries show a comparable conditional default risk with respect to the one simulated for the entire EU banking sector.

EO1043: **Robust bootstrap techniques for claims reserving**
Presenter: **Pieter Segaert**, KU Leuven, Belgium
Co-authors: Tim Verdonck, Stefan Van Aelst, Kris Peremans
Non-life insurers are often faced with the challenge of estimating the future reserves necessary to handle claims that are not fully settled. A well-known and popular technique to obtain reserve estimates for these outstanding claims is the (deterministic) chain-ladder method. Exactly the same estimates can also be obtained by applying special cases of generalized linear models. To obtain a prediction interval for the outstanding claims, we need an estimate for the variance of the resulting reserve estimate. Computing analytic expressions for the variance is notoriously difficult and bootstrap methodology offers a simple and popular alternative. However, classical procedures are very sensitive to the possible presence of outliers. These atypical observations, deviating from the pattern of the majority of the data, may both inflate or deflate classical future claim estimates. Even when paired with robust estimators for generalized linear models, classical bootstrap estimators may break down. We present and discuss some popular and state of the art robust bootstrap procedures in the claims reserving framework. Their performance will be investigated in a simulation study and on real data.

EO302 Room CLO 204 DESIGNS FOR NONLINEAR MIXED EFFECT MODELS Chair: France Mentre

EO0708: **Bayesian Fisher information matrix for predicting estimation error and shrinkage of individual parameters**
Presenter: **Thu Thuy Nguyen**, INSERM and University Paris Diderot, France
Co-authors: Thi Huyen Tram Nguyen, France Mentre
In nonlinear mixed-effects models (NLMEM), individual parameters can be estimated by maximum a posteriori (MAP). We aimed to 1) propose an approach based on Bayesian Fisher information matrix (BFIM) to predict standard errors (SE) and shrinkage for MAP estimation; 2) take into account possible data below the quantification limit (BQL) in BFIM; 3) illustrate these developments with a pharmacokinetic/viral kinetic (PK/VK) model. First, BFIM, expressed as the sum of the expectation of individual Fisher matrix (IFIM) and prior information, was approximated by first-order linearisation and implemented in PFIM4.0 (www.pfim.biostat.fr). Next, in order to take into account contribution of BQL data, the likelihood was decomposed as the product of the usual likelihood for observations and the probability for data to be BQL. A PK/VK simulation study showed good agreement between BFIM-predictions and simulation results, either in presence or absence of BQL data. Lastly, several scenarios were evaluated to study the influence of designs and variability levels on SE and/or shrinkage predicted by BFIM versus IFIM, showing advantage of NLMEM/MAP versus individual regression. In conclusion, BFIM, available in PFIM, enables efficient evaluation/optimisation of Bayesian designs. The proposed method for handling BQL data correctly predicted SE and shrinkage in presence of these data.

EO0955: **Optimal design calculations for nonlinear mixed effect models with long computation times**
Presenter: **Andrew Hooker**, Uppsala University, Sweden
One key property of optimal design calculations are that they should be fast enough so that a vast number of designs in a design space can be considered within the time that one clinical trial simulation of one design scenario can be considered. This is still the case with nonlinear mixed effect (NLME) models, however, in many situations the evaluation of one design scenario can still be time intensive, often on the order of minutes or even hours. Thus, it is essential that optimization of these designs evaluate the objective function as few times as possible, which often excludes gradient based methods of optimization. We discuss methods for speeding up and stabilizing the computation of the Fisher information matrix for NLME models and investigate and compare methods for optimization. We do this with two examples of designs that are slow to compute, (1) a design optimization for a drug treatment with a biological compound that displays target mediated drug disposition and (2) the computation of robust optimal designs by introducing priors about potential model parameter values.

EO0936: **Computation of the Fisher information matrix for discrete nonlinear mixed effect models**
Presenter: **Sebastian Ueckert**, INSERM and Universite Paris Diderot, France
Co-authors: France Mentre
Despite an increasing use of optimal design methodology for non-linear mixed effect models (NLMEMs) during the clinical drug development process, examples involving discrete data NLMEMs remain scarce. One reason is the limitation of existing approaches to calculate the Fisher information matrix (FIM) which are either model dependent and based on linearization or computationally very expensive. The main challenges in the computation of the FIM for discrete NLMEMs evolve around the calculation of two integrals. First, the integral required to calculate the expectation over the data, and second, the integral of the likelihood over the distribution of the random effects. Monte Carlo (MC), Latin-Hypercube (LH) and Quasi-Random (QR) sampling for the calculation of the first as well as adaptive Gaussian quadrature (AGQ) and QR sampling for the calculation of the second integral are proposed. The resulting methods are compared for a number of discrete data models and evaluated in the context of model based adaptive optimal design.

EO1274: **Supersaturated split-plot screening experiments**
Presenter: **Emily S Matthews**, University of Southampton, United Kingdom

A key step in early industrial experimentation is screening to identify those factors that, when their levels are varied, have a substantive impact on the measured response. It is common for such experiments to investigate many factors in only a small number of runs, and increasingly supersaturated designs, with fewer runs than factors, are employed. The runs of the experiment are also often grouped by the levels of one or more hard-to-change, or whole-plot, factors, resulting in a split-plot structure and the need to account for correlation between observations from units in the same whole-plot. Motivated by an example from materials engineering, new approaches to both the design and analysis of supersaturated split-plot experiments are discussed. A linear mixed model is assumed to describe the response, and methods for optimal design, model selection, and variance-component estimation are developed and presented. The methodology is demonstrated on a range of examples, reflecting realistic industrial and scientific experiments, including the motivating engineering application.

EO1403: **Assessing the random effects part of mixed models**
Presenter: **Reza Drikvandi**, Imperial College London, United Kingdom

Correctly specifying the random effects part of mixed models is crucial for both the design and the power considerations. Since random effects are latent and unobservable quantities, it is challenging to decide which random effects should be included into the model. Inclusion or exclusion of random effects from the model is equivalent to test whether or not their variance components equal zero. However, test for zero variance components is a nonstandard problem since the null hypothesis in on the boundary of the parameter space. We propose a permutation test for variance components of random effects which avoids issues with the boundary of the parameter space. Another challenging task regarding the random effects part is to make sure the assumed distribution for random effects is correctly specified. An inappropriate random-effects distribution would result in model misspecification which could lead to biased parameter estimates as well as a poor power. We introduce a likelihood-based diagnostic test based on the so-called gradient function to assess the random-effects distribution. We establish asymptotic properties of our diagnostic test and also develop a parametric bootstrap algorithm for small sample situations. Our strategy can be used to check the adequacy of any distribution for random effects in a wide class of mixed models, particularly non-linear mixed models, with univariate as well as multivariate random effects.

EO166 Room CLO 306 RECENT DEVELOPMENT OF EXPERIMENTAL DESIGNS Chair: Chang-Yun Lin

EO1351: **Response surface methodology using split-plot definitive screening designs**
Presenter: **Chang-Yun Lin**, National Chung Hsing University, Taiwan
Co-authors: Po Yang

Definitive screening designs are a new class of three-level designs. We investigate the performance of definitive screening designs in split-plot structures for one-step response surface methodology. The result of the projection eligibility and the study of D-efficiency and I-efficiency show that split-plot definitive screening designs perform well when the number of important factors is small. To reduce the risk of being unable to fit second-order models for response surfaces, we provide the column indexes of projections. Experimenters can assign potentially important factors to those columns to avoid ineligible projections. An example is presented to demonstrate how to analyze data for response surface methodology using the split-plot definitive screening design.

EO0410: **Generalized resolution for factorial designs**
Presenter: **Ulrike Groemping**, Beuth University of Applied Sciences Berlin, Germany

The confounding structure of factorial designs can be investigated using the generalized word length pattern, which implies the resolution as the length of the shortest word. For factorial designs with only two-level factors, resolution was previously refined to generalized resolution (GR): GR increases the resolution by the distance to worst case confounding and has an interesting geometrical interpretation. We have recently developed squared canonical correlations and average R^2 values as means for supplying two types of generalized resolution for general factorial designs, including designs for factors at more than two levels and mixed level designs. The proposed generalized resolutions are coding invariant, i.e., they are particularly suitable for designs with qualitative factors. The squared canonical correlations and average R^2 values are also of interest in themselves, and can also be used for factor-specific generalized resolutions, i.e., factor specific worst-case considerations.

EO0317: **Blocked semifoldovers of two-level orthogonal designs**
Presenter: **Po Yang**, University of Manitoba, Canada
Co-authors: Chang-Yun Lin, William Li

Follow-up experimentation is often necessary to the successful use of fractional factorial designs. When some effects are believed to be significant but cannot be estimated using an initial design, adding another fraction is often recommended. As the initial design and its foldover (or semi-foldover) are usually conducted at different stages, it may be desirable to include a block factor. We study the blocking effect of such a factor on foldover and semifoldover designs. We consider two general cases for the initial designs, which can be either unblocked or blocked designs. In both cases, we explore the relationships between semifoldover of a design and its corresponding foldover design. More specifically, we obtain some theoretical results on when a semifoldover design can estimate the same two-factor interactions or main effects as the corresponding foldover. These results can be important for those who want to take advantage of the run size savings of a semifoldover without sacrificing the ability to estimate important effects.

EC1462: **Near optimum estimation of variance components in mixed effects models**
Presenter: **Subir Ghosh**, University of California, United States

The variance component estimation for a mixed effects model is considered for analyzing the data obtained from a block design experiment. The optimum estimation method like uniformly minimum variance quadratic unbiased estimation (UMVQUE) is often not possible to carry out. Two near optimum estimation methods are proposed without making any distributional assumptions for the data. Examples are worked out for the real data. For the simulated data generated by assuming the multivariate normal model, the optimal estimators when they exist as well as the near optimal estimators when they do not exist are compared with the maximum likelihood (ML) estimators and restricted maximum likelihood (REML) estimators.

EO1743: **An analysis of the career checkup using selective multi-group principal component regression and ridge regression**
Presenter: **Sho Kawasaki**, Mejiro University, Japan
Co-authors: Genso Watanabe, Takenori Takahashi

Employees job awareness, as evidenced in taking initiatives and self-motivation, became an extremely important topic from a career development perspective. We discuss the principles of career support which affects the job consciousness of employees in small and medium enterprises, based on the Selective Multi-group Principal Component Regression and Ridge Regression. The Career Checkup model developed by the Ministry of Health, Labor and Welfare, the questions covered a wide range and consisted of many items. In analyzing surveys with many questions, the existence of multiple cases of high correlation between questions is not uncommon. We propose the Selective Multi-group Principal Component Regression (SPCR) method based on response data. Furthermore, we compare SPCR with a Ridge estimator to validate the application of SPCR. The results

of the analysis, which used Selective Multi-group Principal Component Regression methods and validation in Ridge Regression, showed that the employees whose perception of their career within the company is midlevel experienced the greatest effect on job consciousness from career support measures which clarify their objectives. In career research until now, most focused on qualitative case studies. Quantitative/scientific methodology is used to make this new approach meaningful.

EG263 Room MAL 540 CONTRIBUTIONS ON CENSORED DATA Chair: Claudia Czado

EC1297: **Nonparametric estimator of the latency in a mixture cure model**
Presenter: **Ana Lopez-Cheda**, University of A Coruna, Spain
Co-authors: Ricardo Cao, Maria Amalia Jacome Pumar
Current cancer treatments caused an increased ratio of cured patients or, at least, a long term survival. In order to accommodate the insusceptible proportion of subjects, a cure fraction can be explicitly incorporated into survival models and as a consequence, cure models arise. The goals in cure models are usually to estimate the cure rate (incidence) and the probability of survival of the uncured patients up to a given point of time (latency). A completely nonparametric approach for mixture cure models is introduced, as an alternative to the current parametric and semiparametric methods in the literature. The nonparametric estimator of the latency is presented, which is the local maximum likelihood estimator. Furthermore, the iid representation and the asymptotic mean squared error are obtained. The choice of the optimal bandwidth is addressed by a bootstrap selection method. Its behavior is assessed in a simulation study. Finally, the methodology is applied to a database of colon cancer of CHUAC (University Hospital of A Coruna) patients to determine the prognosis based on, among other variables, the age.

EC1555: **Vine copula based inference of multivariate survival data**
Presenter: **Nicole Barthel**, Technische Universitaet Muenchen, Germany
Co-authors: Claudia Czado, Paul Janssen, Candida Geerdens
Multivariate survival data are characterized by certain types of incompleteness. For right-censored data the event of interest has not occurred for all observations within the study period. The resulting lack of information about the data at hand needs to be taken into account when modeling the dependence structure of clustered event times. Visualization tools of bivariate right-censored data in the copula space are developed. Further, vine copulas are used to make parametric maximum likelihood inference for four dimensional right-censored survival data. A modified likelihood function in terms of vine copula components is provided. The estimator's small sample performance is investigated through extensive simulations. A real data set from veterinary medicine is used to demonstrate the viability of the derived model and its estimation.

EC1735: **Cumulative/dynamic ROC curve estimation**
Presenter: **Sonia Perez-Fernandez**, University of Oviedo, Spain
Co-authors: Pablo Martinez-Camblor
The ROC curve is a popular graphical method frequently used in order to study the diagnostic capacity of continuous (bio)markers. When the considered outcome is a time-dependent variable, the direct generalization is known as cumulative/dynamic ROC curve. For a fixed point of time t, an individual is allocated to the positive group if the event happens before t and to the negative group if the event has not happened at t. The presence of censored individuals, which cannot be directly assigned to a group, is the main handicap of this approach. The proposed cumulative/dynamic ROC curve estimator assigns a probability of belonging to the negative (positive) group to the individuals censored previously to t. The performance of the resulting estimator is studied from Monte Carlo simulations. Some real- world applications are reported. Results suggest that the new estimators provide a good approximation to the real cumulative/dynamic ROC curve.

EC1425: **Diagnostic checks for mixture cure models with interval censored data**
Presenter: **Sylvie Scolas**, Universite catholique de Louvain, Belgium
Co-authors: Anouar El Ghouch, Catherine Legrand, Abderrahim Oulhaj
In elderly population, studying the time until Mild Cognitive Impairment (MCI) can be crucial since it is considered as a precursor of Alzheimers disease. However, not everyone will develop signs of MCI. We refer to this as the presence of a cure fraction. Another characteristic of these data, also found in other medical studies, is that patients are followed up at scheduled interviews, leading to interval-censored event-times, in addition to the possibility of being right-censored. When considering diagnostic tools in survival models, the idea of residual checking, widely used in linear regression, must be adapted to take account of right-censoring. Proposals have been made, mainly in the context of right-censored Cox and AFT regression models. For example, martingale residuals are commonly used to detect non-linearity of a covariate. While cure models are gaining into popularity, literature on residual-based model checking in this context is very sparse, and to the best of our knowledge nonexistent in the case of interval-censored data and cure. Clearly, an adaptation to mixture cure models is not straightforward and worth further study. We discuss the extension of existing methods to the presence of interval-censoring, and the difficulties encountered when applying these ideas to the cure mixture model. We illustrate our results based on simulations and on real data from an Alzheimer study.

EC1474: **Computing C-statistic with interval-censored data**
Presenter: **Kenichi Hayashi**, Keio University, Japan
Evaluation of statistical models is one of the most important steps in the process of data analysis. In particular, the relationship between a response variable and related covariates must be evaluated to establish better statistical models. This issue often arises when a novel biomarker is investigated in biomedical research. We discuss how to compute the C-statistic when the response is a time-to-event variable subject to interval-censoring. The C-statistic is seen as a generalization of the area under the ROC curve for a binary response. The central issue of interval-censored data is that the event time is not observed exactly but known in an interval. This situation does not enable us to apply the concept of usable pairs presented in the previous research dealing with right-censored time-to-event data. We extend this concept and propose the C-statistic for interval- censored data. The simple algorithm and some numerical examples are given to show how well the proposed method works.

EG041 Room CLO 102 CONTRIBUTIONS IN COMPUTATIONAL AND NUMERICAL STATISTICS Chair: Lutz Duembgen

EC1394: **A stochastic estimation algorithm with efficient sampling in high dimension: Application to deformable template model**
Presenter: **Estelle Kuhn**, INRA, France
Co-authors: Stephanie Allassonniere
Deformable template models aim at summarizing through a template and some geometrical characteristics a population of images representing the same object. The issue is to propose an accurate algorithm with low computational cost and with theoretical guaranties of relevance for estimating simultaneously the template and the geometrical characteristics. We consider the Bayesian mixed effect template model. We propose to use a stochastic approximation of the Expectation-Maximization (EM) algorithm where the simulation of the latent variable is performed through a Monte Carlo Markov Chain (MCMC) method. We introduce a new Anisotropic Metropolis Adjusted Langevin Algorithm to optimize the proposal of the MCMC method. We first prove that this new sampler leads to a geometrically uniformly ergodic Markov chain. We prove then in a general latent variable model that under mild conditions, the estimated parameter sequence generated by the proposed stochastic approximation of EM algorithm converges almost surely toward a local maximum of the likelihood and is asymptotically Gaussian distributed. The proposed estimation

algorithm can be used for a large range of models in many fields of applications such as pharmacology or genetic. The methodology developed is tested on handwritten digits of the USPS database and on some 2D images representing mandibles of mice for the deformable model estimation.

EC1802: **Piecewise monotonic approximation for estimating the extreme values of a function from large sets of noisy measurements**
Presenter: **Ioannis Demetriou**, University of Athens, Greece
We consider the problem of estimating the extrema of a univariate function from noisy measurements of its values by the least squares piecewise monotonic data approximation method. This method makes the smallest change to n data so that the piecewise linear interpolant to the smoothed values consists of k monotonic sections, where k is a prescribed positive integer. The positions of the joins of the monotonic sections are unknowns of the optimization process whose optimal values are determined automatically, which is a combinatorial problem that can have $O(n^k)$ local minima. However, the method gives a global solution in $O(n^2+kn\log_2 n)$ computer operations, which is far less than the number of local minima that can occur. We show the efficiency of the method on large sets of noisy observations. Our results suggest some subjects for future research, as for example in automatic peak finding, which is a subject of continuous interest in spectroscopy and chromatography. Other examples arise from financial mathematics, from medical image processing and data science for instance.

EC1485: **IsoGeometric Smoothing: A new approach for smoothing on surfaces**
Presenter: **Matthieu Wilhelm**, Universite de Neuchatel, Switzerland
Co-authors: Luca Dede, Laura Sangalli, Wilhelm Pierre
We propose a new approach for smoothing on surfaces. More precisely, we aim at estimating functions lying on a surface represented by NURBS, which are geometrical representations commonly used in industrial applications. The estimation is based on the minimization of a penalized least-square functional which is equivalent to solve a 4th-order Partial Differential Equation (PDE). We use Isogeometric Analysis (IGA), which is a method for numerically solve PDE, for the numerical approximation of such surface PDE, leading to an IsoGeometric Smoothing (IGS) method. IGA has the great advantage to use the exact geometrical representation of the surface in the analysis, thus avoiding complex meshing procedures. IGA also provides at least globally C^1-continuous basis functions with compact support. We show the performance of the proposed IGS method by means of numerical simulations and we apply it to the estimation of the pressure coefficient, and associated aerodynamic force on a winglet of the SOAR space shuttle.

EC0424: **Simulation and quadrature rules: Implications on the performance of random coefficient models of demand**
Presenter: **Daniel Brunner**, University of Duesseldorf, Germany
The random coefficient model of demand requires the evaluation of market share integrals at every step of an iterative procedure. Taking into account unobserved consumer heterogeneity will result in non analytic solutions to these market share integrals. Simulation techniques are the most frequently used numerical approach to solve such kind of problems. Another method, namely sparse grids, performed better in recent BLP studies in terms of precision and computational cost. A new numerical method for BLP models is suggested. It is referred to as adaptive sparse grids. It increases parameter precision and reduces computational cost of a BLP demand estimation even further. A simulation study evaluates the different numerical integration approaches and gives a clear recommendation of more powerful integration methods.

EC0208: **Computation of optimal stratum boundaries using multiple auxiliary variables**
Presenter: **Karuna Reddy**, The University of the South Pacific, Fiji
Co-authors: M G M Khan, Dinesh Rao
The computation of determining Optimal Stratum Boundaries (OSB) based on the survey variable is not feasible in practice since the variable of interest is unavailable prior to conducting the survey. A methodology is presented to compute the OSB for a study variable based on multiple auxiliary variables, which are readily available and regressible with the study variable. The auxiliary variables used for this computational problem are estimated to follow skewed 3P Weibull distributions. The problem is formulated into a Mathematical Programming Problem (MPP) that seeks minimization of the variance of the estimated population parameter. The formulated MPP is then solved for the OSB using a dynamic programming (DP) technique. A numerical example with a real data set, aiming to estimate the Haemoglobin content in women in a national Iron Deficiency Anaemia survey, is presented to illustrate the methodology. The computational results obtained by the proposed method are compared with other univariate methods and the results reveal that it yields a substantial gain in the precision of the estimates.

EG011 Room MAL 402 CONTRIBUTIONS ON BAYESIAN COMPUTATIONS Chair: Michael Daniels

EC1656: **Reparameterisations for location-scale mixtures**
Presenter: **Kate Lee**, Auckland University of Technology, New Zealand
Mixture models have been used in a wide variety of applications. The construction of reference Bayesian analysis for mixture models has been very challenging and still is not unsolved. We introduce the global mean-variance reparameterisation for mixture models that main consequence is to have all other parameters within a compact space. For Gaussian mixtures, a genuine non-informative prior is developed and the posterior distribution associated with this prior is almost surely proper as it is supported by simulation studies. While we only study the Gaussian case, extension to other classes of location-scale mixtures is straightforward. The MCMC implementation constrained on simplex equations to sample from a complex space is discussed.

EC1805: **Pseudo-marginal MCMC for parameter estimation in α-stable distributions**
Presenter: **Marina Riabiz**, University of Cambridge, United Kingdom
Co-authors: Fredrik Lindsten , Simon Godsill
The α-stable distribution is very useful for modelling data with extreme values and skewed behaviour. The distribution is governed by two key parameters, tail thickness and skewness, in addition to scale and location. Inferring these parameters is difficult due to the lack of a closed form expression of the probability density. We develop a Bayesian method, based on the pseudo-marginal MCMC approach, that requires only unbiased estimates of the intractable likelihood. To compute these estimates we build an adaptive importance sampler for a latent variable representation of the α-stable density. This representation has previously been used in the literature for conditional MCMC sampling of the parameters, and we compare our method with this approach.

EC1298: **Classification of Gaussian Markov random field observations based on Bayes discriminant functions**
Presenter: **Kestutis Ducinskas**, Klaipeda University, Lithuania
Co-authors: Lina Dreiziene
The problem of classifying Gaussian Markov random field (GMRF) observations into one of two populations specified by conditional autoregressive model (CAR) is considered. We deal with classification procedures associated with Bayes discriminant function (BDF) and plug in Bayes discriminant function (PBDF) based on maximum likelihood (ML) estimators of parameters. The overall misclassification probability associated with BDF and actual error for PBDF are derived. These are used as the performance measures of classification procedures. The second and third order neighbourhood schemes on regular 2-dimensional lattice are used for illustrative examples. The effect of Mahalanobis distance and prior probabilities on the performance of proposed classification procedures for various spatial sampling designs is numerically evaluated.

EC1598: **Approximate Bayesian computation for model choice**
Presenter: **Clara Grazian**, Sapienza Universita di Roma, Italy

Approximate Bayesian computation is a class of algorithms which is now an essential tool for handling complex models, for which the likelihood function is intractable. The idea is to compare the observed data set with a simulated data set by using a summary statistics bringing information on the parameter; unfortunately, in most real applications the summary statistics is unlikely to be sufficient. The ability of closely approximating the posterior distribution of the parameter of interest strongly depends on the choice of this statistics; some works which study how to choose the summary statistics already exist for inferential problems, nevertheless this aim is much more complicated to achieve in problems of model choice. It is now well known that there is a substantial discrepancy between the true Bayes factor and its ABC approximation, except for very few cases. We propose a way to compare observed and simulated data set inside an ABC algorithm based on the Bayes factor for a simpler version of the considered models and apply the methodology in the particular setting of quantiles distributions, a very flexible tool to model data far from normality (highly skewed or kurtotic), but which are difficult to manage without introducing ABC. The methodology only needs to be able to approximate the available models with simpler models; the distance between the models to compare and the models used as approximations gives a measure for the performance of the algorithm.

EC1582: **The Bayesian analysis toolkit version 1.0**
Presenter: **Allen Caldwell**, Max Planck Institute for Physics, Germany
Co-authors: Daniel Greenwald, Kevin Kroeninger, Frederik Beaujean, Oliver Schulz

BAT, the Bayesian Analysis Toolkit (http://mpp.mpg.de/bat), is an open-source multi-threaded C++ package providing Markov-chain Monte Carlo sampling of arbitrary user-defined models. It is developed for the particle physics community and contains several ready-to-use common-case models. It has been used extensively in particle, astroparticle, and nuclear physics since its initial release in 2008; and its usage is being evaluated for the fields of biotechnology and medical imaging. We present version 1.0, which includes numerous improvements to both the algorithms and programming interface, highlighting in particular the creation of a community-expandable suite of common priors and the ability to share posterior samples without sharing data or the internal workings of a model.

EC036 Room MAL 539 CONTRIBUTIONS IN APPLIED STATISTICS AND DATA ANALYSIS Chair: Yves Berger

EC1669: **Variable selection in high dimensional regression: A non-parametric procedure for business failure prediction**
Presenter: **Marialuisa Restaino**, University of Salerno, Italy
Co-authors: Alessandra Amendola, Francesco Giordano, Maria Lucia Parrella

The variable selection is a challenging task in statistical analysis. In many real situations, a large number of potential predictors are available and selection among them is recommended. There exists a large number of methods that can be used, however they can be applied only if the assumptions of linearity and additivity hold on and if the number of predictors is less than the number of observations. Therefore it is important to develop a variable selection procedure stable in presence of non-linearity and non-additivity and also when the number of predictors is large. This is particularly important in some real applications, such as in business failure prediction, where the financial ratios have common items in nominator and denominator, leading to multicollinearity and redundancies and consequently the hypotheses of linearity and additivity are often violated. Therefore we evaluate the performance of the non-linear and non-additive method on a sample of firms that have experienced the bankruptcy and for which a set of financial ratios has been observed.

EC1654: **The delay vector variance (DVV) method and recurrence quantification analysis of energy markets**
Presenter: **Emmanuel Senyo Fianu**, Leuphana University of Lueneburg, Germany

Since deregulation, energy markets have experience high amounts of volatility which is evidenced by huge spikes inherent in these markets. The identification of power spikes and the likelihood of increased volatility in energy prices is a topic of major concern. The recently proposed delay vector variance (DVV) method is employed, which examines local predictability of a signal in the phase space to detect the presence of deterministic and non-linearity in time series. The DVV approach utilizes optimal embedding parameters, which are obtained via a differential entropy, based method using wave-based surrogates. The concept of (cross)-recurrence quantification analysis is used to study energy markets to locate hidden patterns, non-stationarity, presence of potential and well-pronounced spikes and to examine the nature of these plots in the event of crisis in the energy sector as well as financial crisis. Specifically, the recurrence plots are utilized to detect and characterize seasonal cycles. The focus is on some emerging European power markets that could serve as a benchmark for analyzing other power markets in the region and beyond. The employed methodology is shown to be useful in the diagnosis and detection of potential power spikes, which is significantly impacted by economic downturns and other main drivers of imbalances in matching demand to supply in the power markets in recent years.

EC1390: **Towards an education monitoring index**
Presenter: **Audrone Jakaitiene**, Vilnius university, Lithuania
Co-authors: Dovile Stumbriene, Rimantas Zelvys

Information about education monitoring indices is very diverse and limited, especially over time. Given that education is one milestones, which guarantees the well-being of a country, the understanding about the education system status might be crucial. However there are plenty of education variables and a single index, which incorporates big data and gives notion about the system status, might be essential. The European Commission within the Education and Training Monitor initiative calculates indices towards Europe 2020 strategy. Each country's indices are calculated normalizing data with respect to EU28 average. We develop a methodology to construct an Education Monitoring Index for a country using country or region data. For methodologically consistent data for a country, we use Europe 2020 strategy main indicators as starting dataset. The objective is to calculate an Education monitoring index for Lithuania, Latvia, Estonia, United Kingdom, Finland and Germany available over time.

EC1407: **Decomposition of gender wage inequalities through calibration: Application to the Swiss structure of earnings survey**
Presenter: **Mihaela Catalina Anastasiade**, University of Neuchatel, Switzerland
Co-authors: Yves Tille

The wage decomposition method assumes that the difference between the average wage of men and the average wage of women comprises two parts, an explained part (due to differences in the characteristics of the two groups) and an unexplained part (due to discrimination). A linear relationship between the wage and the characteristics is assumed. A key element in the computation of the two parts is the counterfactual average wage of women, regarded as the average wage of women if they had the same average characteristics of men. A semi-parametric method based on a logistic model has been proposed, allowing the construction of the entire counterfactual wage distribution of women. This enables the measurement of the unexplained part of the wage difference at other points, such as quantiles. However, the unexplained part obtained with this method is either over- or underestimated, depending on the model assumed. We propose calibration to generate the counterfactual wage distribution of women without making assumptions about the model behind the data. We present two calibration instances, in an attempt to show that they represent an improvement to the two methods mentioned above.

EC1797: **How to find the best publication strategy**

Presenter: **Lisiane Schnegelsberg**, Martin-Luther-University Halle-Wittenberg, Germany
Co-authors: Steffen Liebscher, Thomas Kirschstein

The amount of publications in peer-reviewed scientific journals and their quality are popular metrics for assessing a researcher's performance. Hence, an individual aiming to optimize his/her performance has to make decisions about the journals where his/her research is to be published. The journals differ w.r.t. various criteria such as impact, review time, review system, etc. Based on the individual's priorities (and other constraints) different publication strategies can be deduced. The Graphical Evaluation and Review Technique (GERT) is used for describing the publication process. This technique allows to model each step in the publication process taking into account stochastic components (i.e. probability of acceptance or review time). This way the publication process can be assessed by performance indicators (as expected publication time or acceptance probability) which helps selecting an appropriate journal considering individual preferences. The applicability of the model to derive the optimal publication strategy will be illustrated with real-world data.

EC0266: **Modelling rice yield in the Philippines using dynamic spatio-temporal models**

Presenter: **Stephen Jun Villejo**, University of the Philippines, Philippines

One stimulus of the development of these spatio-temporal models is in modelling agricultural yield which accounts for spatial and temporal dependencies. Three spatial-temporal models with varying dynamic characteristics of the parameters are postulated with a different estimation procedure for the agricultural yield in irrigated areas of the Philippines each. One of them is a robust estimation procedure using forward search algorithm with bootstrap in a backfitting algorithm framework. The other two algorithms also used the backfitting algorithm but infused with the Cochranne-Orcutt procedure. The robust estimation procedure and the other one considering varying parameter across space gave competitive predictive abilities and are better than the ordinary linear model. Moreover, simulation studies show the superiority of the robust estimation procedure over the Cochranne-Orcutt procedure and ordinary linear model in the presence of structural change.

Sunday 13.12.2015 08:45 - 10:25 Parallel Session G – CFE-CMStatistics

CO510 Room MAL B29 ECONOMETRIC ANALYSES OF SPILLOVERS AND LIQUIDITY Chair: Pierre Siklos

CO0161: **Monetary policy spillovers: A global empirical perspective**
Presenter: **Pierre Siklos**, Wilfrid Laurier University, Canada
Co-authors: Domenico Lombardi, Samantha St Amand
New light is shed on the spillovers from US unconventional monetary policies by examining the behavior of select financial asset prices at the daily frequency and by incorporating a crucial element in the conduct of monetary policy, namely the tone of verbal announcements by central banks. We eschew the event study approach and adopt a time series perspective. Monetary policy surprise easings are found to decrease yields in most economies since October 2008. The size of the response varies considerably across the economies examined, but for 10-year government bonds, they range from a 15 to 30 basis points reduction in yields in the United States, the United Kingdom and the Euro zone. The impact was found to be larger on long-term sovereign bonds than on shorter- term assets, and stronger in major economies, such as the United Kingdom and the Euro zone, and in Canada. The analysis also provides evidence that the tone of US Fed communication and a countrys own central bank policy statements, constructed using text analysis software, have a significant impact on asset prices. The impact of the tone of these statements has been notably stronger since the start of the crisis, with financial market reactions appearing to be relatively more sensitive to the content of statements during the post-crisis period.

CO0352: **Price discovery in European agricultural markets: When do futures contracts matter?**
Presenter: **Philipp Adaemmer**, University of Muenster, Germany
Co-authors: Martin T Bohl, Oliver von Ledebur
The literature on price discovery in agricultural commodity markets is predominantly devoted to North America. The aim is to extend the analysis to Europe whose changes in market structure offer an interesting setting to investigate the relevance of futures markets for the price discovery process during times of price turmoil. By applying a time-varying vector error correction model with heteroscedastic disturbances we show that the relative contribution of the futures market to price discovery was higher during the first period of price turmoil (from 2007 to 2009) than during the second one (from 2010 to 2013). This is interesting as liquidity was low during the first period but high during the latter. Our findings refine the argument of a positive relationship between trading volume and price discovery. There is no doubt that liquidity is crucial for the efficient functioning of futures markets. However, once a certain level of trading activity is reached, other factors become more influential. Previous research implies that agricultural markets were hit by surprise during the first period of price turmoil but they anticipated the second one. Futures contracts therefore especially matter during times of unanticipated price shocks.

CO0418: **Asymmetric connectedness on the U.S. stock market: Bad and good volatility spillovers**
Presenter: **Evzen Kocenda**, Charles University, Czech Republic
Co-authors: Jozef Barunik, Lukas Vacha
The purpose is to suggest how to quantify asymmetries in volatility spillovers that emerge due to bad and good volatility. Using data covering most liquid U.S. stocks in seven sectors, we provide sample evidence of the asymmetric connectedness of stocks at the disaggregate level. Moreover, the spillovers of bad and good volatility are transmitted at different magnitudes that sizably change over time in different sectors. While negative spillovers are often of substantial magnitudes, they do notstrictly dominate positive spillovers. We find that the over all intra-market connectedness of U.S. stocks increased substantially during the financial crisis.

CO1181: **Financial integration, capital flows and economic performance: Evidence from a global vector error correction model**
Presenter: **Ansgar Belke**, University of Duisburg-Essen, Germany
Co-authors: Joscha Beckmann
We apply a global vector autoregression (GVAR) model which is a suitable framework for empirically identifying spillovers and shock transmission across countries on a global scale. Our model consists of individual country VAR models describing the countrys economy which are then consistently linked into a single multi-country model using weights relating to the trade and financial linkages across countries. Variables include GDP, inflation, interest rates, monetary aggregates and capital flows. Using this error-correction model, we analyse the role of international capital flows and their role for the international transmission of cyclical shocks based on the dissection of long-run and short-run dynamics.

CO639 Room MAL G15 NOWCASTING AND FORECASTING UNDER UNCERTAINTY III Chair: Boriss Siliverstovs

CO0293: **Short-term forecasting with mixed-frequency data: An appraisal of supervised factor and shrinkage methods**
Presenter: **Boriss Siliverstovs**, KOF ETHZ, Switzerland
A wide variety of variable selection, dimension reduction and shrinkage methods for forecasting with mixed-frequency data is applied. In particular, we assess the effectiveness of supervised learning using the following algorithms for variable pre-selection: Least Absolute Schrinkage and Selection Operator (LASSO), partial least squares (PLS), principle covariate regression (PCovR), bagging algorithm (short for boostrap aggregation), boosting algorithm, the non-negative garotte (NNG) algorithm as well as recently suggested the three-pass regression filter. Big data become even bigger due to the use of the skip-sampling or blocking approach for frequency conversion of high-frequency variables, making the task of selecting most informative variables regarding the target variable even more challenging. Assessing which method or their combination produces most reliable results bears a direct relevance for forecasting practitioners.

CO1183: **Nowcasting and forecasting with heavy tails in macroeconomics**
Presenter: **Nicholas Fawcett**, Bank of England and Centre for Macroeconomics, United Kingdom
Co-authors: Andrew Harvey, Martin Weale
Although most macroeconomic forecasting models assume that errors follow a Gaussian distribution, in many cases this is invalid: large outliers are much more likely than the models allow. We use a new class of state space models in which errors can have heavy-tailed distributions to produce better nowcasts and short-term forecasts of macroeconomic series. These models offer a natural way of down-weighting large outliers when they occur, which provides robustness to noise, allowing us to distinguish better between structural breaks and outliers. We apply the framework to a range of series, including nominal wages, GDP, and a small macroeconomic system including oil prices, GDP, inflation and interest rates. The framework allows for variation in the underlying location and scale variables over time.

CO1316: **The Bayesian estimation, analysis and regression toolbox for forecasting and policy analysis**
Presenter: **Bjoern van Roye**, European Central Bank, Germany
We have developed a new toolbox for forecasting and policy analysis, the so-called "Bayesian Estimation and Analysis Regression Toolbox (BEAR)". BEAR is a comprehensive (Bayesian) VAR toolbox (based on MATLAB) where the programme is easy for non-technical users to understand, augment and adapt. In particular, the BEAR toolbox includes a user-friendly graphical interface which allows the tool to be used by country desk economists. Furthermore, the toolbox is well documented, both within the code as well as including a detailed theoretical and user's

guide. The toolbox includes state-of-the art applications such as sign and magnitude restrictions, unconditional and conditional forecasts, Bayesian forecast evaluation measures (continuous ranked probability scores), tilting of distribution, etc.

CO1174: **On the use real-time data to evaluate macroeconomic forecasting indicators**
Presenter: **Rolf Scheufele**, Swiss National Bank, Switzerland

When looking at typical nowcast problems, e.g. to evaluate the relative importance of hard versus soft data, the average forecasting performance of using last vintage data and the use of current vintage data give very close results. We investigate the forecasting performance of typical indicator models for nowcasting German GDP. We then document the difference between final and current data-releases. Our leading indicator models consist of MF-VARs where we can directly take into account the ragged-edge and mixed-frequency problem. More specifically, we analyse the relative predictive power of surveys (ifo and zew) relative to hard data (industrial production, orders) when data revisions are taken into account. One question we ask is whether data revisions on ip and orders (which are substantially revised over time) have any impact on their relative importance compared to soft data which are rarely revised.

CO544 Room MAL B33 MODELLING AND COMPUTATION IN MACRO-ECONOMETRICS Chair: Eric Eisenstat

CO0316: **Permanent and transitory components of house prices fluctuations**
Presenter: **Luca Benati**, University of Bern, Switzerland

Fluctuations in U.S. real house prices are overwhelmingly transitory. Whereas cointegrated SVARs identified via long-run restrictions produce fragile and sometimes implausible results, VARs in levels robustly identify three large, statistically significant, and highly persistent transitory deviations of house prices from their stochastic trend since the early 1970s. Evidence that this methodology would have allowed policymakers to detect the recent housing bubble in real time is however weak and inconclusive. Permanent shocks to house prices generate transitory housing market booms, with hump-shaped increases in both hours in construction and housing starts. We document a fundamental asymmetry between the inflating and unravelling phases of the recent housing bubble: whereas a single shock accounts for between one half and two-thirds of the deviation of real house prices from their stochastic trend during the former phase, no shock seems to have played a comparable dominant role during the latter phase.

CO0672: **Forecasting commodity currencies: The role of fundamentals with short-lived predictive content**
Presenter: **Francesco Ravazzolo**, Norges Bank and BI Norwegian Business School, Norway
Co-authors: Claudia Foroni, Pinho Ribeiro

Recent evidence highlights that commodity price changes exhibit a short-lived, yet robust contemporaneous effect on commodity currencies, which is only detectable in daily-frequency data. We use MIDAS models in a Bayesian setting to include mixed-frequency dynamics while accounting for time-variation in predictive ability. Using the Metropolis-Hastings technique as a new tool to estimate our class of MIDAS regressions, we find that for most of the commodity currencies in our sample, exploiting this short-lived relationship yields to statistically more precise out-of-sample exchange rate forecasts relative to the no-change benchmark. Further, the usual low-frequency predictors, such as money supplies and interest rates differentials, typically receive little support from the data at monthly forecasting horizons. In contrast, models featuring daily commodity prices are highly likely.

CO1056: **Estimation of continuous piecewise linear models using Bayesian clustering**
Presenter: **Eric Eisenstat**, The University of Queensland, Australia
Co-authors: Fabrizio Carmignani, Rodney Strachan, Rabee Tourky

We develop methods for efficiently estimating continuous piecewise linear functions based on a Bayesian approach to clustering. Specifically, we design MCMC algorithms that rely on clustering to generate hyperplanes and combine these with the Reisz estimator to construct a joint M-H proposal. The probabilistic approach circumvents the need to search exhaustively through a vast number of possible hyperplanes and thus surmounts the curse of dimensionality associated with existing methods. We use our method to estimate a nonlinear growth equation, nested within a VAR involving fiscal policy, monetary policy and output growth.

CO0879: **Reduced sources of error in time-varying parameter models**
Presenter: **Rodney Strachan**, The University of Queensland, Australia
Co-authors: Eric Eisenstat, Joshua Chan

There have been several advances in the estimation of large vector autoregressive models (VARs). Increasing the dimension of time-varying parameter VARs (TVP-VARs) remains a challenge. We propose an approach to increasing the number of variables we can model in a TVP-VAR that takes advantage of the strong correlations among the states. It has been shown, using principal component analysis, that the states can be modelled with a few factors. We specify a TVP-VAR with a reduced rank covariance matrix for the states such that we are able to significantly reduce the dimension of the states without reducing the dimension of the VAR. The specification of the reduced rank model induces manifolds as supports for the parameters. Using a judicious selection of parameter expansions and priors for the expanding parameters, we develop a specification that is fast, efficient and easy to compute. The original model specification and mapping to the expanded model uses homeomorphic transformations such that estimation and inference is invariant to the ordering of the states in the model.

CO468 Room MAL B34 MACROECONOMICS, ASSET PRICING, AND ROBUSTNESS Chair: Alexander Meyer-Gohde

CO0347: **Ambiguity and financial uncertainty in a real Business Cycle Model**
Presenter: **Hening Liu**, University of Manchester, United Kingdom
Co-authors: Yuzhao Zhang

Financial uncertainty, measured by the risk-neutral variance, is negatively related to aggregate quantities and the equity valuation, but is positively related to volatilities of quantities. We examine a production-based asset pricing model in which productivity growth follows a Markov process with a time-varying conditional mean and volatility, and the representative agent has ambiguity aversion. When the model is calibrated to unconditional moments of quantities and asset returns, the model can reproduce the relations between the risk-neutral variance and both the level and variation of quantities and returns. The model can generate a sizable variance risk premium.

CO0409: **Long-run risk is the worst-case scenario**
Presenter: **Rhys Bidder**, Federal Reserve Bank of San Francisco, United States
Co-authors: Ian Dew-Becker

The aim is to study an investor who is unsure of the dynamics of the economy. Not only there are parameters unknown, but the investor does not even know what order model to estimate. She estimates her consumption process non-parametrically and prices assets using a pessimistic model that minimizes lifetime utility subject to a constraint on statistical plausibility. The equilibrium is exactly solvable and we show that the pricing model always includes long-run risks. With risk aversion of 4.8, the model matches major facts about asset prices, consumption, and investor expectations. A novel link between ambiguity aversion and non-parametric estimation is provided.

CO0852: **Understanding expected returns**
Presenter: **Andrea Tamoni**, London School of Economics, United Kingdom
Co-authors: Daniele Bianchi
The predictive power of forecasting variables studied in the literature varies with the horizons. For example, the forecasting power of consumption-wealth ratio is particularly strong at short to intermediate horizons. On the other hand the dividend-price ratio tracks longer-term tendencies in asset markets. We show how to use the medium- and long-term information content of standard predictors to extract short-term expected returns. To combine information across horizons (or levels of resolution) we adopt a framework that relies on multi-scale models for time series. We show that our filtered expected return series exhibits aggregation properties which differs wildly from those of a standard autoregressive process of order one. In particular, our filtered monthly expected return series aggregated over the same temporal scale of the predictor used in the estimation, has an autocorrelation function which decays much slower than that obtained from aggregating a persistent, monthly AR(1) process. This result shows the importance of considering the scale at which the predictor conveys information about returns, as this may lead to different conclusions with respect to the high-frequency dynamics of the expected return process.

CO0780: **Model uncertainty and generalized entropy**
Presenter: **Alexander Meyer-Gohde**, Hamburg University, Germany
A generalization of the standard measure of entropy in multiplier preferences of model uncertainty is considered. Using this measure, we derive a generalized exponential certainty equivalent, which nests the standard exponential certainty equivalent, pervasive in Hansen-Sargent model uncertainty formulations, and the power certainty equivalent, common to Epstein-Zin-Weil-type of recursive preferences. Besides providing as a model uncertainty rationale to the power certainty equivalent of risk-sensitive preferences, the two parameter generalized exponential equivalent provides additional flexibility in modeling risk sensitivity and uncertainty. Applying the generalized exponential equivalent to the neoclassical growth model, we close the gap to the Hansen-Jagannathan bounds with a plausible detection error probability.

CO446 Room MAL B20 BAYESIAN NONPARAMETRIC ECONOMETRICS Chair: John Maheu

CO0354: **A Bayesian time-varying approach to risk neutral density estimation**
Presenter: **Roberto Casarin**, University Ca Foscari of Venice, Italy
Co-authors: Enrique ter Horst, German Molina
The purpose is to expand the literature of risk neutral density estimation across maturities from implied volatility curves, usually estimated and interpolated through cubic smoothing splines. The risk neutral densities are computed through the second derivative, which we extend through a Bayesian approach to the problem, featuring: (1) an extension to a multivariate setting across maturities and over time; (2) a flexible estimation approach for the smoothing parameter, traditionally assumed common to all assets, known and fixed across maturities and time, but now potentially different between assets and maturities, and over time; and (3) information borrowing about the implied curves and risk neutral densities not only across different option maturities, but also dynamically.

CO0501: **Particle learning for Bayesian non-parametric Markov switching stochastic volatility model**
Presenter: **Audrone Virbickaite**, Universidad Carlos III de Madrid, Spain
Co-authors: Hedibert Lopes, Concepcion Ausin, Pedro Galeano
The aim is to design a Sequential Monte Carlo (SMC) algorithm for estimation of Bayesian non-parametric Stochastic Volatility (SV) models for financial data. In particular, we make use of one of the most recent particle filters called Particle Learning (PL). The performance of this particle method is then compared with the standard Markov Chain Monte Carlo (MCMC) methods for non-parametric SV models. PL performs as well as MCMC, and at the same time allows for on-line type inference. The posterior distributions are updated as new data is observed, which is prohibitively costly using MCMC. Further, a new non-parametric SV model is proposed that incorporates Markov switching jumps. The proposed model is estimated by using PL and tested on simulated data. Finally, the performance of the two nonparametric SV models, with and without Markov switching, is compared by using real financial time series. The results show that including a Markov switching specification provides higher predictive power in the tails of the distribution.

CO0453: **A Bayesian semiparametric approach in a random coefficient demand framework**
Presenter: **Yong Song**, University of Melbourne, Australia
Co-authors: Dong-Hyuk Kim
A semiparametric Bayesian method is developed to analyze market share data of highly differentiated products. In particular, we use a Dirichlet process mixture (DPM) to account for the distribution of product-level unobserved heterogeneity in the random coefficient demand framework. While DPM is nonparametric, we obtain inferential efficiency from its parsimonious representation. Our empirical analysis shows promise in both model fitting and forecasting as well as provides an optimal market design for policymakers. Finally, the method is fast as it employs a Gibbs sampler with (known or simple) densities for all the model parameters and latent consumer choices.

CO0516: **Empirical relevance of ambiguity in first price auction models**
Presenter: **Dong-Hyuk Kim**, Vanderbilt University, United States
Co-authors: Gaurab Aryal
The aim is to study the identification and estimation of first-price auction models where bidders have ambiguity about the valuation distribution and their preferences are represented by maxmin expected utility. When entry is exogenous, the distribution and ambiguity structure are non-parametrically identified, separately from risk aversion, CRRA. We propose a flexible Bayesian method based on Bernstein polynomials. Monte Carlo experiments show that our method estimates parameters precisely, and chooses reserve prices with nearly optimal revenues, whether there is ambiguity or not. Furthermore, if the model is misspecified incorrectly assuming no ambiguity among bidders it may induce estimation bias with a substantial revenue loss.

CO1840: **Bayesian nonparametric time varying vector autoregression**
Presenter: **Maria Kalli**, Canterbury Christ Church University, United Kingdom
Co-authors: Jim Griffin
Although stationary time series models are theoretically appealing, macroeconomists consider them to be too restrictive. A popular alternative framework is time varying vector autoregression, with or without stochastic volatility (TVP-VAR or TVP-SV-VAR). Under this framework the parameters of the stationary vector autoregressive (VAR) model are allowed to change over time. This accounts for nonlinearity in the conditional mean, and heteroscedasticity in the conditional variance. We considered a Bayesian nonparametric stationary VAR (BayesNP-VAR) model and found that it outperformed the TVP-SV-VAR in terms of out-of-sample prediction for monthly macroeconomic series from the USA and Eurozone. Our aim is to extend the BayesNP-VAR model to a time varying parameter specification, creating a nonparametric, non-stationary model.

CO512 Room MAL B30 TOPICS IN FINANCIAL ECONOMETRICS Chair: Leopold Soegner

CO0376: **Forecasting discrete dividends by no-arbitrage**
Presenter: **Sascha Desmettre**, University of Kaiserslautern, Germany
Co-authors: Sarah Gruen, Frank Thomas Seifried
The aim is to develop and showcase a simple no-arbitrage methodology for the prediction of discrete dividend payments, based exclusively on market prices of options via the put-call parity. Our approach integrates all available option market data and simultaneously calibrates the market-implied discount curve, thus ensuring consistency across spot and derivative markets. We illustrate our method using stocks from the German blue-chip index DAX.

CO0577: **Stylized facts for regime-switching models**
Presenter: **Elisabeth Leoff**, University of Kaiserslautern, Germany
Co-authors: Joern Sass
Regime-switching models, in particular Hidden Markov Models, are widely-used in financial applications, due to their tractability and good econometric properties. We consider regime-switching models in discrete and continuous time with both constant and switching volatility. We examine which stylized facts they can reproduce and what this implies for the choice of model parameters. Our analysis is done both theoretically and with simulations, as for models with more than 2 states the transition probabilities cannot be calculated explicitly. Special attention will be paid to the structure of the linear and absolute autocorrelations, in particular for the model with 2 states.

CO0620: **Modeling creditworthiness using a generalized linear mixed-effects approach**
Presenter: **Laura Vana**, WU Wirtschaftsuniversitaet Wien, Austria
Co-authors: Kurt Hornik, Bettina Gruen
Based on a multivariate generalized linear mixed-effects approach, the aim is to model creditworthiness using key financial ratios from an extensive list of potential variables that drive a firm's financial health as covariates and ratings and default information as dependent variables. The model allows to estimate the latent creditworthiness as well as rater-specific biases. While the origin of bias in credit ratings has appeared in the literature under different assumptions, the proposed model allows to empirically quantify the rater-specific deviations from a firm's latent creditworthiness and analyze to which financial ratio the bias might be attributable. To assess if including variable selection based on a spike-and-slab prior and explicitly including financial ratios to model rater bias improves the model, the out-of-sample predictive performance of these models is compared.

CO0888: **NASDAQ trading pauses: Pacifiers or amplifiers**
Presenter: **Akos Horvath**, University of Vienna, Austria
Co-authors: Nikolaus Hautsch
In response to the Flash Crash a new type of circuit breakers, called trading pauses, were introduced on U.S. stock exchanges. Using a unique source of NASDAQ order book data, we take advantage of this quasi-natural-experiment situation and investigate for the first time whether the recently implemented control mechanism fulfills the intended regulatory role of addressing extraordinary volatility. Calculating high-frequency statistics and gathering a control group of pre-regulation extreme pricemovements, we implement a difference-in-differences framework to find that trading pauses enhance price discovery, however, this beneficial effect comes at the cost of higher volatility and bid-ask spread.

CO470 Room MAL 402 QUANTITATIVE ASSET MANAGEMENT Chair: Rafael Molinero

CO0421: **Using quantitative risk management as a trading tool in a commodities trading company**
Presenter: **Salim Boutaleb**, Invivo Trading, France
The aim is to review practical Risk Management tools we use in our trading department to compute, e.g., the position allocation size and the trailing stop-losses distances. Our Risk Management techniques are used as trading tools to optimize asset allocation across different time frames, markets, strategies and traders.

CO0907: **A speculative volume based covariance model for currency portfolios**
Presenter: **Guillaume Bagnarosa**, ESC Rennes, France
Co-authors: Gareth Peters, Matthew Ames
We propose a new forecasting model for the variance covariance matrix within which we parametrize the covariance matrix as a function of explanatory variables. This model displays interesting properties such as coping with the heteroskedasticity effect across the various explanatory variables. Furthermore, it can be estimated through a random effect representation making the EM-algorithm estimation possible. We finally apply this model to currency portfolio and consider the informative content of speculative volumes as predictive indicators of dependence among the portfolio components. We thus include different dependences structure for the high interest rates and low interest rates currencies which are linked together by the highly leveraged carry trade strategy.

CO0753: **Serial correlation and time-varying liquidity in the hedge fund industry**
Presenter: **Adrien Becam**, Paris-Dauphine University, France
Co-authors: Serge Darolles, Gaelle Le Fol
A well-known statistical feature of hedge funds returns is their serial-correlation. As the industry is lightly regulated, hedge funds use dynamic strategies involving illiquid and OTC assets, so there are two possible sources for this serial-correlation: the illiquidity of the underlying portfolios or the artificial smoothing of the reported returns. A new econometric model is proposed to disentangle between these two effects, and enables to estimate the dynamic process of serial correlation in hedge fund returns.

CO0769: **Active risk-based investing**
Presenter: **Emmanuel Jurczenko**, Ecole Hoteliere de Lausanne, Switzerland
Co-authors: Jerome Teiletche
Risk-based investment solutions are seen as incorporating no views. We propose an analytical framework that allows the introduction of explicit active views on expected asset returns in risk-based solutions. Starting from a Black-Litterman approach, we derive closed-form formulas for the weights of the active risk-based portfolio, and identify their main determinants. We discuss implementation aspects and show that our framework encompasses several other popular active investing methodologies. We illustrate the methodology with a multi-asset portfolio allocation problem using views based on macroeconomic regimes over the period 1974-2013.

CO502 Room MAL 415 PANEL DATA MODELS WITH COMMON FACTORS: THEORY AND APPLICATIONS Chair: Vasilis Sarafidis

CO0573: **A two stage approach to spatio-temporal analysis with strong and weak cross-sectional dependence**
Presenter: **Natalia Bailey**, Queen Mary University of London, United Kingdom
Co-authors: Sean Holly, Hashem Pesaran
An understanding of the spatial dimension of economic and social activity requires methods that can separate out the relationship between spatial units that is due to the effect of common factors from that which is purely spatial even in an abstract sense. The same applies to the empirical analysis of networks in general. We use cross unit averages to extract common factors (viewed as a source of strong cross-sectional dependence) and compare the results with the principal components approach widely used in the literature. We then apply multiple testing procedures to the de-factored observations in order to determine significant bilateral correlations (signifying connections) between spatial units and compare this to an approach that just uses distance to determine units that are neighbours. We apply these methods to real house price changes at the level of Metropolitan Statistical Areas in the USA, and estimate a heterogeneous spatio-temporal model for the de-factored real house price changes and obtain significant evidence of spatial connections, both positive and negative.

CO1213: **Nonlinear panel models with interactive effects**
Presenter: **Martin Weidner**, University College London, United Kingdom
Co-authors: Mingli Chen, Ivan Fernandez-Val
Estimation and inference are considered on semiparametric nonlinear panel single index models with predetermined explanatory variables and interactive individual and time effects. These include static and dynamic probit, logit, and Poisson models. Fixed effects conditional maximum likelihood estimation is challenging because the log likelihood function is not concave in the individual and time effects. We propose an iterative two-step procedure to maximize the likelihood that is concave in each step. Under asymptotic sequences where both the cross section and time series dimensions of the panel pass to infinity at the same rate, we show that the fixed effects conditional maximum likelihood estimator is consistent, but it has bias in the asymptotic distribution due to the incidental parameter problem. We characterize the bias and develop analytical and jackknife bias corrections that remove the bias from the asymptotic distribution without increasing variance. In numerical examples, we find that the corrections substantially reduce the bias and rmse of the estimator in small samples, and produce confidence intervals with coverages that are close to their nominal levels.

CC1398: **Least squares estimation of large dimensional threshold factor models**
Presenter: **Daniele Massacci**, Einaudi Institute for Economics and Finance, Italy
Large dimensional factor models are estimated under the maintained assumption that the factor loadings do not change over time. The aim is to study least squares estimation of large dimensional factor models subject to regime shifts in the loadings parameterized according to the threshold principle. We propose to estimate the unknown threshold value by concentrated least squares, and factors and loadings by principal components. The estimator for the threshold value is superconsistent, with convergence rate that depends on both the times series and the cross-sectional dimensions of the available panel, and it does not affect the estimators for factors and loadings: these have the same convergence rate they would have if the threshold was known. We further propose model selection criteria robust to the threshold effect. Empirical application of the model documents an increase in connectedness in financial markets during periods of high economic policy uncertainty.

CC1197: **A simple estimator for short panel data models with common factors**
Presenter: **Vasilis Sarafidis**, Monash University, Australia
A new methodology is proposed for estimating panel data models with a multifactor error structure using the GMM approach. The underlying idea involves substituting the unknown factors with time specific weighted averages of the variables included in the model. The gains of such strategy are threefold. First, the resulting estimation procedure becomes considerably simpler, since the unobserved factors are superseded with observed data. Second, given that the model is effectively parameterized in a more parsimonious way, the resulting estimators are more efficient than existing ones. Finally, the resulting moment conditions can be linearized in a straightforward way. Using simulated data we show that the performance of the proposed estimators is more than satisfactory for a wide range of specifications.

CO354 Room MAL B35 MEASURING AND FORECASTING DEFAULT RISK Chair: Alessandra Amendola

CO0765: **Bankruptcy survival: Evidence from France**
Presenter: **Lara Abdel Fattah**, Univerity of Paris Ouest Nanterre La Defense, France
Co-authors: Sylvain Barthelemy, Nadine Levratto, Benjamin Trempont
Despite the strong emphasis of the French legislator to the reorganization and restructuring of distressed companies, only a small share of insolvent firms benefit from a reorganization. In addition, among the rescued companies a large majority of these attemps fail as attested by the important number of conversion of reorganization into liquidation. Earlier research has explored the relationship between business survival and a variety of cases or company characteristics. However, little is known on the effectiveness of such a decision in the French case. Using a newly available dataset, we are able to document for the first time the survival of companies that have benefited from reorganization. The novelty of our approach comes from the use of data-driven comprehensive models of those relationships. We seek to explore a double question. The first one concerns the difference between companies which are still operating 3 years after the decision of reorganization by the commerce court and those which have been liquidated. The second one explores the probability of liquidation 1, 2 and 3 years after the judgement. In other terms it analyses the variables that help in predicting the probability of success of a reorganization. To answer these questions we implement support vector machines and random forest.

CO0811: **Variable selection methods for forecasting multiple business exits in Europe**
Presenter: **Alessandra Amendola**, University of Salerno, Italy
Co-authors: Marialuisa Restaino, Luca Sensini, Francesca Ametrano
The difficulties experienced by firms and institutions during the Global Financial Crisis (GFC) demonstrated the importance of understanding the determinants of financial default risk and investigating the differences between causes of failure and between industries, regions and countries. The aim is to identify the main variables that drive the financial distress across the European countries paying attention to the different reasons that may cause the exit from the market. An approach that takes into account different causes of failure has been implemented at both national and European levels, allowing us to study the single- country specificities as well as the between-country interdependencies. The most significant variables have been selected by means of some variable selection methods (stepwise, lasso, adaptive lasso, and so on), and all methods have been compared in terms of predictive ability by means of some accuracy measures, widely used in the business failure literature, in order to assess which procedure outperforms. Then, the sign of the selected variables is compared for each country model, in order to evaluate the differences in the determinants of financial distress and in the predictive ability of the model set-ups and to give an economic evaluation and interpretation of the models.

CO1228: **Advances on the use of mixture cure models in the credit risk context**
Presenter: **Lore Dirick**, KU Leuven, Belgium
Co-authors: Gerda Claeskens, Bart Baesens
Since the Basel accords, a higher need for more accurate credit risk calculations has arisen. This is where survival analysis gained a lot of importance in credit risk modeling. In this particular context, the survival function $S(t) = P(T > t)$ can be interpreted as the probability that a debtor will still be repaying his loan at timepoint t. With a high proportion of right-censoring because default does not take place in the entire loan lifetime, mixture cure models have become increasingly popular. This model uses the EM algorithm to divides the entire population into two parts: one part with debtors that are susceptible to default, and another part with debtors that are not. We propose an adequate model selection criterion based on Akaike's Information Criterion in order to perform variable selection in mixture cure models. As the exact likelihood cannot be computed in models when using the EM algorithm, this adapted criterion is based on the use of the complete-data log likelihood. Secondly, the model is adapted in order to make the inclusion of time-dependent covariates possible. Because of that, macro-economic factors that change over time can be included in the model.

CO1327: **Cyclicality and firm-size in private firm defaults**
Presenter: **Mamdouh Medhat**, Cass Business School, United Kingdom
The Basel II/III and CRD-IV accords treat the default probabilities of small firms as less sensitive to macroeconomic cycles than those of large firms, and therefore permits lower capital charges on bank loans to smaller firms. We test this assumption in a default intensity regression framework using a large sample of loans to private Danish firms. We find that controlling only for firm-size, the default probabilities of small firms are less cyclical than the default probabilities of large firms. However, accounting for firm characteristics other than size, we find that small firms default probabilities are equally cyclical or even more cyclical than the default probabilities of large firms. These results hold using both a multiplicative Cox model and an additive Aalen model. Our findings indicate that reducing capital charges on loans to smaller firms cannot be justified from a risk perspective.

CO518 Room MAL B36 FINANCIAL VOLATILITY AND COVARIANCE MODELLING Chair: Genaro Sucarrat

CO0784: **Generalized dynamic conditional score model**
Presenter: **Bilel Sanhaji**, Aix Marseille University, France
Co-authors: Sebastien Laurent
We propose a likelihood ratio test for the null hypothesis of dynamic conditional score model with Skewed-Student distribution of the innovations. The alternative hypothesis considers different degree of freedom and skewness parameter in the score. We study the size and power properties in finite samples through Monte Carlo experiments. Finally, we show some empirical illustrations using financial data.

CO1005: **A multivariate stochastic volatility model applied to a panel of S&P500 stocks in different industries**
Presenter: **Serda Selin Ozturk**, Istanbul Bilgi University, Turkey
Co-authors: Thanasis Stengos
We estimate a multivariate stochastic volatility model for a panel of stock returns for a number of S&P500 firms from different industries. As in the case of a univariate model we use an efficient importance sampling (EIS) method to estimate the likelihood function of the given multivariate system that we analyze. As opposed to univariate methods where each return is estimated separately for each firm, our results are based on joint estimation that can account for potential common error term interactions based on industry characteristics that cannot be detected by univariate methods. Our results reveal that there are important differences in the industry effects, something that suggests that differential gains to portfolio allocations in the different industries that we examine. There are differences due to idiosyncratic factors and the common industry factors that suggest that each industry requires a separate treatment in arriving at portfolio allocations. Therefore investors should treat their portfolio allocation differently based on the industry they are investing in.

CO0622: **Fitting vast dimensional time-varying covariance models**
Presenter: **Cavit Pakel**, Bilkent University, Turkey
Co-authors: Kevin Sheppard, Neil Shephard, Robert Engle
Building models for high dimensional portfolios is important in risk management and asset allocation. We propose a novel and fast way of estimating existing models of time-varying covariances that overcome an undiagnosed incidental parameter problem which has troubled existing methods when applied to hundreds or even thousands of assets. Indeed we can handle the case where the cross-sectional dimension is larger than the time series one. The theory of this new strategy is developed in some detail, allowing formal hypothesis testing to be carried out on these models. Simulations are used to explore the performance of this inference strategy while empirical examples are reported which show the strength of this method. The out of sample hedging performance of various models estimated using this method are compared.

CO0759: **Equation-by-equation estimation of a multivariate log-GARCH-X model of financial returns**
Presenter: **Genaro Sucarrat**, BI Norwegian Business School, Norway
Co-authors: Christian Francq
Estimation of large financial volatility models are plagued by the curse of dimensionality: As the dimension grows, reliable joint estimation of all the parameters becomes infeasible in practice. This problem is compounded if covariates or conditioning variables X are added to each volatility specification, and the problem is particularly acute for non-exponential volatility models (e.g. GARCH models) since there the variables and parameters are restricted to be positive. We propose an estimator for a log-GARCH-X model that avoids all these problems. The model allows for feedback among the equations, admits stationary regressors as conditioning variables in the X-part, including leverage terms, and allows for time-varying covariances. We prove the consistency and asymptotic normality of an equation-by-equation least squares estimator. The results can be used to undertake inference both within and across equations, and the estimator does not rely on financial returns being distributed according to a specific conditional distribution. We illustrate the usefulness of the methods in two empirical applications. In the first one we illustrate how inference about volatility spillover among stock markets can be undertaken both within each volatility equation, and jointly across the equations. In the second application we illustrate how, in less than 1 minute, our estimator provides consistent and numerically reliable estimates of a 50-dimensional volatility model.

CO370 Room MAL 421 MEASURING FINANCIAL RISK Chair: Scott Brave

CO1273: **Taxes, not credit-event premia, explain short maturity investment-grade spreads**
Presenter: **Luca Benzoni**, Federal Reserve Bank of Chicago, United States
Co-authors: Robert Goldstein
We provide theoretical arguments and empirical support for short-maturity, investment-grade (IG) corporate bond spreads possessing a significant tax component, but not a significant credit-event risk component. First, we extend a previous work by permitting management to issue both debt and equity. Rather than defaulting, managers of IG firms receiving bad private signals maximize shareholder value by servicing existing debt via debt issuance. Consistent with empirical observation, this strategy permits low-spread firms to avoid jumping to default. Second, we provide empirical support for significant state and federal tax components to spreads, and reconcile these findings with the extant literature.

CO1284: **Financial stability and optimal interest-rate policy**
Presenter: **Andrea Ajello**, Board of Governors of the Federal Reserve System, United States
We study optimal interest-rate policy in a new Keynesian model in which the economy is at risk of experiencing a financial crisis and the probability of a crisis depends on credit conditions. The optimal adjustment to interest rates in response to credit conditions is (very) small when the model is calibrated to match an estimated historical relationship between credit conditions, output, inflation and the likelihood of financial crises. Given the imprecise estimates of a number of key parameters, we also study optimal policy taking parameter uncertainty into account. We find that both Bayesian and robust central banks will respond more aggressively to financial stability risks when the probability and severity of financial crises are uncertain.

CO1429: **Impact of the change of survival function on CVA**
Presenter: **Jakub Cerny**, Charles University in Prague, Czech Republic
A new banking regulation, Basel III, introduces the concept of the credit valuation adjustment (CVA). The role of the CVA is to include the counterparty credit risk into the market price of OTC derivatives. Basel III, as usual, comes with a standardized CVA formula using piecewise constant default intensity. Many papers consider stochastic default intensity which is a correct approach, but it forces us to do a lot of time-consuming simulations. A compromise is the assumption of time-varying default intensities. In particular, we investigate an interest rate swap CVA formula based on the assumption of Gaussian copula with constant correlation between the exposure and the default time (wrong-way risk) using different parametric models for the survival function calculation.

CC1334: **A note on the Vasicek's model with the logistic distribution**
Presenter: **Jiri Witzany**, University of Economics in Prague, Czech Republic
We argue that it would be natural to replace the standard normal distribution function by the logistic function in the regulatory Basel II (Vasicek's) formula. Such a model would be in fact consistent with the standard logistic regression PD modeling approach. An empirical study based on US commercial banks loan historical delinquency rates from the period 1985-2012 re-estimates the default correlations and unexpected losses for the normal and logistic distribution models. The results indicate that the capital requirements could be up to 100% higher if the normal Vasicek's model was replaced by the logistic one.

CO352 Room MAL 414 NONLINEAR MODELLING OF FINANCIAL TIME SERIES Chair: Wei Wei

CO0478: **A stochastic price duration model for estimating high-frequency volatility**
Presenter: **Wei Wei**, Aarhus University, Denmark
Co-authors: Denis Pelletier
A class of stochastic price duration models is proposed to estimate high-frequency volatility. A price duration measures how long it takes for the asset price to change by a given amount. It is directly linked to volatility from the passage time theory for Brownian motions. Modeling with price durations renders more effecient sampling scheme compared to return-based estimators. Also, our parametric approach allows us to estimate intraday spot volatility and incorporate additional information such as trade durations.

CO0639: **Modelling sovereign debt contagion: A smooth transition approach**
Presenter: **Susana Martins**, University of Minho, Portugal
Co-authors: Cristina Amado
The unprecedented recent sovereign debt crisis has shown persistent negative effects reaffirming financial contagion as a mechanism that makes instability and systemic risk so widespread. Since the spread of shocks is not immediate, occurring only when the markets are already stressed, a decrease in cross-market dependence is expected before an increase beyond its initial level. The extent and transmission mechanisms of sovereign debt contagion are investigated using daily returns on 10-year government bonds for nine Eurozone countries. Sovereign debt contagion is measured and detected by significant changes in the cross-market correlations that are modelled assuming a nonlinear Smooth Transition Conditional Correlation (STCC-) GARCH process. Over time, the cross-market correlations change smoothly between two states according to an observable transition variable. The time-varying and nonlinear correlation structures are purged from volatility movements and control for country or market-specific effects.

CO1449: **Discrete time nonlinear diffusion models**
Presenter: **Rickard Sandberg**, Stockholm School of Economics, Sweden
A class of Smooth Transition (STR) discrete time diffusion process is considered where the drift component is allowed not only to have a regime-switching type of behavior but also a time-varying behavior. Such features can simply be motivated by the fact that the drift of the return series, say, depends on the phases of the business cycle (the regime-switching behavior) as well as the business cycle may be subject to structural changes over time (the time-varying behavior). Simple tests to identify such features of the data are also derived. Maximum Likelihood (ML) techniques to estimate the proposed discrete time STR diusion models are discussed. In an application to S&P500 and NYSE data, we find, using our tests, overwhelming evidence in favour of that a drift component is subject to both types of nonlinearities. For an example return series, the ML techniques are demonstrated.

CO0303: **Estimation of STAR-GARCH models with iteratively weighted least squares**
Presenter: **Murat Midilic**, Ghent University, Belgium
The Iteratively Weighted Least Squares (IWLS) algorithm is applied to a STAR model with conditional variance. Monte Carlo simulations are carried out to measure the performance of the algorithm. The performance of the algorithm is compared with the performances of fmincon function of MATLAB, maxLik function of R, and an estimation method that separately estimates the mean and the variance. Simulation results show that the IWLS algorithm improves the estimation of the model parameters when the real value of the slope parameter is low. In the empirical part, STAR-GARCH model is used to estimate daily crude oil prices and volatilities.

CG469 Room MAL 541 CONTRIBUTIONS IN DSGE MODELLING Chair: Luca Fanelli

CC0300: **Exchange rates and fundamentals: Closing a two-country model**
Presenter: **Takashi Kano**, Hitotsubashi University, Japan
In an influential paper claims that the near random-walk behavior of nominal exchange rates is an equilibrium outcome of a present-value model of a partial equilibrium asset approach when economic fundamentals follow exogenous first-order integrated processes and the discount factor approaches one. Subsequent empirical studies further confirm this proposition by estimating a discount factor that is close to one under distinct identification schemes. We argue that the unit market discount factor implies the counterfactual joint equilibrium dynamics of random-walk exchange rates and economic fundamentals within a canonical, two-country, incomplete market model. In particular, at the limit of the unit discount factor, the model predicts that a dominant driver of random-walk exchange rates is a permanent shock to the two-country money supply differential, the two-country consumption differential is perfectly correlated with the real exchange rate, and estimated money monetary disturbances are counterfactually volatile. Bayesian posterior simulation exercises based on post-Bretton Woods data from Canada and the United States reveal difficulties in reconciling the equilibrium random-walk proposition within the two-country model; in particular, the market discount factor is identified as being much lower than one.

CC1338: **Fundamental shock selection in DSGE models**
Presenter: **Stefano Grassi**, University of Kent, United Kingdom
Co-authors: Filippo Ferroni, Miguel Leon Ledesma
DSGE models are typically estimated assuming the existence of certain structural shocks that drive macroeconomic fluctuations. We analyze the consequences of introducing non-fundamental shocks for the estimation of DSGE model parameters and propose a method to select the structural shocks driving uncertainty. We show that forcing the existence of non-fundamental structural shocks produces a downward bias in the estimated internal persistence of the model. We then show how these distortions can be reduced by allowing the covariance matrix of the structural shocks to be rank deficient using priors for standard deviations whose support includes zero. The method allows us to accurately select fundamental shocks and estimate model parameters with precision. Finally, we revisit the empirical evidence on an industry standard medium-scale DSGE model model and find that government, price, and wage markup shocks are non-fundamental.

CC1661: **Misspecification and expectations correction in new Keynesian DSGE models**
Presenter: **Giovanni Angelini**, University of Bologna, Italy
Co-authors: Luca Fanelli
The focus is on the dynamic misspecification that characterizes the class of small scale New-Keynesian models currently used in monetary and business cycle analysis, and provides a remedy for the typical difficulties these models have in accounting for the rich contemporaneous and dynamic correlation structure of the data. We suggest using a statistical model for the data as a device through which it is possible to adapt the econometric specification of the New-Keynesian model such that the risk of omitting important propagation mechanisms is kept under control. A pseudo-structural form is built from the baseline system of Euler equations by forcing the state vector of the system to have the same dimension as the state vector characterizing the statistical model. The pseudo-structural form gives rise to a set of cross-equation restrictions that do not penalize the autocorrelation structure and persistence of the data. Standard estimation and evaluation methods can be used. We provide an empirical illustration based on U.S. quarterly data and a small-scale monetary New Keynesian model.

CC1793: **Learning expectations using multi-period forecasts**
Presenter: **Demetris Koursaros**, Cyprus University of Technology, Cyprus
The macroeconomic implications of introducing perpetual learning to a simple search and matching model are investigated. The model provides amplification when agents are assumed to form multiperiod forecasts for the costs and benefits of future job positions. Forecasts of future cash flows are subject to periods of either excess optimism or excess pessimism which boosts volatility in the model. The model can replicate moments from quarterly US data from 1955Q1 to 2010Q4 and thus account for the amplification puzzle. Alternatively, if it is assumed that agents merely need to provide forecasts up to a period ahead, the model predictions are very close to the rational expectations case. The rational for using multiperiod forecasts besides fitting the model better, it appears to be a more realistic description of how agents form beliefs. We examine the patterns created by forecasting errors from the survey of professional forecasters from 1968Q1 to 2015Q2 and seem to be significantly autocorrelated. The assumption that agents form multiperiod forecasts is the only specification for expectations that delivers autocorrelated forecasting errors in the model. Rational expectations and a period in advance forecasting models imply a near zero autocorrelation for simulated forecasting errors.

CG499 Room MAL 540 CONTRIBUTIONS ON FINANCIAL TIME SERIES AND RISK PREMIA Chair: Philip Hans Franses

CC1319: **UK term structure decompositions at the zero lower bound**
Presenter: **Sarah Mouabbi**, Banque de France, France
Co-authors: Andrea Carriero, Elisabetta Vangelista
A Zero Lower Bound (ZLB) consistent shadow-rate model is employed to decompose UK nominal yields into expectation and term premia components. Compared to a standard affine term structure model, it performs relatively better in a ZLB setting and effectively captures the countercyclical nature of term premia. The ZLB model is then exploited to estimate inflation expectations and risk premia. This entails jointly pricing and decomposing nominal and real UK yields. We find that the asset purchases programme, launched in March 2009, was successful in re-anchoring inflation expectations in the aftermath of Lehman Brothers bankruptcy.

CC1400: **Size and power of tests based on permanent-transitory component models**
Presenter: **Fabrizio Casalin**, Newcastle University, United Kingdom
Recently, the literature has proposed a new type of tests for the Efficient Market Hypothesis based on Permanent-Transitory Component Models. We compare the power of these statistics with conventional tests based on linear regressions. Simulation results suggest that the former dominate the latter for a wide range of data generating processes. We propose an application to spot and forward interest rates. Empirical results show that the two types of tests can yield conflicting results which can be explained by the size distortions and reduced power which affect the statistics based on linear regressions.

CC1626: **On the tail risk premium in the crude oil futures market**
Presenter: **Reinhard Ellwanger**, Bank of Canada, Canada
Oil prices are notoriously difficult to forecast and exhibit wild swings or excess volatility that are difficult to rationalize by changes in fundamentals alone. An explanation is offered for these phenomena based on time varying disaster probabilities and disaster fears. Using information from crude oil options and futures we document economically large jump tail premia in the crude oil futures market: a substantial fraction of the average futures and variance risk premium is accounted for by the compensation for rare events. These tail risk premia vary substantially over time and significantly forecast crude oil futures and spot returns. The results suggest that oil futures prices overshoot (undershoot) in the presence of upside (downside) tail fears in order to allow for smaller (larger) risk premia thereafter. We show that this overshooting (undershooting) is amplified for

the spot price because of time varying benefits from holding inventory that work in the same direction. The novel oil price uncertainty measures yield additional insights into the relationship between the oil market and macroeconomic outcomes.

CC1655: **Forecasting the equity risk remium: The ups and the downs**
Presenter: **Nick Baltas**, Imperial College Business School - QMUL, United Kingdom
Co-authors: Dimitrios Karyampas
Recent academic advances in predicting the market equity risk premium (ERP) employ constraints motivated by economic theory in order to generate non-negative ERP forecasts. We show that the improved accuracy of these constrained models is primarily due to the fact that the equity markets have exhibited positive average excess returns across the business cycle. However, we find that in recessionary periods, in periods of negative ERP realisations and in periods of increased volatility, they perform poorly and underperform the basic unconstrained linear forecasting model. This finding constitutes a significant challenge for the practical benefit of constrained forecasting models when used by a mean-variance investor who allocates dynamically between the equity market and a risk-free asset. We verify the validity of our arguments using a long sample period for the US market, from 1927 to 2013.

CG551 Room MAL 539 CONTRIBUTIONS ON PORTFOLIO SELECTION Chair: Chulwoo Han

CC1697: **Dynamic shrinkage estimates for portfolio selection**
Presenter: **Francesco Lautizi**, University of Rome Tor Vergata, Italy
Dynamic shrinkage estimators are proposed for both the covariance matrix and its inverse, the precision matrix. The idea is to obtain a series of estimates that do not vary too much across time, designed to provide a smooth sequence of portfolio weights across time when used as input in the global minimum variance portfolio policy. The proposed estimation technique for the covariance (precision) matrix is an iterative procedure, designed for empirical applications based on rolling estimation window. Given an estimation date t, this method shrinks the sample covariance (precision) matrix, computed in t, towards two targets: the covariance (precision) matrix implied by the CAPM computed in the same date and the the estimate of the covariance (precision) matrix computed in $t-1$. Two cross-validation algorithms are proposed in order to compute the shrinkage parameters of interest. These estimators produce global minimum variance portfolios that are more stable across time, considerably reducing the average turnover of portfolio weights; this translates into a reduction of transaction costs that, combined with good performance both in the standard deviation and the mean of out-of-sample portfolio returns, leads to higher Sharpe ratios compared to standard shrinkage estimators.

CC1703: **Approximating expected shortfall for heavy tailed distributions**
Presenter: **Simon Broda**, University of Amsterdam, Netherlands
Co-authors: Marc Paolella, Jochen Krause
The Basle Committee's proposed move from Value at Risk to expected shortfall as the mandated risk measure in its market risk framework necessitates practical methods for evaluating said measure. Defined as a partial expectation of the return distribution standardized by the tail probability, expected shortfall cannot be obtained in explicit form for many distributions of interest. A saddlepoint approximation is derived for the expected shortfall associated with certain random variables, without assuming the existence of a moment generating function. This includes, in particular, certain heavy tailed distributions, to which earlier results were not applicable. The new approximation can be evaluated quickly and reliably, and provides excellent accuracy. We also establish a link between our proposed approximation and mean-expected shortfall portfolio optimization. Numerical examples and an empirical application are presented.

CC1539: **Stochastic spanning**
Presenter: **Stelios Arvanitis**, RC-AUEB, Greece
Co-authors: Thierry Post, Mark Hallam
A theory and method are developed and implemented for analyzing whether introducing new securities or relaxing investment constraints improves the investment opportunity set for risk averse investors. We develop a test procedure for stochastic spanning for two nested polyhedral portfolio sets based on subsampling and Linear Programming. The procedure is statistically consistent and asymptotically exact for a class of weakly dependent processes. Using the stochastic spanning tests, we accept market portfolio efficiency but reject two-fund separation in standard data sets of historical stock market returns. The divergence between the results of the two tests illustrates the role for higher-order moment risk in portfolio choice and challenges representative- investor models of capital market equilibrium.

CC1094: **Time varying integration and diversification strategies: The case of emerging and frontier stock markets**
Presenter: **Francesco Guidi**, University of Greenwich, United Kingdom
The aim is to investigate to what extent frontier as well as emerging equity markets are integrated with a leading stock market such the USA. By using both static and dynamic cointegration models we found that frontier equity markets are weakly integrated with the US market. In order to investigate the speed at which this integration is taking place, we use smooth transition models. Our findings show that the speed of integration in very low in most of the frontier markets. As the latter are located in economies characterized by very high rate of economic growth, the lack or weak integration with leading markets might be a valid reason for US investors willing to diversify their portfolio across these markets. Thereby, we explored such as hypothesis by using popular diversification strategies. Our results show that US investors would benefits in terms of portfolio gains if they had invested in frontier equity markets.

EO214 Room Gordon WAVELET METHODS IN STATISTICS Chair: Jean-Michel Poggi

EO0194: **Poisson functional regression for the analysis of Next Generation Sequencing data**
Presenter: **Franck Picard**, CNRS Lyon, France
Next Generation Sequencing experiments have now become standard for the analysis of genome-wide molecular phenomenon. The characteristic of produced data is their volume, and when mapped on a reference genome, they are spatially organized along chromosomes. Another particularity of these data is their discrete nature as they are made of counts usually modeled by Poisson of Negative Binomial distributions. To account of the spatial structure, Hidden Markov models have been considered to detect reads enrichments, but the flexibility of such approaches is limited, as well as the performance to detect sharp enrichments in the data. We develop a Poisson functional regression framework based on wavelets to analyze mapped NGS data. This framework is very flexible as it allows a multiscale representation of the signal, and allows for the introduction of replicates and covariates for normalization purposes. The statistical challenge lies in the selection of wavelet coefficients which depends on penalty weights that need to be calibrated. We provide data-driven weights for the Lasso and the group-Lasso derived from concentration inequalities adapted to the Poisson case. We show that the associated Lasso and group-Lasso procedures are theoretically optimal in the oracle approach. Our method is then illustrated on the identification of replication origins in the human genome.

EO0312: **Wavelet estimation and variable selection for additive partial linear models**
Presenter: **Anestis Antoniadis**, University Joseph Fourier, France
Co-authors: Umberto Amato, Italia De Feis
Additive partial linear models with nonparametric additive components of heterogeneous smoothness are studied. To achieve optimal rates in large sample situations we use block wavelet penalisation techniques combined with adaptive (group) LASSO procedures for selecting the variables in the linear part and the the additive components in the nonparametric part of the models. Numerical implementations of our procedures for proximal like algorithms are discussed. Large sample properties of the estimates and of the model selection are presented and the results are illustrated with simulated examples and a real data analysis.

EO0390: **Robust and efficient regression**
Presenter: **Guy Nason**, University of Bristol, United Kingdom
It is known that L1/median/quantile based regressions perform better than L2/mean-based regressions in the presence of outliers, but that the latter is more statistically efficient for light tailed data. We develop a new method that is almost as efficient as L2 methods on light tailed data and almost as robust as L1 methods in the presence of outliers and adapts automatically to the different regimes. The new method relies on a new multiscale measure of location which is fast to compute and has an interesting non-multiscale interpretation in terms of standard statistics. We exhibit the new estimator on a range of regression problems.

EO1200: **Locally stationary approaches to complex time series**
Presenter: **Idris Eckley**, Lancaster University, United Kingdom
Many time series, whilst not stationary, can often be successfully modelled as a locally stationary series. There has been a concerted effort to develop wavelet-based methods for analysing such time series over the last 15 years. We will consider the case of locally stationary series with correlated increments, describing what might be done to understand the time-varying second order structure within such time series.

EO072 Room SH349 THEORETICAL FOUNDATION OF BIG DATA Chair: Hao Chen

EO0206: **Fusion learning by individual to clique (FLIC): Enhancing inference by adaptive combination of confidence distributions**
Presenter: **Min-ge Xie**, Rutgers University, United States
Learning from multiple studies can often be fused together to yield a more effective overall inference than individual studies alone. Such effective fusion learning is of vital importance, especially in light of the trove of data nowadays collected routinely from various sources in all domains and at all time. A new approach is presented, named Fusion Learning by Individual-to-Clique (FLIC), to enhancing inference of an individual study through adaptive combination of confidence distributions obtained from its clique (namely peers of similar studies). Roughly speaking, FLIC begins with obtaining inference for each individual study (or a subset of data), then adaptively forming a clique, and finally obtaining a combined inference from the clique. FLIC can be performed without accessing the entire dataset and thus allow the so-called split-and-conquer approach to be implemented on individual studies and reduce substantially computational expense. Drawing inference from the clique allows borrowing strength from similar studies to enhance the inference efficiency for individual studies. We also provide supporting theories for FLIC and its applications in personalized medicine and financial profiling of companies.

EO0980: **Sequential change-point detection based on nearest neighbors**
Presenter: **Hao Chen**, University of California at Davis, United States
We propose a novel approach for the detection of change-points as data are generated. The approach utilizes nearest neighbor information and can be applied to high-dimensional data sequences and object data sequences. An analytic expression is obtained for the average run length when there is no change, facilitating the application of the approach to real problems. Simulations reveal that the proposed approach has shorter expected detection delay than existing approaches when the dimension of the data is moderate to high.

EO1146: **Inference for misspecified high dimensional Cox regression**
Presenter: **Shengchun Kong**, Purdue University, United States
Co-authors: Guang Cheng, Xianyang Zhang
We consider high-dimensional inference under a potentially misspecified Cox proportional hazard model. We propose a de-sparsified Lasso estimator for the regression parameters based on the partial likelihood function, and show that it converges to a well-defined constant vector under some conditions. In addition, each component of the de-sparsified Lasso estimator is asymptotically normal with a variance that can be consistently estimated. We provide some correct interpretations of the asymptotic limit constant vector, and valid statistical inferences of the corresponding covariate effect can be derived based on the asymptotic distribution for the de-sparsified Lasso estimator given the asymptotic limit can be interpreted meaningfully.

EO0824: **High dimensional factor model: The blessing of dimensionality**
Presenter: **Quefeng Li**, University of North Carolina - Chapel Hill, United States
Co-authors: Jianqing Fan, Guang Cheng, Yuyan Wang, Quefeng Li
Factor model is an essential tool for exploring the covariance structure of subjects driven by latent variables. Much progress has been made for the estimation of the covariance matrix from a high-dimensional factor model. However, practitioners may only care about a small subset of those subjects. If our ultimate goal is to estimate a low-dimensional covariance matrix related to the few subjects-of-interest, shall we employ additional subjects in our covariance matrix estimation procedure? We provide sufficient conditions for an affirmative answer, and further quantify its gain in terms of Fisher information and convergence rate. In fact, we can even obtain an oracle-like result (as if all the factors were known) when the number of total subjects is large enough. A divide-and-conquer algorithm is proposed to largely alleviate the computational burden in processing a large amount of data, but shown not to sacrifice any statistical inferential accuracy in comparison with the pooled analysis. Simulation studies further advocate the utilization of more data and demonstrate the effectiveness of the divide-and-conquer algorithm. Finally, our proposal is applied to a microarray data example to show the blessing of utilizing more data.

EO152 Room Chancellor's Hall ESTIMATION AND INFERENCE IN HIGH DIMENSIONAL AND MIXTURE COPULAS Chair: Robert Kohn

EO0218: **Generalized information matrix tests for copulas**
Presenter: **Ulf Schepsmeier**, Technische Universitaet Muenchen, Germany
Co-authors: Artem Prokhorov, Yajing Zhu
A family of goodness-of-fit tests for copulas is proposed. The tests use generalizations of a previous information matrix (IM) equality and so relate to previous copula tests. The idea is that eigen spectrum-based statements of the IM equality reduce the degrees of freedom of the test statistics asymptotic distribution and lead to better size-power properties, even in high dimensions. The gains are especially pronounced for vine copulas, where additional benefits come from simplifications of score functions and the Hessian. We derive the asymptotic distribution of the generalized tests, accounting for the non-parametric estimation of the marginals and apply a parametric bootstrap procedure, valid when asymptotic critical values are inaccurate. In Monte Carlo simulations, we study the behavior of the new tests, compare them with several Cramer-von Mises type tests favored in a previous simulation study and confirm the desired properties of the new tests in high dimensions.

EO0240: **Estimating high dimensional skew-t copulas using Langevin and Hamiltonian based Sequential MCMC**
Presenter: **Gareth Peters**, University College London, United Kingdom
Co-authors: Francois Septier
Nonlinear non-Gaussian state-space models arise in numerous applications in statistics and signal processing. In this context, one of the most successful and popular approximation techniques is the Sequential Monte Carlo (SMC) algorithm, also known as particle filtering. Nevertheless, this method tends to be inefficient when applied to high dimensional problems. We focus on another class of sequential inference methods, namely the Sequential Markov Chain Monte Carlo (SMCMC) techniques, which represent a promising alternative to SMC methods. After providing a unifying framework for the class of SMCMC approaches, we propose novel efficient strategies based on the principle of Langevin diffusion and Hamiltonian dynamics in order to cope with the increasing number of high-dimensional applications. Simulation results show that the proposed algorithms achieve significantly better performance compared to existing algorithms.

EO0427: **On copulas and finite mixtures**
Presenter: **Mohamad Khaled**, University of Queensland, Australia
Co-authors: Robert Kohn
Finite mixtures have been used as a tool for constructing flexible families of copulas. Yet the approximation properties of finite mixtures of copulas are not well known. We explore some of those properties. In particular, three different types of statistical models are studied: finite mixture models where the component distributions are given by Archimidean copulas, finite mixtures of elliptical copulas and copulas of finite mixtures of multivariate normal distributions. Identification and statistical inferential procedures are explored in each case.

EO0532: **Fast estimation for copula models by Variational Bayes**
Presenter: **Minh-Ngoc Tran**, University of Sydney, Australia
Co-authors: Robert Kohn
A fast Variational Bayes approach for estimating copula models is discussed. The method is generic in the sense that it can be applied to almost all copula models without requiring a model-based derivation. We illustrate the methodology with simulated and real data.

EO615 Room CLO 102 ADVANCED COMPUTATIONAL METHODS FOR COMPLEX DATA Chair: Eva Petkova

EO0283: **Identifying biosignatures for placebo response using high dimensional functional data**
Presenter: **Thaddeus Tarpey**, Wright State University, United States
Co-authors: Eva Petkova, Todd Ogden
Placebo response rates often tend to be quite high when treating many mental illnesses, such as depression. Modeling placebo response in the presence of active treatments continues to present a major statistical challenge. Quantifying the efficacy of active drug treatments is hampered by the presence of nonspecific placebo effects. Consequently, the problem of finding biomarkers for placebo response as well as specific drug response is of critical importance for advancing precision medicine in mental health treatment. The availability of modern high dimensional baseline modalities such as brain imaging data has reinvigorated the goal of finding these potential biomarkers. We explore the use of high dimensional functional baseline functional covariates to help distinguish specific drug response from placebo response using data from a depression trial.

EO0342: **A functional data approach for using EEG Data to select treatment for major depressive disorder**
Presenter: **Adam Ciarleglio**, New York University School of Medicine, United States
Co-authors: Eva Petkova, Thaddeus Tarpey, Todd Ogden
Major depressive disorder (MDD) is a disease characterized by substantial heterogeneity. This heterogeneity exists in both the symptoms associated with MDD and in the responses that MDD patients have to various forms of treatment: what works for one patient may be ineffective or harmful for another. This makes treatment selection a difficult task particularly because there are no widely accepted biomarkers for MDD treatment response. Recently, the search for such biomarkers has broadened to include measures derived from neuroimaging modalities such as MRI, fMRI, and EEG that are collected at baseline. This seems justified since various aspects of brain structure and function have been implicated in depressive symptoms and in response to treatment. We propose a functional data approach for using baseline EEG data to both select the best MDD treatment for the individual and provide interpretable measures of the relationship between the EEG signal and treatment response. The approach will be evaluated in several realistic settings using synthetic data and will be applied to real data arising from a multi-center clinical trial comparing two treatments for MDD in which subjects had their EEG data collected at baseline.

EO1801: **Nonparametric generated effect modifiers for placebo response**
Presenter: **Eva Petkova**, New York University, United States
The problem of identifying baseline covariates that predict placebo response and distinguish placebo response from specific drug response is an on-going problem in clinical research that has evaded satisfactory solutions. Much of the difficulty is due to the fact that typical baseline predictors individually have very little modifying effects in differentiating placebo response from specific drug response. The aim is to propose a novel approach to this problem whereby several baseline predictors are combined to form a single powerful moderator a generated effect modifier (GEM). Previous attempts have considered GEMs in linear regression models. In order to add flexibility and increase the power of a modifying effect, we investigate GEM models with quadratic effects as well as nonparametric regression models where the outcome is modeled as a function of treatment and a smooth nonparametric function of the GEM predictor. Results will be illustrated via simulation studies and an application to a placebo controlled depression trial.

EO0529: **Regression analysis of complex outcome data from two-stage randomized studies based on cumulative incidence function**
Presenter: **Yu Cheng**, University of Pittsburgh, United States
Co-authors: Idil Yavuz, Abdus Wahed
The focus is on modeling covariate effects on disease progression that is subject to competing risk censoring by death. The complication of the data is that subjects were further randomized to different treatment options per their response status to the initial treatment to which they were assigned in the first randomization. The objective is to model covariate effects on the cumulative incidence of disease progression over time for different dynamic treatment strategies. Some of the popular existing regression models, such as the multi-state model, the Fine and Gray, and the Scheike et al. regression models, are extended to the two-stage randomization setting. Two regression strategies are proposed based on inverse probability weighting and pattern mixture estimation, and ways are provided to implement the proposed models in R using the existing packages. The improvement of the proposed methods is illustrated through numerical studies.

EO238 Room CLO B01 FUNCTIONAL AND OBJECT DATA WITH COMPLEX DEPENDENCIES Chair: Laura Sangalli

EO0302: **Spatial variable selection via soft-thresholded Gaussian processes**
Presenter: **Brian Reich**, North Carolina State University, United States
Co-authors: Kang Jian, Ana-Maria Staicu
A novel Bayesian framework is proposed to do scalar-on-image regression by modeling the regression coefficient through a thresholded latent Gaussian processes. The thresholding of the process will ensure that the result is sparse while its smoothness property will guarantee the smoothness of the effect. Our methods are computationally tractable, accommodate additional covariates, and can be extended to generalized responses. We illustrate the approach through simulations and apply it to the data from an electroencephalography (EEG) study of alcoholism, where we study the relation between the alcoholism status and the electrical brain activity over time.

EO0763: **Analysis of big medical motion data using functional LARS**
Presenter: **Jian Qing Shi**, Newcastle University, United Kingdom
Co-authors: Yafeng Cheng
How to extend the idea of LARS to select functional variables in a scalar-on-function regression model is discussed. Two new stopping rules are proposed. The new algorithm works very efficiently even for the case with very large number of functional variables. We discuss the idea, method, implementation and the result of a comprehensive simulation study. The method has been applied to evaluate the function of upper limbs for patients after stroke by using motion data. The data include scalar response, more than 200 scalar covariates and 500 functional covariates. Different models are compared.

EO1300: **Additive mixed models for generalized functional data**
Presenter: **Fabian Scheipl**, Ludwig-Maximilians-Universitaet Muenchen, Germany
Co-authors: Jan Gertheiss, Sonja Greven
We propose a comprehensive framework for additive regression models for non-Gaussian functional responses, allowing for multiple (partially) nested or crossed functional random effects with flexible correlation structures for, e.g., spatial, temporal, or longitudinal functional data as well as linear and nonlinear effects of functional and scalar covariates that may vary smoothly over the index of the functional response. Our implementation handles functional responses from any exponential family distribution as well as many others like Beta- or scaled non-central t-distributions. Development is motivated by and evaluated on an application to large-scale longitudinal feeding records of pigs. Results in extensive simulation studies as well as replications of two previously published simulation studies for generalized functional mixed models demonstrate the good performance of our proposal. The approach is implemented in well-documented open source software in the `pffr` function in R-package refund.

EO1787: **On averaging a set of biomedical curves**
Presenter: **Sofiane Boudaoud**, Sorbonne University-University of Technology of Compiegne, France
In biomedical applications, averaging is a current operation applied on a set of curves or time series. It allows the access to a prototype function that can be considered as centered on the present variability dispersion and should guaranty representativeness for shape analysis or classification purposes. Classical averaging treats solely with amplitude variability and is biased by other kinds of variability (phase and shape) that are often present in biomedical curves. However, recent averaging techniques, bearing on functional formalism, emerged to include these kinds of variability and provide additional information on the complex studied living system. Some of them suppose the presence of a common structure inside the studied set and others work in a more freely manner with heterogeneous shapes. Several averaging approaches will be presented, by firstly describing mathematical background, and illustrated on simulated and real data to highlight on their usefulness and limitations in studying complex biomedical curves.

EO290 Room CLO 203 ADVANCED METHODS FOR NEUROIMAGING DATA ANALYSIS Chair: Hongtu Zhu

EO0314: **Spatial Bayesian latent factor regression modeling of multiple sclerosis lesion data**
Presenter: **Timothy Johnson**, University of Michigan, United States
Co-authors: Silvia Montagna, Thomas Nichols
Multiple Sclerosis (MS) is an autoimmune disease affecting the central nervous system. The immune system attacks and destroys the myelin sheath – a fatty substance that surrounds nerve fibers and acts as an insulator for neurotransmission. The resulting lesions are visible on structural magnetic resonance images. One current research question is whether the spatial pattern of lesions can distinguish and predict which subtype of MS a patient has. To address this question we propose a Bayesian hierarchical model for neuroimaging point pattern data. We model the lesion locations from each patient as a spatial Cox process where each subject specific log intensity function is modeled as a linear combination of a high-dimensional basis set. This functional representation leads to computational speed-ups over traditional log-Gaussian Cox processes, as a sparse representation of the intensities is guaranteed through sparse latent factor modeling of the basis coefficients. Furthermore, the latent factors are used to link the random intensities to MS subtype, as part of a scalar-on-image regression. We apply our methodology to a cohort of MS patients.

EO0417: **Evaluating dynamic bivariate correlations in resting-state fMRI**
Presenter: **Martin Lindquist**, Johns Hopkins University, United States
To date, most functional Magnetic Resonance Imaging (fMRI) studies have assumed that the functional connectivity (FC) between time series from distinct brain regions is constant across time. However, recently, there has been increased interest in quantifying possible dynamic changes in FC during fMRI experiments, as it is thought this may provide insight into the fundamental workings of brain networks. We focus on the specific problem of estimating the dynamic behavior of pair-wise correlations between time courses extracted from two different regions of the brain. We critique the commonly used sliding-windows technique, and discuss some alternative methods used to model volatility in the finance literature that could also prove useful in the neuroimaging setting. In particular, we focus on the Dynamic Conditional Correlation (DCC) model, which provides a model-based approach towards estimating dynamic correlations. We investigate the properties of several techniques in a series of simulation studies and find that DCC achieves the best overall balance between sensitivity and specificity in detecting dynamic changes in correlations. We

also investigate its scalability beyond the bivariate case to demonstrate its utility for studying dynamic correlations between more than two brain regions. Finally, we illustrate its performance in an application to test-retest resting state fMRI data.

EO0640: **SILFM: Single index latent factor model based on high-dimensional features**
Presenter: **Hongtu Zhu**, University of North Carolina at Chapel Hill, United States
The aim is to develop a single-index latent factor modeling (SILFM) framework to build an accurate prediction model for clinical outcomes based on a massive number of features. We develop a three-stage estimation procedure to build the prediction model. SILFM uses an independent screening method to select a set of informative features, which may have a complex nonlinear relationship with outcome variables.

EO0795: **Modelling time-varying brain connectivity using multiregression dynamic models**
Presenter: **Thomas Nichols**, University of Warwick, United Kingdom
We consider the use of Bayesian time series models for modelling and inference on time-varying connectivity in resting-state Functional Magnetic Resonance Imaging (fMRI). In contrast to moving-window approaches, this method poses a single generative model for all nodes, all time points. Known as a "Multiregression Dynamic Model" (MDM), it comprises an extension of a traditional Bayesian Network by posing latent time-varying regression coefficients of a given node on its parent nodes. After estimating a single "discount factor", reflecting the balance of observation and latent variance, the model estimates and posterior have a closed form; thus different proposed network structures can be readily compared with Bayes Factors. While originally developed for directed acyclic graphs, a modified version can also accommodate directed (possibly cyclic) graphs as well. We describe this model and illustrate it on a set of resting- and constant-task fMRI data. Not only it does find biologically plausible networks, we show how its formulation as a set of regressions readily allows diagnostics and possible elaboration of basic models.

EO308 Room Montague TOPICS ON DISTANCE CORRELATION Chair: Konstantinos Fokianos

EO0364: **Distance correlation coefficients for Lancaster distributions**
Presenter: **Dominic Edelmann**, Heidelberg University, Germany
Co-authors: Johannes Dueck, Donald Richards
The problem of calculating distance correlation coefficients between random vectors whose joint distributions belong to the class of Lancaster distributions is considered. We derive under mild convergence conditions a general series representation for the distance covariance for these distributions. To illustrate the general theory, we apply the series representation to derive explicit expressions for the distance covariance and distance correlation coefficients for the bivariate and multivariate normal distributions, and for the bivariate gamma and Poisson distributions which are of Lancaster type. The results are presented as functions of the natural parameters which parametrize these distributions.

EO0152: **Local Gaussian autocorrelation and tests of serial independence**
Presenter: **Virginia Lacal Graziani**, University of Bergen, Norway
The traditional and most used measure for serial dependence in a time series is the autocorrelation function. This measure gives a complete characterization of dependence for a Gaussian time series, but it often fails for nonlinear time series models such as, for instance, the GARCH model, where it is zero for all lags. The autocorrelation function is an example of a global measure of dependence. The purpose is to apply a well-defined local measure of serial dependence, called the local Gaussian autocorrelation, to time series. It generally works well for nonlinear models, and it can distinguish between positive and negative dependence. Asymptotic properties are given and both univariate and bivariate time series are considered. Tests of serial independence based on the local Gaussian correlation are compared to other tests such as the Brownian distance correlation in a number of simulation experiments. The new tests perform well.

EO0362: **Consistent testing for pairwise dependence in time-series**
Presenter: **Konstantinos Fokianos**, University of Cyprus, Cyprus
The problem of testing pairwise dependence for stationary time series is considered. We suggest the use of a Box-Ljung type test statistic which is formed after calculating the distance covariance function among pairs of observations. The distance covariance function is a suitable measure for detecting dependencies among data and it is based on the distance between the characteristic function of the joint distribution of the random variables to the product of the marginals. We show that, under the null hypothesis of independence and under mild regularity conditions, the test statistic converges to a normal random variable. The results are complemented by several examples.

EC1504: **Modelling multivariate serially correlated count data in continuous time**
Presenter: **Almut Veraart**, Imperial College London, United Kingdom
A new continuous-time framework for modelling serially correlated count and integer-valued data is introduced in a multivariate setting. The key modelling component is a multivariate integer-valued trawl process which is obtained by kernel smoothing of an integer-valued Levy basis. Various ways of describing both serial and cross dependence are discussed in detail in such a setting and we describe how the corresponding model parameters can be estimated. Simulation studies reveal a good finite sample performance of the proposed methods and applications to financial econometrics are discussed.

EO074 Room Court ADVANCES IN BAYESIAN NONPARAMETRIC METHODS Chair: Taeryon Choi

EO0405: **Generalised Gaussian process regression model for non-Gaussian functional data**
Presenter: **Bo Wang**, University of Leicester, United Kingdom
Co-authors: Jian Qing Shi
A generalized Gaussian process concurrent regression model for functional data is discussed where the functional response variable has a binomial, Poisson or other non-Gaussian distribution from an exponential family while the covariates are mixed functional and scalar variables. The proposed model offers a nonparametric generalized concurrent regression method for functional data with multidimensional covariates, and provides a natural framework on modeling common mean structure and covariance structure simultaneously for repeatedly observed functional data. The mean structure provides an overall information about the observations, while the covariance structure can be used to catch up the characteristic of each individual batch. The prior specification of covariance kernel enables us to accommodate a wide class of nonlinear models. The definition of the model, the inference and the implementation as well as its asymptotic properties are discussed. Several numerical examples with different types of non-Gaussian response variables are presented.

EO0786: **On the two-parameter Poisson-Dirichlet process truncation error**
Presenter: **Pierpaolo De Blasi**, University of Torino and Collegio Carlo Alberto, Italy
We study the asymptotic distribution of the truncation error in the stick-breaking representation of the two-parameter Poisson-Dirichlet process. First, we consider the limit distribution of a (properly centered) log transform of the truncation error and we show that it is infinitely divisible and can be described in terms of the σ-diversity of the process, thus obtaining back a result due to Jim Pitman. Second, we derive a large deviation principle for the tail behavior of the truncation error.

EO0813: **Nonparametric goodness of fit via cross-validation Bayes factors**
Presenter: **Jeff Hart**, Texas AM University, United States
Co-authors: Taeryon Choi
A nonparametric Bayes procedure is proposed for testing the fit of a parametric model for a distribution. Alternatives to the parametric model are kernel density estimates. Data splitting makes it possible to use kernel estimates for this purpose in a Bayesian setting. A kernel estimate indexed by bandwidth is computed from one part of the data, a training set, and then used as a model for the rest of the data, a validation set. A Bayes factor is calculated from the validation set by comparing the marginal for the kernel model with the marginal for the parametric model of interest. A simulation study is used to investigate how large the training set should be, and examples involving astronomy and wind data are provided. Evidence for Bayes consistency of the proposed test is also given.

EO0992: **Iterative algorithms for constructing priors**
Presenter: **Fumiyasu Komaki**, The University of Tokyo, Japan
Two iterative algorithms to construct prior densities for Bayesian prediction based on parametric models are discussed. First, an algorithm to construct latent information priors is introduced. The latent information prior is defined as a prior maximizing the conditional mutual information between the quantity to be predicted and the unknown parameter given the data. Bayesian predictive densities based on the latent information priors achieve minimaxity under the Kullback-Leibler loss in many examples. The algorithm is a generalization of the Arimoto-Blahut algorithm to obtain channel capacity in information theory. Next, an algorithm to obtain Bayes projection of a non-Bayesian predictive density is introduced. The Bayes projection is defined as a divergence projection of a predictive density to the space of Bayesian predictive densities. The Bayes projection of a predictive density is superior to the original predictive density under the Kullback-Leibler loss. The two introduced algorithms are closely related to each other.

EO170 Room Jessel BAYESIAN METHODS IN OFFICIAL STATISTICS Chair: Brunero Liseo

EO0438: **Unit level small area model with covariates perturbed for disclosure limitation**
Presenter: **Serena Arima**, Sapienza University of Rome, Italy
Co-authors: Silvia Polettini
Small area estimation deals with the problem of estimating area level aggregates, when the sampling design is not guaranteed to produce sufficiently large subsamples for all areas of interest. Direct survey estimates may be unreliable and improved estimates can be obtained using mixed effects regression models that link the small areas and borrow strength from similar domains. We focus on small area model in which covariates are measured with error. We consider the situation in which errors in covariates are artificially introduced by a mechanism of disclosure protection of the respondents. Statistical disclosure control (SDC) is commonly applied to prevent reidentification of respondents. An area model that includes auxiliary variables perturbed by disclosure limitation methods has been proposed recently. However, in a disclosure limitation context unit-level models arise more naturally since the perturbation is usually performed at the unit-level. We extend the aforementioned model as a unit-level model: we investigate the performance of the model in estimating the regression parameters and predicting the small area means. We also study the model capability in predicting the true value of the perturbed variables for each unit in order to reidentificate respondents.

EO0458: **Enhancing information quality via official statistics data integration**
Presenter: **Luciana Dalla Valle**, University of Plymouth, United Kingdom
Co-authors: Ron Kenett
Information quality, or InfoQ, is the potential of a data set to achieve a specic (scientic or practical) goal by using a given empirical analysis method. This concept does not simply correspond to data and analysis quality, but is more broad and articulated. InfoQ is based on the identification of four interacting components: the analysis goal, the data, the data analysis, and the utility, and it is assessed through eight dimensions: data resolution, data structure, data integration, temporal relevance, generalizability, chronology of data and goal, construct operationalization and communication. We present a novel strategy to increase InfoQ, based on the integration of official statistics data with administrative data using nonparametric Bayesian Networks. Official statistics, despite being very informative sources of information, are sometimes neglected, leading to a poor level of InfoQ, implying low valued statistical analyses and the lack of sufficiently informative results. By improving temporal relevance and chronology of data and goals, the use of nonparametric Bayesian Networks allow us to calibrate official statistics with administrative or organizational data, thus strengthening the quality of the information derived from the official statistics surveys, and, overall, enhancing InfoQ. We show how to design and implement such a calibration strategy with two case studies on education and transport safety.

EO1099: **Assessing good fitting models in nonparametric disclosure risk estimation**
Presenter: **Silvia Polettini**, Sapienza Universita di Roma, Italy
Co-authors: Cinzia Carota, Maurizio Filippone
We report ongoing research on identification of models with good predictive performance for dislcosure risk estimation. Disclosure risk is usually estimated through parametric models on contingency tables of key variables that permit re-identifying sampled individuals. It has been proposed a Poisson model with rates explained by a log-linear mixed model with Dirichlet process (DP) random effects, to account for lack of fit, allowing to identify good fitting models with a simple fixed-effects structure. Given the size of contingency tables involved, the severe sparsity issues, and the complexity of the proposed Bayesian approach, formal model selection is challenging. The nonparametric main effects model is a starting point for the identification of an "optimal" model; as opposite to parametric models, sensitivity to model specification is low, and often adding a single two-way interaction, no matter which, leads to satisfactory performances. Note that the significant reduction of the space of models to be examined makes the issue of selection bias less important. We consider criteria to assess model's predictive performance. We investigate the effectiveness of such criteria when estimating the disclosure risk, in applications to real data and focus on the role of nonparametric random effects in reducing the selection bias.

EO1070: **Bayesian treatment of a multivariate Fay-Herriot functional measurement error model with applications**
Presenter: **Carolina Franco**, US Census Bureau, United States
Co-authors: Serena Arima, William Bell, Gauri Datta, Brunero Liseo
We explore Bayesian treatment of a multivariate Fay-Herriot measurement error model under the functional error setup. For noninformative uniform priors for both the true unknown covariate values that are measured with error and the regression coefficients, and for a class of noninformative priors for the model variance matrix, we establish conditions for the propriety of the resulting posterior distributions. We illustrate the model by applying it to estimating rates of school-aged children in poverty for U.S. counties for 2011 using data from the US Census Bureau's American Community Survey (ACS). A comparison of the posterior means and standard errors from the measurement error model to those from the "naive" model, which includes the measurement error variable as if it were a known regressor with no error, and from the "null" model, which excludes the measurement error variable, reveals interesting patterns.

EO114 Room Senate ADVANCES IN LATENT VARIABLE MODELS Chair: Angela Montanari

EO0492: **On the adoption of a mixed effects model to classify multiple samples represented by mixture distributions**
Presenter: **Geoffrey McLachlan**, University of Queensland, Australia

The modelling of multivariate observations in a random sample from a population composed of one or more components that have asymmetric distributions is considered. The distribution for each component is modelled by the skew normal and various skew $t-$distributions. In the case where there are multiple samples taken on objects with significant inter-object variation, we consider the alignment problem of matching the sample-specific clusters across the available samples. We discuss the joint clustering and matching (JCM) procedure, which accomplishes this task using random effects terms to allow for inter-sample variation among the objects. The clustering of the samples and the matching of the clusters are carried out simultaneously. This approach is to be illustrated by an application to a problem in flow cytometry where the aim is to assign an unclassified sample to one of two or more predefined classes corresponding to different states of a disease.

EO0722: **Using mixtures in seemingly unrelated linear regression models with non-normal errors**
Presenter: **Giuliano Galimberti**, University of Bologna, Italy
Co-authors: Gabriele Soffritti

Seemingly unrelated regression equations (SURE) refers a set of equations for modelling the dependence of D variables ($D \geq 1$) on one or more regressors in which the error terms in the different equations are allowed to be correlated. In such situations, the equations should be jointly considered. Most of the methods that have been developed to deal with SURE are based on the assumption that the distribution of the error terms is multivariate Gaussian. In order to allow departures from this assumption, the use of multivariate Gaussian mixtures is proposed. This approach has the advantage of capturing the effect of unobserved nominal regressors from the model and obtaining robust estimates of the regression coefficients when the distribution of the error terms is non- normal. Identifiability conditions for this new class of models are provided. The score vector and the hessian matrix of the corresponding log-likelihood function are derived. Maximum likelihood estimates for the unknown parameters are computed using an Expectation-Maximisation algorithm. In order to select the number of mixture components, the use of information criteria is suggested. The properties and the usefulness of the proposed methods are illustrated through the analysis of simulated and real datasets.

EO0580: **Approximate likelihood inference in generalized linear latent variable models based on integral dimension reduction**
Presenter: **Silvia Bianconcini**, University of Bologna, Italy
Co-authors: Silvia Cagnone, Dimitris Rizopoulos

Latent variable models represent a useful tool for the analysis of complex data when the constructs of interest are not observable. A problem related to the estimation of these models is that the integrals involved in the likelihood function cannot be solved analytically. We propose a new approach, referred to as Dimension Reduction Method (DRM), that consists in a dimension reduction of the multidimensional integral that makes the computation feasible in situations in which the quadrature-based methods are not applicable. We discuss the advantages of DRM compared with other existing approximation procedures in terms of both computational feasibility as well as asymptotic properties of the resulting estimators.

EO1135: **Estimating large covariance matrices via low rank plus sparse decomposition**
Presenter: **Matteo Farne**, University of Bologna, Italy

A large dimensional covariance matrix estimation method under the approximate factor model assumptions is presented. Existing methods perform estimation by extracting principal components and then applying a soft thresholding algorithm. In the described method, the low rank plus sparse decomposition for the covariance matrix is recovered by least squares minimization under nuclear norm plus l_1 norm penalization. The non-smooth convex minimization procedure is based on subdifferential methods, and results in a singular value thresholding plus a soft thresholding algorithm. Non asymptotic error rates can be derived under different theoretical assumptions on the eigenvalues and the sparsity pattern of Σ. It is possible to show that the unshrinkage of the estimated eigenvalues improves the performance of our estimator considerably. An ad- hoc model selection criterion which detects the optimal point in terms of composite penalty is proposed. A wide simulation study where various low rank plus sparse settings are simulated according to different parameter values is described and the improvements upon existing methods are outlined in detail.

EO130 Room Bedford ROBUSTNESS, BIAS REDUCTION, MULTIVARIATE EXTREME Chair: Armelle Guillou

EO0515: **Outlier detection for generalized Pareto distribution**
Presenter: **Mikhail Zhelonkin**, University of Lausanne, Switzerland

The problem of outlier detection is well known both in statistical and machine learning literature. In spite of the vast amount of developments the issue of outlier detection in extreme value theory has not been sufficiently addressed. The aberrant values can be legitimate extremes but can also be outliers. We assume that the true data generating process is a gross error model, which means that the majority of observations are generated from an assumed model, and a small portion comes from an unknown contaminating model. Such situations can naturally arise in practice. The typical examples are in risk management, when several sources of risk can be grouped together. The possible consequences are inflated capital requirements and inefficient asset allocation. In such situations it is desirable to detect the possible atypical observations in order either to remove them if they are technical errors or to mark them for careful examination and possible model adjustments. The classical distance- or depth-based methods are not applicable in the tails, as well as outlier tests. We propose a quantile-based outlier detection algorithm for data following the Generalized Pareto Distribution, study its properties and provide an illustrative example with real data.

EO0319: **Robust and bias-corrected estimation in multivariate extreme value statistics**
Presenter: **Yuri Goegebeur**, University of Southern Denmark, Denmark
Co-authors: Christophe Dutang, Yuri Goegebeur, Armelle Guillou

A robust and asymptotically unbiased estimator is introduced for the coefficient of tail dependence in multivariate extreme value statistics. The estimator is obtained by fitting a second order model to the data by means of the minimum density power divergence criterion. The asymptotic properties of the estimator are investigated. As an extension, the estimation of the probability of a tail region is considered. The efficiency of our methodology is illustrated on a small simulation study and by a real dataset from the actuarial context.

EO0170: **Estimation of the multivariate conditional-tail-expectation for extreme risk levels: Illustration on a rainfall data-set**
Presenter: **Elena Di Bernardino**, CNAM, France

The problem of estimating the multivariate version of the conditional-tail-expectation already introduced in the bivariate framework is considered. We propose a new semi-parametric estimator for this risk measure, essentially based on statistical extrapolation techniques, well designed for extreme risk levels. We prove a central limit theorem. We illustrate the practical properties of our estimator on simulations. The performance of our new estimator is discussed and compared to the ones of the empirical Kendall's process-based estimator, previously proposed. We conclude with an application to a rainfall data-set.

EO0398: **Estimation of extreme depth-based quantile regions**
Presenter: **Yi He**, Tilburg University, Netherlands
Co-authors: John Einmahl
Consider the extreme quantile region, induced by the halfspace depth function HD, of the form $Q = \{\mathbf{x} \in R^d : HD(\mathbf{x}, P) \leq \beta\}$, such that $PQ = p$ for a given, very small $p > 0$. This region can hardly be estimated through a fully nonparametric procedure since the sample halfspace depth is 0 outside the convex hull of the data. Using Extreme Value Theory, we construct a natural, semiparametric estimator of this quantile region and prove a refined consistency result. A simulation study clearly demonstrates the good performance of our estimator. We use the procedure for risk management by applying it to stock market returns.

EO226 Room CLO 101 DEPTH AND FUNCTIONAL DATA ANALYSIS Chair: Juan Romo

EO0584: **Exact computation of the Tukey depth via linear optimization**
Presenter: **Pavlo Mozharovskyi**, University of Cologne, Germany
The Tukey depth of z w.r.t. $X = \{x_1, ..., x_n\}$ in $\mathbb{R}^d$ is defined as the smallest fraction of X lying in a closed halfspace with z on its boundary. Taking into account the geometry of X in a nonparametric and highly robust way, it possesses a number of attractive properties. Obtaining the exact value of the Tukey depth is a nontrivial computational task, to which a substantial part of the literature on this depth notion is devoted. Its exact calculation requires up to $\sum_{i=0}^{d-1} \binom{n-1}{d}$ separations of X by a hyperplane through z to be considered. One possibility is to fill $\mathbb{R}^d$ with a number of direction cones of constant 1-dimensional depth. We present a new algorithm that examines these cones by the breadth-first search, applying linear programming and binary coding of the cones. This provides high execution speed when working with a cone segmentation and allows for further generalizations. The algorithm and its slight modifications can be used to compute further depths, e.g. regression depth, solve the densest hemisphere problem, find a hyperplane with minimal empirical risk for binary supervised classification.

EO0663: **Robust outlier detection methods for functional data**
Presenter: **Francesca Ieva**, Universita degli Studi di Milano, Italy
Co-authors: Nicholas Tarabelloni
Functional data is a very attractive model for modern applications, yet in view of practical analyses of functional datasets a major issue is the identification of their extreme observations. Due to their infinite-dimensional nature, even after a dimensional reduction, functional data are always high-dimensional compared to their typical sample size. For this reason, even a little contamination of the dataset may lead to unreliable inferential conclusions. From an operational point of view, two main kinds of outlying behaviours are generally considered: the first is related to the amplitude of data (amplitude outliers); the second is related to the phase of data (shape outliers), which can be due to different running times of the units. In principle, outliers can be addressed either by building robust estimators for the quantities involved in the inferential procedures, or by suitably robustifying the dataset, through the use of ad-hoc techniques; an example is the adjusted functional boxplot, based on the notion of statistical depth and adapted to the dataset at hand, which targets amplitude outliers. We explore the possibility to combine the strengths of both approaches in order to build a robust version of the adjusted functional boxplot.

EO0918: **Homogeneity test for functional data**
Presenter: **Ramon Flores**, Universidad de Sevilla, Spain
Co-authors: Rosa Lillo, Juan Romo
In the context of functional data analysis, we propose new two-sample tests for homogeneity. Based on some well-known depth measures, we construct four different statistics in order to measure distance between the two samples. A simulation study was performed to check the efficiency of the tests when confronted with shape and magnitude perturbation. Finally, we applied these tools to measure the homogeneity in some samples of real data, as tecator data or growth data.

EO1124: **Depth-based ordering of general functions and applications**
Presenter: **Sara Lopez Pintado**, Columbia University, United States
Biomedical diagnosis is increasingly reliant on the analysis of complex data, such as functions and images. In some applications the data consists of correlated functions for each sample subject. Examples are the changes in weight and height over time of individuals, or multiple lead recordings from an electrocardiogram. Data can also consist of multiple brain images, such as functional magnetic resonance imaging (fMRI) recordings, for individual subjects, in different neurophysiological states. Developing new statistical tools to analyze these rich data sets has become a limiting factor for the advancement of many disciplines. In all of the applications mentioned above the basic unit of observation can be considered as a general function which is defined in a subset of either the real line or a higher dimensional space, taking values in a univariate or multivariate space. We propose a general definition of depth for functions defined in spaces of arbitrary dimensions, which we call multivariate modified volume depth. The theoretical properties of this depth will be established and it will be used as a building block for extending robust inferential methods to the analysis of general complex functions. In particular a depth-based non-parametric test for comparing different group of images will be proposed and used for testing if there are differences in the brain structure or function between healthy individuals and patients with major depressive disorders.

EO110 Room CLO 306 MODEL SELECTION WITH APPLICATIONS IN GENETICS Chair: Florian Frommlet

EO0588: **Performance of a blockwise approach in variable selection using linkage disequilibrium information**
Presenter: **Alia Dehman**, Evry Val d Essonne University, France
Co-authors: Christophe Ambroise, Pierre Neuvial
Genome-wide association studies (GWAS) aim at finding genetic markers that are significantly associated with a phenotype of interest. Single nucleotide polymorphism (SNP) data from the entire genome are collected for many thousands of SNP markers, leading to high-dimensional regression problems where the number of predictors greatly exceeds the number of observations. Moreover, these predictors are statistically dependent, in particular due to linkage disequilibrium (LD). We propose a three-step approach that explicitly takes advantage of the grouping structure induced by LD in order to identify common variants which may have been missed by single marker analyses (SMA). In the first step, we perform a hierarchical clustering of SNPs with an adjacency constraint using LD as a similarity measure. In the second step, we apply a model selection approach to the obtained hierarchy in order to define LD blocks. Finally, we perform Group Lasso regression on the inferred LD blocks. We investigate the efficiency of this approach compared to state-of-the art regression methods. Our results show that the proposed method is efficient not only at the level of LD blocks by inferring well the underlying block structure but also at the level of individual SNPs.

EC0231: **Detection of gene-environment effects in GWAS using logistic regression with latent exposure**
Presenter: **Gregory Nuel**, CNRS, France
Co-authors: Flora Alarcon
The detection of gene-environment (GE) interactions is of utmost interest in genetic epidemiology, particularly as it leads to a better understanding of underlying disease etiology and subsequently the development of disease prevention and intervention strategies. However, to date, only few loci that interact with the environment have been discovered, demonstrating that the problem is very challenging due to various causes, including

the presence of confounding factors. Indeed, it is very common that environment exposure is only observed through proxy covariates that often act as confounding factors and cause a drastic loss of power in GE interaction detection. We suggest accounting for these confounding factors by introducing a binary latent exposure in the logistic regression. This logistic regression with latent exposure (LRLE) model enables a significant power gain in the detection of GE interactions. The usefulness of the model is illustrated through a simulation framework where its properties and performances are studied and compared to two standard approaches for detecting GE interactions: the cases-controls and the cases-only tests. Through these simulations, LRLE proves itself to provide a dramatic power improvement over existing approaches and thus appear as a promising new method for detecting GxE interactions in GWAS.

EO1333: **Adaptive selection via SLOPE**
Presenter: **Damian Brzyski**, Jagiellonian University, Poland

Sorted L-One Penalized Estimation (SLOPE) is a relatively new convex optimization procedure which enables the adaptive selection of regressors under sparse high dimensional designs. The method was designed to control the expected proportion of irrelevant regressors among all selected predictors (false discovery rate, FDR). It was shown that SLOPE performs well in situations when explanatory variables are nearly orthogonal. The control over FDR could be however lost in the presence of strong correlations, which usually happens in the context of Genome Wise Association Studies (GWAS). We show that SLOPE can be successfully used in GWAS after introducing the additional clustering step, where clusters correspond to the groups of strongly correlated SNPs. We use computer simulations to show that after this modification SLOPE controls FDR defined at the clusters level and that it compares favorably to other popular methods for GWAS.

EC1533: **Variable selection in binomial regression with latent Gaussian field models for analysis of epigenetic data**
Presenter: **Aliaksandr Hubin**, University of Oslo, Norway
Co-authors: Geir Olve Storvik

Epigenetic observations data is represented by the total amount of reads from a particular cell and the amount of methylated reads, which are reasonable to model via a Binomial distribution. There are numerous factors that might influence the probability of success from a particular region. We might also expect spatial dependence of these probabilities. We incorporate dependence on the covariates and spatial dependence of probability of being methylated for observation from a particular cell by means of a binomial regression with latent Gaussian field model. We use INLA approach for calculating posterior marginal likelihoods for fixed models and carry out efficient MCMC with locally optimized mode jumping proposals across the models in order to draw from the posterior distributions of parameters and models jointly. During these MCMC walks we also simultaneously perform model selection with respect to different criteria to discover the optimal choice of covariates that influence methylation structure in the regions along the genome.

EO080 Room Woburn MULTIVARIATE SURVIVAL MODELS Chair: Roel Braekers

EO0608: **Parametric maximum likelihood inference for copula models with dependently left-truncated data**
Presenter: **Takeshi Emura**, National Central University, Taiwan
Co-authors: Chi-Hung Pan

Traditionally, most literature on the left-truncated samples considers statistical inference by assuming that left-truncation time L is independent of the lifetime X. However, dependence between L and X occurs, for instance in a community health and a field reliability study. A copula-based parametric model is considered for dependent truncation. Then we consider the maximum likelihood estimator (MLE) of the unknown parameters. To calculate the MLE, a key technical challenge is to obtain the form of the inclusion probability $Pr(L \leq X)$ and its partial derivatives with respect to parameters. We first show that, under the copula model on (L,X), the probability $Pr(L \leq X)$ is expressed as a Reimman integral of the $h-$function on the unit interval. With these new expressions, we propose the Newton-Raphson algorithm to maximize the log-likelihood and the Hessian matrix to calculate standard errors. Simulations are conducted to examine the performance of the proposed method. Real data from a field reliability study on the brake pad lifetimes are analyzed for illustration.

EO0792: **A new nonparametric copula estimator for clustered right-censored event time data**
Presenter: **Candida Geerdens**, Hasselt University, Belgium

In many scientific studies the response of interest is the time until a predefined event (e.g., the time to tumor appearance). Often, this event time is right-censored for some items in the study sample, i.e., only a lower time bound for the event is observed (e.g., by the end of the study still no tumor has appeared). A further complexity can be the grouping of data (e.g., the study might include only twins, each twin then acts as a group of size two). Since clustered items share common traits, their event times show correlation. Copulas provide a popular tool to describe the association in grouped time-to-event data. In a data setting where it is less evident to predetermine a parametric copula, one may opt to apply a nonparametric copula. We define a new nonparametric copula estimator for the joint survival function of grouped right-censored event time data. We consider two right-censoring schemes: univariate censoring and copula censoring. For the new nonparametric copula estimator, we establish the consistency and we assess the finite sample performance in various data settings via a simulation study.

EO0791: **Using Archimedean copulas to model multivariate survival data with multilevel clustering**
Presenter: **Leen Prenen**, Hasselt University, Belgium

Lifetime data can sometimes be grouped in clusters within clusters. Consider for example the survival times of cancer patients in hospitals, clustered within geographical areas. To model this hierarchical dependency structure in multivariate survival data, we can use hierarchical copulas. Hereby the joint survival function of the lifetimes is expressed as a copula function, in which some arguments are replaced by another nested copula function, that is evaluated in the marginal survival functions. Contributions to the loglikelihood of a (sub)cluster are given by partial derivatives of the joint survival function w.r.t. the uncensored observations. We use hierarchical Archimedean copulas to model time to infection of the four parts of a cow udder and study the effects of incorporating more complicated association structures in the model. Models with a higher level of nesting are compared to simpler models using the likelihood ratio test. We conclude that the association between infection times within two front/rear udder parts is different from the association between one front and one rear udder part. Estimation procedures comprise one-stage and two-stage parametric estimation and two-stage semiparametric estimation. Simulation studies are used to estimate power and size of the tests.

EO1453: **Copula-based estimation of a distribution function under interval sampling**
Presenter: **Carla Moreira**, University of Vigo, Spain
Co-authors: Jacobo de Una-Alvarez, Roel Braekers

We consider the problem of estimating a distribution function from doubly truncated data, when the right-truncation variable equals the left-truncation variable plus a constant (that is, under interval sampling). It is assumed that the dependence structure between the left-truncation variable and the variable of interest is ruled by a parametric copula. Maximum-likelihood estimation of both the copula parameter and the marginal distributions of the variable of interest and the left-truncation variable is introduced. For the computation of the estimators, an iterative algorithm and a simplification of it are proposed and compared in an extensive simulation study. The main conclusions are that the simplified algorithm is recommended due to its good relative performance and computational speed, and that both algorithms outperform the Efron-Petrosian estimator for doubly truncated data when dependence is severe. Real data illustrations are provided too.

EO340 Room CLO 204 DESIGN AND ANALYSIS OF CLINICAL TRIALS Chair: Sergei Leonov

EO0625: **A comparison of five approaches to decision making for a first clinical trial of efficacy**
Presenter: **Simon Kirby**, Pfizer, United Kingdom
The first trial of clinical efficacy is an important step in the development of a compound. Such a trial gives the first indication of whether a compound is likely to have the efficacy needed to be successful. Good decisions dictate that good compounds have a large probability of being progressed and poor compounds have a large probability of being stopped. We consider and contrast five approaches to decision making that have been used. To illustrate the use of the five approaches we conduct a comparison for two plausible scenarios with associated assumptions for sample sizing. The comparison shows some large differences in performance characteristics of the different procedures. Which decision making procedures and associated performance characteristics are preferred will depend on the focus of interest and the decision-makers attitude to risk.

EO0695: **Asymptotic properties of certain compromise-optimal covariate-adjusted allocations in clinical trials**
Presenter: **Asya Metelkina**, CNRS, France
Co-authors: Luc Pronzato
Treatment allocation is considered when the success probabilities depend on observed covariates through generalized linear models. We assume that the objective of the clinical trial is statistical inference about success probabilities as functions of covariates, through the estimation of the parameters of each treatment model. We also aim at taking individual ethics into account by favouring allocation of the best treatment for each patient. A concave function of the Fisher information matrix is used to measure precision on model parameters, a regret function is used to account for the ethical constraint. The locally optimal covariate-adjusted allocation is then deterministic; it corresponds to a particular design measure that maximizes a convex combination of these two functions. It can be constructed explicitly by exploiting necessary and sufficient conditions for optimality (Equivalence Theorem). However, (i) it requires knowledge of the distribution of covariates, (ii) the construction seems too complicated for trials involving non-scalar covariates. We show how these two difficulties can be avoided by using a covariate-adaptive allocation rule based on empirical allocation measures, which converges to the optimal allocation. A response-adaptive version, that uses current parameter ML estimates, is also shown to converge to the optimal allocation measure for the true unknown model parameters. Comparison with recent results of the literature on CARA design will be given.

EO1136: **Constrained dose selection and estimation of pharmacokinetic parameters in adaptive clinical trials**
Presenter: **Barbara Bogacka**, Queen Mary University of London, United Kingdom
We present an adaptive design for dose finding in phase I/II trials, where the probabilities of efficacy and toxicity as well as pharmacokinetic (PK) information are considered in the dose-selection procedure. The local D-optimal design for estimating PK parameters is found and applied in each step of an adaptive trial. Also, the responses to the drug are measured and the PK and dose-response models are updated accordingly. A new, optimum dose is then selected, based on the updated information. An ethical approach is considered in this case, where the dose for the next cohort is optimized for efficacy of the response with some constraints on toxicity. We also consider the total exposure to the drug as an additional constraint for dose selection. The area under the drug concentration curve reflects the population variability in the drug absorption and elimination and this fact is included in the procedure. This method gives efficient designs for dose finding from the point of view of population PK parameter estimation and ethical dose selection with maximum probability of efficacy while keeping the chances of toxicity under control.

EO1067: **Modelling of correlated random variables in clinical trials**
Presenter: **Sergei Leonov**, ICON Clinical Research, United States
Modelling and simulation of correlated random variables is important for evaluating operating characteristics of clinical trial designs that involve multiple endpoints, some of which may be discrete and some continuous (e.g., number of events vs. change of a continuous score). There exist efficient algorithms to address the problem of generating multivariate distributions with given marginals and given correlation structure, in particular NORTA (NORmal To Anything). To implement such algorithms, it is important to know the extreme values of pairwise correlations, which may be less than 1 in absolute value. We provide closed-form expressions for several classes of multivariate distributions that involve both discrete and continuous endpoints and illustrate the performance of algorithms via several examples.

EO112 Room Torrington RECENT ADVANCES IN TIME SERIES ANALYSIS Chair: Piotr Fryzlewicz

EO0668: **Estimation for nonlinear state space models**
Presenter: **David Stoffer**, University of Pittsburgh, United States
Ever wonder how the Mars One one-way trip to Mars will actually get to the planet without winding up on, say Venus? The tracking devices will use a nonlinear state space model. While inference for the linear Gaussian model is fairly simple, inference for nonlinear models can be difficult and often relies on derivative free numerical optimization techniques. A promising method that we discuss is based on particle approximations of the conditional distribution of the hidden process given the data. This distribution is needed for both classical inference (e.g., Monte Carlo EM type algorithms) and Bayesian inference (e.g., Gibbs sampler). Particle methods are an extension of sequential importance sampling (SIS). Although the SIS algorithm has been known since the early 1970s, its use in nonlinear problems remained largely unnoticed until the early 1990s. Obviously the available computational power was too limited to allow convincing applications of these methods, but other difficulties plagued the technique. Time series data are typically long and particles have a tendency to die young. Consequently, the approach is cursed by dimensionality. But as Shakespeare noted, if dimensionality curseth, a better algorithm useth.

EO0835: **Log-concavity and quasi-maximum likelihood estimation**
Presenter: **Yining Chen**, London School of Economics and Political Science, United Kingdom
We propose a general quasi-maximum likelihood approach in time series analysis. Our quasi-likelihood function is constructed by assuming that the innovations follow a log-concave distribution. This framework can be easily adapted to many well-known time series models, including autoregressive moving-average (ARMA) models, threshold autoregressive (TAR) models, generalised autoregressive conditional heteroscedasticity (GARCH) models, and ARMA-GARCH models. Furthermore, we show that the estimator under this new framework is consistent in all the above-mentioned settings, even when the innovations are not log-concave. We demonstrate its promising finite sample performance via a thorough simulation study and apply it to model the daily return of FTSE 100 index.

EO0871: **Multiscale autoregression in time series modelling**
Presenter: **Piotr Fryzlewicz**, London School of Economics, United Kingdom
Co-authors: Rafal Baranowski
Motivated by the practice of simultaneously considering multiple time scales in financial time series modelling and forecasting, we introduce the concept of "multizoom" autoregressive time series models, in which the process in question is regressed on certain features of its own past, each living on a different time scale. This is in contrast to classical autoregressive modelling, in which the process is modelled as a function of a few most recent observations. In our approach, the time scales are not necessarily pre-determined and can be estimated from data. We describe some simple probabilistic properties of the new class of models, and provide an estimation theory for it. One appealing aspect of this way of modelling is that

it is able to (nontrivially) capture the "lack of serial autocorrelation" feature commonly observed in returns on the prices of financial instruments. We illustrate the new approach and show its benefits in an extensive high- and mid-frequency forecasting exercise involving transaction prices on equities.

EO0761: **Bootstrap-based simultaneous confidence bands for sample autocorrelations**
Presenter: **Jens-Peter Kreiss**, TU Braunschweig, Germany
Co-authors: Efstathios Paparoditis, Han Xiao
A bootstrap procedure is proposed and investigated, which allows for a rather general class of stationary time series an asymptotically correct approximation of the distribution of the maximum of sample autocorrelations over an increasing set of lags. This result immediately leads to simultaneous confidence bands for sample autocorrelations. The proposed bootstrap procedure is a hybrid procedure which consists of a combination of an autoregressive sieve bootstrap on the one hand and a block bootstrap procedure on the other hand. Simulations will be presented which demonstrate the finite sample behaviour of the proposed procedure and further possible applications of the proposed resampling scheme will be indicated.

EO611 Room Athlone STATISTICAL METHODS FOR IMPERFECT DATA Chair: Andrea Wiencierz

EO1089: **Reliable analysis of categorical data under epistemic data imprecision**
Presenter: **Julia Plass**, Ludwig-Maximilians University, Germany
Co-authors: Thomas Augustin, Marco Cattaneo, Georg Schollmeyer
In many problems of data analysis for a part of the data only coarsened versions of the true values are observable. Compared with the practical relevance of the issue, these so-called coarse data under epistemic imprecision have obtained little attention in traditional statistical methodology so far. Thus, it is still common to impose strict assumptions as coarsening at random, although a substantial bias may result. The cautious statistical modelling of categorical data under epistemic data imprecision is considered by discussing the case of independent and identically distributed variables as well as regression under categorical covariates. Thereby, an observation model formalizing the information available on the coarsening process constitutes the core of the approach leading to (typically set valued) maximum likelihood estimators of the underlying distributions. If one benefits from additional auxiliary information, sensitivity parameters allow the refinement of these estimators and furthermore offer an additional confidence interval based treatment of additional statistical sampling uncertainty. Alternatively, knowledge about the coarsening may be included via sets of priors, ending up in a generalized Bayesian approach. The methodology is illustrated by the data from the German panel study Labour Market and Social Security, where the income is partially observed in a coarse way.

EO1353: **M-estimation when data values are not completely known**
Presenter: **Marco Cattaneo**, University of Hull, United Kingdom
Real data often do not have the level of precision required by conventional statistical methods. In particular, a data point can be incompletely observed, in the sense that the only available observation is a set known to contain the data point. An important problem is then how to perform statistical estimation, and in particular regression, when some (or all) data points are incompletely observed. In the case of precisely observed data, most statistical estimation methods can be expressed as M-estimators (or slight generalizations thereof), i.e. the estimate minimizes some kind of distance from the data. Now, M-estimators can be generalized to the case in which data points are incompletely observed by minimizing the distance from the set of all possible data (i.e. all precise data compatible with the imprecise observations). Different estimators with different properties can be obtained depending on how the distance from a set is defined. Of particular interest are the minimax and minimin estimators obtained by minimizing the distance from the worst-case and best-case possible data, respectively.

EC1092: **Partial identification in linear models: Regression with interval-valued data**
Presenter: **Georg Schollmeyer**, Ludwig Maximilaians University Munich, Germany
Co-authors: Thomas Augustin
In many situations of statistical analysis, the variables of actual interest cannot be observed in the precision that is needed to meet the requirements to justify the application a classical standard procedure of statistical data analysis. One prominent example is the request of the income of participants in a survey. There, respondents often refute an answer and one has to deal with a big amount of missing values that could not be treated as missing at random. One way to circumvent this problem is to allow respondents to give a categorized answer about their income to decreases non-response. In this situation, parametric statistical models are generally only partially identified, meaning that also with an infinite amount of data, there is still a rest of systematic uncertainty about the true parameter(s) because different parameter settings could lead to exactly the same distribution of the observed coarsed income. Instead of imposing further, often unjustified assumptions to enforce model identification, the methodology of partial identification deals with the identification problem by looking at the set of parameter settings compatible with the observable random variables and the underlying assumed model. We compare approaches for estimating partially identified linear models based on moment inequalities with other approaches that essentially determine the envelopes of estimates arising from all potential data completions.

EC1746: **Cramer-Von Mises test for interval data**
Presenter: **Sebastien Destercke**, Universite de Technologie de Compiegne, France
Co-authors: Olivier Strauss, Hugo Saulnier
We are interested in extending the classical Cramer-Von Mises homogeneity test to compare two samples of interval-valued observed measurements. In such a case, the test statistic becomes itself set-valued, and the main difficulty is to find the bounds of this set. We propose conservative approximations of these bounds that arise from a $p-$box (pairs of upper and lower cumulative distributions) representation of the samples. We also perform some first experiments to determine the quality of these easy-to-compute bounds.

EO328 Room Bloomsbury GRAPHICAL MODELS AND CAUSALITY Chair: Dominik Rothenhaeusler

EO1104: **Recovering Information from Missing Data using Graphical Causal Models**
Presenter: **Karthika Mohan**, University of California Los Angeles, United States
Datasets in which values of one or more variables are missing from recorded observations are a menace to empirical sciences. Based on the underlying cause of missingness these datasets can be broadly classified into Missing Completely At Random (MCAR), Missing At Random (MAR) data and Missing Not At Random (MNAR). We focus primarily on the relatively unexplored MNAR datasets. Our recent work, which employs graphical models to portray the data generation process, enables us to determine the feasibility of recovering i.e. computing a consistent estimate of, certain parameters as if no data were missing. We further devise procedures that yield closed form estimands of parameters that are deemed recoverable.

EO0855: **Causal discovery from real-world data sets: The next big challenge**
Presenter: **Tom Claassen**, Radboud University Nijmegen, Netherlands
Despite recent advances in speed, power, and accuracy of methods that, in principle, are able to learn valid causal models from purely observational

data, the step towards effective, commonplace application to real-world data sets is still hampered by the fact that in practice many of the crucial underlying assumptions are not satisfied. We tackle ways to handle two prominent problems associated with causal inference in such settings: complex mixtures of discrete and continuous variables with missing data, and so-called faithfulness violations. First we introduce a copula-inspired approach into a state-of-the-art Bayesian technique in order to transform the required non-standard independence tests between mixed variables into a straightforward MVG scoring problem on small subgraphs. Secondly we identify ways to recognize and account for certain types of apparent faithfulness violations. We show how additional variables may be used to decide between conflicting statements, and use this to measurably improve the accuracy of the output model. We demonstrate the effectiveness of the method on an overlapping set of neuro-psychological studies into psychopathic and aggressive traits (a typical example involving questionnaires, expert assessments, external socio-economical factors, hormonal measurements etc. with partial compliance in follow-up tests), as well as on artificial data from models that simulate realistic answer and behavioural patterns of participants.

EO1320: **Conditional instrumental variables: What they are and how to find them**
Presenter: **Johannes Textor**, University of Utrecht, Netherlands

Instrumental variables (IVs) are widely used to identify causal effects. For this purpose, IVs have to be exogenous. However, exogenous variables can be hard to find. A generalized IV method has been proposed, which only requires exogeneity conditional on a set of covariates. This leads to a wider choice of potential IVs, but is rarely used yet. We address two issues with conditional IVs. First, they are conceptually rather distant to standard IVs; even variables that are independent of X could qualify as conditional IVs. We propose a new concept called ancestral IV, which interpolates between the two existing notions. Second, only exponential-time algorithms had been known to find conditional IVs in a given causal diagram, and we prove that this problem is indeed NP-hard. Nevertheless, we show that whenever a conditional IV exists, so does an ancestral IV, and ancestral IVs can be found in polynomial time. Together this implies a complete and constructive solution to causal effect identification using IVs in linear causal models.

EO1322: **Understanding consistency in hybrid causal structure learning**
Presenter: **Preetam Nandy**, ETH Zurich, Switzerland
Co-authors: Alain Hauser, Marloes Maathuis

We consider causal structure learning from observational data. The main existing approaches can be classified as constraint-based, score-based and hybrid methods, where the latter combines aspects of both constraint-based and score-based approaches. Hybrid methods often apply a greedy search on a restricted search space, where the restricted space is estimated using a constraint-based method. The restriction on the search space is a fundamental principle of the hybrid methods and makes them computationally efficient. However, this can come at the cost of inconsistency or at least at the cost of a lack of consistency proofs. We demonstrate such inconsistency in an explicit example. In spite of the lack of consistency results, many hybrid methods have empirically been shown to outperform consistent score-based methods such as greedy equivalence search (GES). We present a consistent hybrid method, called adaptively restricted GES (ARGES). It is a modification of GES, where the restriction on the search space depends on an estimated conditional independence graph and also changes adaptively depending on the current state of the algorithm. Although the adaptive modification is necessary to achieve consistency in general, our empirical results show that it has a relatively minor effect. This provides an explanation for the empirical success of (inconsistent) hybrid methods.

EO122 Room Holden METHODS AND MODELS FOR SOCIAL NETWORKS DATA Chair: Giuseppe Giordano

EO1109: **Large-scale Bayesian estimation for component exponential random graph models**
Presenter: **Johan Koskinen**, University of Manchester, United Kingdom

While the likelihood for the Exponential random graph model (ERGM) in general is doubly intractable, ML inference may be based on the method of moments, where moments are approximated by their Monte Carlo estimates. As the number of tie-variables grows quadratically with the number of nodes, numerical optimisation based on simulated quantities quickly becomes cumbersome for large networks. Bayesian inference for ERGM also relies on MCMC involving simulation of networks and therefore does not scale. For non-network data, Pseudo-marginal MCMC (PMCMC) has been proposed as a way in which to scale up MCMC-based inference. PMCMC does however require the likelihood to factor in order for subsampling of data to be permissible. A well-known problem with networks is unfortunately that ERGMs do not marginalise in the sense that subgraphs of a graph that follows an ERGM do not follow ERGMs. Recently it was proved that a particular class of ERGM, component ERGM (CERGM), does marginalize in the sense that, under certain conditions, the conditional likelihood for a graph factors across its components. We demonstrate the application of PMCMC to CERGM and discuss the use of PMCMC for approximate analysis of regular ERGM for large networks.

EO1050: **A stochastic actor oriented approach to model policy-driven innovation network dynamics**
Presenter: **Federica Rossi**, Birkbeck University of London, United Kingdom
Co-authors: Domenico De Stefano, Margherita Russo, Annalisa Caloffi, Susanna Zaccarin

Stochastic actor oriented models (SAOM) are of growing importance to study network dynamics underlying the theoretical micro mechanisms that induce the evolution of relations among a set of social actors. SAOM represents a suitable methodological framework to investigate the evolution of policy driven innovation networks among heterogeneous agents. The empirical application focuses on a set of policies that support networking among SMEs and other partners implemented by the Tuscany region (Italy) in the programming period 2000/2006. We consider a set of about 350 agents involved in 168 funded projects of variable duration ranging from 4 up to 18 months. Model specification accounts for several structural hypotheses on strategic partner choice in undirected networks (transitivity, popularity, homophily), controlling for agents characteristics (e.g. agent type, localisation) and project duration. In order to realistically frame how the decision to create or dissolve collaboration ties is taken, we consider either one-sided or two-sided initiative models.

EO0830: **Exploring and modelling multigraphs**
Presenter: **Termeh Shafie**, University of Konstanz, Germany

A multigraph approach to analyse networks with multiple edges and edge loops is introduced. Multigraph data structure is described with examples of their natural appearance, together with a description of the possibility to obtain multigraphs using blocking, aggregation and scaling. A novel way of representing multigraphs using edge multiplicities is presented and graph complexity is quantified by the distribution of edge multiplicities. Using this representation, a random multigraph model based on independent edge assignments (IEA) to sites of vertex pair is given and several complexity statistics under IEA are derived. It is described how these statistics can be used to analyse local and global network properties and to convey structural dependencies in social networks. Further, some natural extensions to this approach are presented including (i) an alternative random multigraph model called random stub matching (RSM) which is a special kind of preferential attachment model, (ii) information theoretic tools that may be used to explore interdependencies among network variables, and (iii) an application of these tools to select general exponential random graph models.

EO1321: **Exploratory data analysis and contiguity relations: An outlook**
Presenter: **Giuseppe Giordano**, Universita degli Studi di Salerno, Italy
Co-authors: Gilbert Saporta, Maria Prosperina Vitale

The use of Multidimensional Data Analysis in the framework of Social Network Analysis in order to explore roles and positions of actors in a network will be addressed. Emphasis will be put on data handling and visualization. Firstly, the use of Contiguity Analysis able to deal with relational data will allow defining data analysis able to synthesize and visualize the patterns of social relationships in a metric space. Then, since interactions are often derived from the presence of actors at events or activities (two-mode network data), in addition to the relational data, the availability of external information gathered on both actors and events will be considered. To explore the effect of external information about ties, we show how to decompose the original network data matrix and representing external information with a suitable matrix coding. This allows obtaining peculiar relational data matrices that include the effects of such new information. The derived adjacency matrices can then be analyzed from the network analysis perspective. In particular, we look for groups of structurally equivalent actors obtained through clustering methods. Illustrative examples in the framework of scientific collaboration and Marketing will give a major insight into the proposed strategy.

EO598 Room G21A MATRIX COMPUTATIONS AND PARALLELISM IN STATISTICS Chair: Cristian Gatu

EO1291: **Efficient scalable Monte Carlo and quasi-Monte Carlo preconditioners and hybrid methods for solving SLAE**
Presenter: **Aneta Karaivanova**, IICT-BAS, Bulgaria
Co-authors: Vassil Alexandrov, Oscar Esquivel

An extensive study will be presented on designing efficient highly parallel Monte Carlo preconditioners and hybrid methods for solving systems of linear algebraic equations. Thus we present a Monte Carlo preconditioner based on sparse Monte Carlo matrix inversion, which uses Markov Chain Monte Carlo methods to compute a rough approximate matrix inverse. This rough inverse is further used as a preconditioner while solving systems of linear algebraic equations. The proposed method is compared with the enhanced deterministic preconditioners computed by the optimized parallel variant- Modified SParse Approximate Inverse Preconditioner (MSPAI). Results of a study comparing the efficiency of sparse parallel Monte Carlo preconditioners with the MSPAI will be presented. Some results comparing sparse approximate Monte Carlo and sparse approximate quasi-Monte Carlo preconditioners will also be presented. Further results on how these can be optimized by an iterative filter process and a parallel refinement, to enhance the accuracy of the inverse and the preconditioner respectively will be presented. Computational experiments with hybrid methods using Monte Carlo preconditioners and deterministic methods such as BiGSTAB and GMRES to solve systems of linear equations as well as comparison with MSPAI/BiGSTAB and MSPAI/GMRES showing the efficiency of our approach will be demonstrated on classes of dense and general sparse matrices.

EC1498: **Comprehensive factor analysis by a matrix decomposition approach**
Presenter: **Kohei Adachi**, Osaka University, Japan
Co-authors: Nickolay Trendafilov

A new factor analysis (FA) procedure has recently been presented which can be called matrix decomposition FA (MDFA). All FA model parameters (common and unique factors, loadings, and unique variances) are treated as fixed unknown matrices. The main purpose is to show that: [1] MDFA can be presented by a comprehensive FA model in which the unique factors are separated from the model errors in contrast to the currently prevalent FA model; and [2] the model and the error parts can be estimated. As in the standard FA, the common and unique factors cannot be uniquely determined since their estimation involves more variables than the original ones available from the data matrix. The new mathematical results are formally arranged as theorems, and are followed by numerical examples.

EC1732: **An iterative hard thresholding estimator for low rank matrix recovery with explicit limiting distribution**
Presenter: **Arlene Kyoung Hee Kim**, University of Cambridge, United Kingdom
Co-authors: Alexandra Carpentier

We consider the problem of low rank matrix recovery in a stochastically noisy high dimensional setting. We propose a new estimator for the low rank matrix, based on the iterative hard thresholding method, and that is computationally efficient and simple. We prove that our estimator is optimal both in terms of the Frobenius risk, and in terms of the entry-wise risk uniformly over any change of orthonormal basis. This result allows us, in the case where the design is Gaussian, to provide the limiting distribution of the estimator, which is of great interest for constructing tests and confidence sets for low dimensional subsets of entries of the low rank matrix.

EC1838: **lmSubsets: An R package for regression subset selection**
Presenter: **Cristian Gatu**, Alexandru Ioan Cuza University of Iasi, Romania
Co-authors: Marc Hofmann, Erricos John Kontoghiorghes, Achim Zeileis

The R package "lmSubsets" for regression subset selection is introduced and described. Computationally efficient algorithms to compute the subsets are based on regression trees and employ branch-and-bound techniques and heuristics strategies. The main numerical tool that has been employed is the QR factorization and its modification. This yields in a numerically stable and efficient sub-model estimation procedure. The package aims to provide a versatile tool for subset regression. The algorithms are described in detail and examples are provided.

Sunday 13.12.2015 10:55 - 13:00 Parallel Session H – CFE-CMStatistics

CI022 Room Beveridge Hall SPECIAL SESSION ON MODELLING HETEROSKEDASTICITY Chair: Liudas Giraitis

CI0466: **Testing and modelling the unconditional variance component in multiplicative time-varying GARCH models**
Presenter: **Timo Terasvirta**, Aarhus University, Denmark
The topic is testing the hypothesis of constant unconditional variance in GARCH models against the alternative that the unconditional variance changes deterministically over time. Tests of this hypothesis have previously been performed as misspecification tests after fitting a GARCH model to the original series. It is found by simulation that the positive size distortion present in these tests is a function of the kurtosis of the GARCH process. Adjusting the size by numerical methods is considered. The possibility of testing the constancy of the unconditional variance before fitting a GARCH model to the data is discussed. The power of the ensuing test is vastly superior to that of the misspecification test and the size distortion minimal. The test has reasonable power already in very short time series. It would thus serve as a test of constant variance in conditional mean models. An application to exchange rate returns is included. It is also possible to turn the whole specification process around and specify the unconditional variance component before considering the GARCH component. Examples of this are given.

CI0613: **Testing for mean stability of heteroskedastic time series**
Presenter: **Liudas Giraitis**, Queen Mary University of London, United Kingdom
Co-authors: Violetta Dalla, Peter CB Phillips
Time series models are often fitted to the data without preliminary checks for stability of the mean and variance, conditions that may not hold in much economic and financial data, particularly over long periods. Ignoring such shifts may result in fitting models with spurious dynamics that lead to unsupported and controversial conclusions about time dependence, causality, and the effects of unanticipated shocks. In spite of what may seem as obvious differences between a time series of independent variates with changing variance and a stationary conditionally heteroskedastic (GARCH) process, such processes may be hard to distinguish in applied work using basic time series diagnostic tools. We develop and study some practical and easily implemented statistical procedures to test the mean and variance stability of uncorrelated and serially dependent time series. Application of the new methods to analyze the volatility properties of stock market returns leads to some unexpected surprising findings concerning the advantages of modeling time varying changes in unconditional variance.

CI1544: **A semiparametric multiplicative GARCH-X model: Adopting economic variables to explain volatility**
Presenter: **Dennis Kristensen**, University College London, United Kingdom
Co-authors: Heejoon Han
A multiplicative GARCH-X model is investigated which has a nonparametric long-run component induced by an exogenous covariate and a parametric GARCH short-run component. Compared to the usual additive GARCH-X model that includes an additional exogenous covariate in the GARCH specification, the covariate enters through a nonlinear function that is multiplied to the GARCH specification. When the covariate is nonstationary, i.e. integrated, near-integrated or fractionally integrated, the model can explain various stylized facts of financial time series. We propose a kernel-based estimation procedure for the parametric and nonparametric components and analyze its asymptotic properties. The asymptotic analysis is non-standard when the included covariates are non-stationary, and involves novel techiques for nonparametric and semiparametric estimation. An empirical application studies the linkage between US and European stock market volatilities using the VIX index as a covariate in our multiplicative GARCH-X model. It is shown that the model outperforms standard models both in terms of in-sample fitting and out-of-sample forecasting.

CO584 Room Holden EMPIRICAL MACROECONOMICS Chair: Daniel Kaufmann

CO0211: **Global or domestic: Which shocks drive inflation in European small open economies**
Presenter: **Jacek Kotlowski**, Warsaw School of Economics, Poland
Co-authors: Aleksandra Halka
The aim is to investigate which shocks drive inflation in small open economies. We proceed in two steps. Firstly, we use the SVAR approach to identify the global shocks. In the second step we regress the disaggregated price indices for selected European economies - the Czech Republic, Poland and Sweden - on the global shocks controlling for the domestic variables. Our results show that in two out of three analyzed countries the fluctuations of inflation are to the largest extent determined by the cyclical movements of the domestic output gap with the commodity shock being the second important source of inflation variability. We find that the direct impact of the global demand shock on the price dynamics is negligible, while it affects the country's inflation mainly through the domestic output gap. The role of the non-commodity global supply shock is less prominent, however, this shock, interpreted to some extent as a globalization shock, for most of the analyzed period lowers the prices of semi-durable and durable goods and therefore the inflation. Nonetheless, in the aftermath of the global financial crisis, this shock reversed what may be interpreted as a weakening of the globalization process.

CO0883: **Monetary policy, exchange rates and global inflation spillovers to Switzerland: An investigation with disaggregate data**
Presenter: **Matthias Gubler**, Swiss National Bank, Switzerland
Co-authors: Gregor Baeurle
We investigate the impact of global inflation on Swiss inflation relying on a structural dynamic factor model. The model relates a large set of disaggregate prices to the key domestic and global macroeconomic factors. Particular focus is placed on the role of monetary policy and the exchange rate in the pass-through of global inflationary pressures to Swiss inflation. First tentative results suggest that an increase in global trend inflation, i.e. an increase in global inflation which is not fully counteracted by foreign monetary policy, leads to tighter monetary policy in Switzerland. As a result, the Swiss franc appreciates, mitigating spillovers to Swiss inflation. However, if global inflationary pressures are temporary as foreign monetary policy becomes sufficiently restrictive, the pass-through to short-term Swiss inflation tends to be accompanied by a slight depreciation of the Swiss franc. Despite of that, long-run inflation in Switzerland is stabilized. While aggregate inflation responds qualitatively similar to both shocks, the different channels through which global trend and temporary inflation transmit impacts on the cross-sectional distribution of price changes. Specifically, the opposite exchange rate responses lead to opposite adjustments of exchange rate sensitive goods prices.

CO0403: **Tracking the global transmission dynamics of US monetary policy**
Presenter: **Jesus Crespo Cuaresma**, Vienna University of Economics and Business, Austria
Co-authors: Florian Huber, Martin Feldkircher, Gernot Doppelhofer
A new class of Bayesian Global Vector Autoregressive (BGVAR) models with drifting parameters is proposed to track the dynamic effects of US monetary policy on other economies over the last three decades. In addition to assessing global linkages explicitly, the model specification accounts for time-varying parameters and stochastic volatility in order to analyze the spillovers from US monetary policy on the global economy in a very flexible manner. Our results suggest that a contractionary shock to US monetary policy (50 basis points increase in short-term US interest rate)

leads to a persistent global contraction and a drop in global inflation rates, together with a rise in global interest rates, and a real depreciation of currencies with respect to the US dollar. We find evidence for important heterogeneity of the spillovers across countries and for changes in the transmission of monetary policy shocks over time.

CO0397: **Dealing with dynamic covariances in high-dimensional time series: A Bayesian approach**
Presenter: **Gregor Kastner**, WU Vienna University of Economics and Business, Austria
Dynamic covariance estimation for multivariate time series suffers from the curse of dimensionality; this renders parsimonious approaches essential for conducting reliable statistical inference. We address this issue by modeling the underlying dynamics of a time series vector through a lower dimensional collection of latent factors that allow for time-varying stochastic volatilities. Furthermore, we apply a Normal-Gamma prior to the elements of the factor loadings matrix. This hierarchical shrinkage prior is a generalization of the Bayesian lasso and effectively pulls the factor loadings of unimportant factors towards zero, thereby increasing sparsity even more. To guarantee efficiency of the estimation procedure, we employ a fully Bayesian yet computationally feasible approach to obtain draws from the high-dimensional posterior and predictive distributions via Markov chain Monte Carlo (MCMC) samplers. The latent vectors of time-varying volatilities are drawn "all without a loop" (AWOL), and we utilize several variants of an ancillarity-sufficiency interweaving strategy (ASIS) to boost efficiency when sampling the factor loadings as well as the parameters driving the time-varying volatilities. Additionally to being a stand-alone tool, the algorithm is designed to act as a "plug and play" extension for other MCMC samplers.

CC1395: **Trend fundamentals and exchange rate dynamics**
Presenter: **Florian Huber**, Oesterreichische Nationalbank, Austria
Co-authors: Daniel Kaufmann
We estimate a multivariate unobserved-components stochastic-volatility model to explain the behavior of a panel of six exchange rates against the US Dollar. The empirical model is based on the assumption that both countries monetary policy strategy may be well described by a Taylor rule with a time-varying inflation target and interest rate smoothing. This implies that the real exchange rate is determined by expectations about the relative trend inflation rates, the inflation gaps, the unemployment gaps and the short-term interest rates. The estimates closely track major movements along with important time series properties of the real and nominal exchange rate across all countries considered. The model generally outperforms a simple benchmark model that does not account for changes in trend inflation and trend unemployment.

CO641 Room Woburn MODELLING AND FORECASTING CYCLICAL FLUCTUATIONS II Chair: Josef Hollmayr

CO0254: **Idiosyncratic volatility puzzle: Influence of macro-finance factors**
Presenter: **Charlotte Christiansen**, Aarhus University, Denmark
Co-authors: Nektarios Aslanidis, Christos Savva, Neophytos Lambertides
The aim is to scrutinize the cross-sectional relation between idiosyncratic volatility and stock returns. As a novelty, the idiosyncratic volatility is obtained by conditioning upon macro-finance factors as well as upon traditional asset pricing factors. The macro-finance factors are constructed from a large pool of macroeconomic and financial variables. Cleaning for macro-finance effects reverses the puzzling negative relation between returns and idiosyncratic volatility documented previously. Portfolio analysis shows that the effects from macro-finance factors are economically strong. The relation between idiosyncratic volatility and returns does not vary with the NBER business cycles. The empirical results are highly robust.

CO0262: **A simple multivariate filter for estimating potential output**
Presenter: **Fan Zhang**, International Monetary Fund, United States
Co-authors: Roberto Garcia-Saltos, Patrick Blagrave
Estimates of potential output are an important ingredient of structured forecasting and policy analysis. Using the Philips curve and Okun's relationship, along with information from consensus forecasts to help tackle end of sample problem, it improves the conventional structured multivariate filter approach. We provide examples of how the methodology is applied and analyzed, by presenting analysis for one advanced economy and one emerging economy. We also applied the framework to a wide range of countries. However, we emphasize that, although the estimates in real time are more robust relative to those of statistical filters, there is still significant uncertainty around the estimates. We realize this is a simple filter, which only observes a few chosen economic signals; information from those macroeconomic indicators that are outside the scope of this filter should always be incorporated through experts' judgments.

CO0280: **Oversampling of stochastic processes**
Presenter: **Stephen Pollock**, University of Leicester, United Kingdom
Discrete-time ARMA processes can be placed in a one-to-one correspondence with a set of continuous-time processes that are bounded in frequency by the Nyquist value of π radians per sample period. It is well known that, if data are sampled from a continuous process of which the maximum frequency exceeds the Nyquist value, then there will be a problem of aliasing. However, if the sampling is too rapid, then other problems will arise that will cause the ARMA estimates to be severely biased.The nature of these problems are revealed and it is shown how they may be overcome.

CO0346: **Learning about banks' net worth and the slow recovery after the crisis**
Presenter: **Josef Hollmayr**, Deutsche Bundesbank, Germany
Co-authors: Michael Kuehl
Imperfect information about the net worth of banks and its consequences for the real economy is discussed. In a first part, we show empirically that expectations about the net earnings of banks (as growth of net worth) are truly biased during and in the aftermath of the financial crisis. This forecast error of professional investors cannot be attributed to information rigidities but to noisy information. This leads investors to follow a learning behavior about past forecast errors in forming their expectations about future earnings during the crisis. In a second part, by drawing on a New Keynesian general equilibrium model with a banking sector we demonstrate that incorporating this type of information updating and expectations formation about the net worth of banks can produce a slow recovery compared to a full information rational expectation case. We therefore argue that the slow recovery after the financial crisis in the US can be partly traced back to imperfect information about the net worth of banks.

CO0716: **An overview of the factor-augmented error-correction model**
Presenter: **Anindya Banerjee**, University of Birmingham, United Kingdom
Co-authors: Igor Masten, Massimiliano Marcellino
The Factor-augmented Error Correction Model (FECM) generalizes the factor-augmented VAR (FAVAR) and the Error Correction Model (ECM), combining error-correction, cointegration and dynamic factor models. It uses a larger set of variables compared to the ECM and incorporates the long-run information lacking from the FAVAR because of the latter's specification in differences. We review the specification and estimation of the FECM, and illustrate its use for forecasting and structural analysis by means of empirical applications based on Euro Area and US data.

CO382 Room Chancellor's Hall NEW DEVELOPMENTS IN FINANCIAL TIME SERIES ANALYSIS Chair: Yasuhiro Omori

CO0286: **Statistical estimation for optimal dividend barrier with insurance portfolio**
Presenter: **Hiroshi Shiraishi**, Keio University, Japan
Co-authors: Zudi Lu
A problem where an insurance portfolio is used to provide dividend income for the insurance company's shareholders is considered. This is an important problem in application of risk theory. Whenever the surplus attains the level barrier, the premium income is paid to the shareholders as dividends until the next claim occurs. We consider the classical compound Poisson model as the aggregate claims process, under which the dividends are paid to the shareholders according to a barrier strategy. Optimal dividend barrier is defined as the level of the barrier that maximizes the expectation of the discounted dividends until ruin, which was initially proposed time ago in the discrete time model and thereafter discussed in the classical risk model with the foundation laid. Although this problem was actively studied around 2000, it is less discussed with the estimation of the unknown "barrier". We propose the estimation problem of the optimal dividend barrier, which is critical in application. We establish an estimator of the expected discounted dividends and the estimated optimal dividend barrier as its maximizer, which are shown to be consistent. Numerical simulation experiments demonstrate that the proposed estimators work well with reasonable size of samples.

CO0548: **On the need of estimating the parameters of GARCH models**
Presenter: **Cindy Shin Huei Wang**, CORE and NTHU, Taiwan
A new estimating tool is proposed for a regression model with serial correlation and generalized autoregressive conditional heteroscedasticity (GARCH). This tool is constructed on the basis of the least absolute deviation (LAD) and the autoregressive (AR) approximation,namely the COAR-LAD estimation. The estimation is easy to implement and avoids issues commonly found in time series literature. Neither does the COAR-LAD estimation need to estimate many parameters of the GARCH models nor does it need to tackle specification problems one at a time. We further show that a convergent usual t-statistic based on our new estimator can be constructed for previously analyzed spurious regression cases, even though the exact form of the error term is unknown in practice. The simulation results indicate that the finite sample performance of our methodology is promising even in sample sizes. This finding sheds some new light on the nature of a regression with serial correlation and GARCH effect simultaneously.

CO0742: **Stock return predictability of the market variance risk premium in Japan**
Presenter: **Toshiaki Watanabe**, Hitotsubashi University, Japan
Co-authors: Masato Ubukata
The variance risk premium (VRP) is defined as the difference between the forecast of return variance (volatility) under the risk neutral measure and that under the physical measure. Recent empirical evidence suggests that the VRPs of stock indexes predict the excess return of those indexes but it is not true for the Nikkei 225 stock index in Japan. The stock return predictability of the VRP of the Nikkei 225 index is re-examined by predicting the volatility under the physical measure based on the monthly data on the index of industrial production (IIP) in Japan as well as the daily return of the Nikkei 225 index while only financial data is used in the previous literature. The monthly and daily data are combined by employing the GARCH-MIDAS model where the daily volatility is divided into the monthly component that depends on the past monthly data on the IIP and the daily component that follows a GARCH model. It is demonstrated that the new variance risk premium of the Nikkei 225 index predicts the excess return. It is also found that the movement of stock market volatility is counter-cyclical in Japan.

CO0974: **Bayesian restricted likelihood-based instrumental variables regression**
Presenter: **Catherine Forbes**, Monash University, Australia
Co-authors: Zhichao Liu
A Bayesian restricted likelihood framework has been suggested for obtaining posterior inference for structural parameters based on summary statistics rather than a complete sample of observations. An application of the Bayesian restricted likelihood approach will be explored as a robust method for handling outliers in an instrumental variables regression setting. The method retains the usual Bayesian advantage of enabling the incorporation of prior information, while in addition being relatively insensitive to outliers present in the reduced form equations. The approach will be demonstrated using data from a study regarding the effect of country openness on inflation rates.

CO0847: **Multivariate realized stochastic volatility with dynamic pairwise correlations**
Presenter: **Yasuhiro Omori**, University of Tokyo, Japan
Co-authors: Yuta Yamauchi
The multivariate realized stochastic volatility model with dynamic pairwise correlations is proposed. In modeling volatilities of the multivariate asset returns, the order of returns in the return vector is often determined arbitrary, which makes it difficult to interpret the model and its parameters. Such an arbitrary ordering assumption is avoided and instead to model each pairwise dynamic correlations is proposed. By adding the information of the realized variances and correlations to the measurement equations of the pairwise asset returns, we estimate the latent volatilities and correlations more accurately. It should be noted that we only need to observe a pair of intraday returns rather than all asset return simultaneously, and hence such a pairwise modeling does not suffer from non-synchronous trading problem of the multivariate asset returns. We take Bayesian approach and use Markov chain Monte Carlo simulation to estimate parameters as well as to guarantee the positive definiteness of the time-varying covariance matrices. The proposed model is applied to nine U.S. stock returns.

CO364 Room Court MEASURING SYSTEMIC RISK Chair: Monica Billio

CO0288: **An entropy-based early warning indicator for systemic risk**
Presenter: **Monica Billio**, University of Venice, Italy
Co-authors: Michele Costola, Roberto Casarin, Andrea Pasqualini
The purpose is the construction of an early warning indicator for systemic risk using entropy measures. The analysis is based on the cross-sectional distribution of marginal systemic risk measures such as Marginal Expected Shortfall, Delta CoVaR and network connectedness. These measures are conceived at a single institution for the financial industry in the Euro area. We estimate entropy on these measures and test if they show forecasting abilities in predicting banking crises. In this regard, we use a previously presented variable from European Central Bank. Entropy indicators show promising forecast abilities to predict financial and banking crisis. The proposed early warning signals reveal to be effective in forecasting financial distress conditions. In the analysis we consider Shannon, Tsallis and Renyi entropy measures and find that the values of the extensive index greater than one allow us to achieve better model fitting. This implies that entropy changes do to changes in the tail probability of the loss distribution are more important than changes in the common event probabilities in order to detect periods of high systemic risk level.

CO0304: **The impact of network connectivity on factor exposures, asset pricing and portfolio diversification**
Presenter: **Roberto Panzica**, Goethe University House of finance, Italy
Co-authors: Monica Billio, Massimiliano Caporin, Loriana Pelizzon
The need for understanding the propagation mechanisms behind the recent financial crises leads the increased interest for works associated with

asset interconnections. In this framework, network-based methods have been used to infer from data the linkages between institutions. In turn, those connections have implications for the evaluation of systemic risk. We elaborate on this and make a step forward by introducing network linkages into linear factor models, thus allowing for the interdependence between asset connections and systematic risks. Networks are used to infer the exogenous and contemporaneous links across assets, and impacts on several dimensions. From a factor exposure perspective, network links act as inflating factor for systematic exposure to common factors, and allow for cross asset exposures to factors due to the presence of the network. In turn, the presence of networks and factors, has potential implications for pricing. Furthermore, the power of diversification is reduced by the presence of network connections, and we analytically show that network links reduce the diversification potential but at the same time could allow for absorption of risks. By fitting a (misspecified) linear factor model under our data generating process (allowing for the presence of network links), the model provides residuals which are correlated and heteroskedastic, and the factor exposures become time varying. We support our claims with an extensive simulation experiment.

CO0301: **Where the risks lie: A survey on systemic risk**
Presenter: **Christophe Hurlin**, University of Orleans, France
The purpose is to review the extensive literature on systemic risk and connect it to the current regulatory debate. While we take stock of the achievements of this rapidly growing field, we identify a gap between two main approaches. The first one studies different sources of systemic risk in isolation, uses confidential data, and inspires targeted but complex regulatory tools. The second approach uses market data to produce global measures which are not directly connected to any particular theory, but could support a more efficient regulation. Bridging this gap will require encompassing theoretical models and improved data disclosure.

CO0338: **Modeling contagion and systemic risk**
Presenter: **Daniele Bianchi**, University of Warwick, United Kingdom
Co-authors: Monica Billio, Roberto Casarin, Massimo Guidolin
The aim is to model contagion in financial markets as a shift in the strength of cross-firm network linkages, and argue that this provides a natural and intuitive framework to measure systemic risk. We take an asset pricing perspective and dynamically infer the network structure system-wide from the residuals of an otherwise standard linear factor pricing model, where systematic and systemic risks are jointly considered. We apply the model to a large set of daily returns on blue chip companies, and find that systemic risk increased across the period 2001/2002 (i.e. dot.com bubble, 9/11 attacks, financial scandals, Iraq war), the great financial crisis, and the recent major Eurozone sovereign turmoil. Our results show that financial firms are key for systemic risk management, and such network centrality does not depend on market values. In addition, the empirical evidence suggests that those institutions with higher network centrality experience larger financial losses. Consistent with this evidence, our model-implied systemic risk measure provides an early warning signal on aggregate financial distress conditions and is positively correlated with the business cycle.

CO1180: **Financial crises, crises spillovers and the business cycle**
Presenter: **Stefan Straetmans**, Maastricht University, Netherlands
Financial crises seem to occur across different macroeconomic regimes (e.g. recessions as well as expansions; low inflation regimes vs. high inflation regimes etc.). The objective is to quantify the likelihood of financial crises and crisis spill-overs across macroeconomic regimes in order to assess whether and to what extent economic recession episodes are more inclined towards financial crises and crisis comovements than expansion periods. Statistical extreme value analysis (EVT) is put at work to calculate these marginal and joint tail likelihoods for recession and expansion subsamples. We find that tail risk is procyclical for different types of financial assets. Also, systemic risk indicators based on extreme co-movements between bank stocks are found to be procyclical which confirms earlier research on market-based systemic risk measures. Moreover, cross-asset crisis spillovers like flight-to-quality effects between stocks, bonds or gold become much more pronounced during recessions. Finally, we show that diversifying portfolio tail risk becomes more difficult during recessions. To our knowledge, applying EVT techniques to business cycle regimes (or other economically meaningful sample partitions) is novel to the literature on financial extremes and extreme value analysis. EVT measures can also be made dependent on multiple regimes and regime determination can be made endogenous.

CO576 Room Bedford NONSTATIONARY TIME SERIES AND PANELS Chair: Stephan Smeekes

CO0315: **Semiparametrically optimal hybrid rank tests for unit roots**
Presenter: **Ramon Van den Akker**, Tilburg University, Netherlands
Co-authors: Bo Zhou, Bas Werker
The aim is to propose a new class of unit root tests that exploits invariance properties in the Locally Asymptotically Brownian Functional limiting experiment of the unit root model. These invariance structures naturally suggest tests based on the ranks of the increments of the observations, their mean, and an assumed reference density for the innovations. The tests are semiparametric in the sense that the reference density need not equal the true innovation density. For correctly specified reference density, the asymptotic power curve of our test is point-optimal and nearly efficient. When using a Gaussian reference density, our test performs as well as commonly used tests under true Gaussian innovations and better under other distributions, e.g., fat-tailed or skewed. Monte Carlo evidence shows that our test also behaves well in small samples.

CO0636: **Asymptotic inference in the Lee-Carter model for modelling mortality rates**
Presenter: **Simon Reese**, Lund University, Sweden
The most popular approach to modelling and forecasting mortality rates is the Lee-Carter model. The popularity of the model rests mainly on its good fit to the data, its theoretical properties being obscure. Asymptotic results for the Lee-Carter model are provided its inherent weaknesses is illustrated formally. Requirements on the underlying data are established and variance estimators are presented in order to allow hypothesis testing and the computation of confidence intervals.

CO1534: **A comparative study on forecasting performance of high-dimensional time series methods**
Presenter: **Etienne Wijler**, Maastricht University, Netherlands
Co-authors: Stephan Smeekes
We compare the predictive capabilities in a time series context of several high-dimensional methods that perform subset selection, which are broadly categorized into methods using derived inputs, shrinkage estimators and hybrid methods. Comparative forecasting results obtained on simulated as well as empirical data indicate that the forecasting performance of shrinkage estimators is fairly robust against alternative DGP specifications compared with methods using derived inputs and hybrid methods, whose performance deteriorates substantially under sub-optimal conditions such as the absence of latent factors or the presence of non-sphericity. Additionally, we investigate the performance of these methods in the presence of $I(1)$ variables, with an additional focus on the variable selection properties of penalized regression methods in a cointegration framework.

CO1709: **Nonlinear heterogeneity in cointegrated panels**
Presenter: **Jean-Pierre Urbain**, Maastricht University SBE, Netherlands
Co-authors: Peter Pedroni
A method is developed for exploiting cross sectional heterogeneity in cointegrated panels to uncover possible nonlinear relationships. In particular,

individual specific slope estimates from panel regressions are implicitly interacted with the regressors observed at a point in time, as well as with unit specific determinants of the underlying nonlinearities, via a twostage estimation procedure. We first show how the method can be used to obtain estimates of quadratic forms that are common among the units of the panel based on relationships between slope estimates and point in time observations of the regressors. Next, we show how unit specific determinants of cross sectional heterogeneity can be exploited to generalize the approach to obtain estimates of quadratic forms that are idiosyncratic to individual units of the panel. These estimates can in turn be used to evaluate the position of a cross sectional unit relative to the maximum or minimum ofa unit specific quadratic form. The methods are sufficiently general to account for both temporal and spatial dependencies, and to allow for endogeneity. We provide Monte Carlo simulation evidence for the small sample performance, and provide a brief illustrationin support of the broad empirical applicability of the technique.

CC1699: **Testing for panel cointegration using common correlated effects estimators**
Presenter: **Josep Lluis Carrion-i-Silvestre**, Universitat de Barcelona, Spain
Co-authors: Anindya Banerjee
Spurious regression analysis in panel data when the time series are cross-section dependent is performed. We show that consistent estimation of the long-run average parameter is possible once we control for cross-section dependence using cross-section averages in the spirit of a previous common correlated effects approach. This result is used to design a panel cointegration test statistic accounting for cross-section dependence. The performance of the proposal is investigated in comparison with factor-based methods to control for cross-section dependence when strong, semi-weak and weak cross-section dependence may be present.

CO408 Room SH349 MONITORING AND TRACKING DEPENDENCE Chair: Oliver Grothe

CO0334: **Detecting relevant changes in time series models**
Presenter: **Dominik Wied**, TU Dortmund, Germany
Co-authors: Holger Dette
Most of the literature on change point analysis by means of hypothesis testing considers hypotheses of the form "Is there a change or not?" A different perspective is taken, i.e., the null hypothesis of no relevant changes of the form "Is there a small change or not?" is considered. A general approach to problems of this type is developed which is based on the cumulative sum principle. For the asymptotic analysis weak convergence of the sequential empirical process must be established under the alternative of non-stationarity, and it is shown that the resulting test statistic is asymptotically normally distributed. Applications of the methodology include the mean and the covariance matrix. The finite sample properties of the new tests are investigated by means of a simulation study and illustrated by analyzing a data example from portfolio management.

CC1111: **Testing for Cojumps: A multivariate coexceedance-based approach**
Presenter: **Hans Manner**, University of Cologne, Germany
Co-authors: Markus Koesler
We consider the problem of testing the synchronicity of jumps, i.e., the presence of cojumps, using high frequency financial data. This is done in the framework of multivariate log price processes with stochastic volatility and compound Poisson jump processes. Testing for cojumps requires a univariate jump detection procedure that tests for and locates jumps. To this end we apply an approach that relies on univariate returns standardized by a jump-robust local volatility estimator. We propose several test statistics for the null hypothesis of independent jump processes against the alternative of dependent jump processes. The test statistics compare the number of coexceedances with their expected number under the null of independent jump processes. We derive the asymptotic distributions of these test statistics under double asymptotics, increasing both the observation frequency and the length of the sample. For finite samples we suggest a block bootstrap procedure that can mitigate size distortions of some of the test statistics. Furthermore, we extend the tests to the context of non-homogeneous jump occurrences using Hawkes processes to model the jump intensities. A comprehensive Monte Carlo study examines the finite sample properties of the tests in a realistic market scenario and an application to financial high-frequency data illustrates its practical use.

CC1119: **D-Vines estimation for mixed data using penalized splines**
Presenter: **Christian Schellhase**, Bielefeld University, Germany
Co-authors: Goeran Kauermann, Joachim Schnurbus
The estimation of vines for mixed data using penalized B-splines is presented, as we extend the estimation of D-vines and the pair-copula construction to support also ordinary variables. The estimation of probability mass functions in the context of copulas is easily done using constant B-splines, whose nodes are placed at the jump points of the empirical cumulative distribution function of some ordinary data. The independence of the constant B-splines from each other ensures the applicability. We illustrate the potential gains of D-vine estimation for mixed data with penalized splines, as we analyse the rent survey of Munich, analysing whether the rent increase for German major cities is merely a matter of rising demand that is exploited by flat owners. The copula-based analysis of several current releases of the Munich rent survey allows to disentangle the rent increase over time into two effects. First, the rent increase caused by an improvement of the flats and second, the increase in terms of equivalent flats simply getting more expensive.

CO0961: **Assessing cross-sectoral spillover potential among banks, shadow banks and insurance companies**
Presenter: **Cosimo Pancaro**, European Central Bank, Germany
Co-authors: Dawid Zochowski, Marco Gross
An analytical framework is developed that aims at assessing the evolution of the spillover potential within and across traditional (regulated) banks, shadow banks, and insurance companies over time. This analytical framework is based on a non-parametric variant of delta-Conditional Expected Shortfall (dCoES) introduced previously. The analysis is carried out using a database of Expected Default Frequencies (EDFs) at daily frequency provided by Moodys KMV for a significant sample of about 2,000 financial institutions, from 20 EU countries, covering the period between 1 Jan 2007 and 28 Feb 2015. The results suggest that there are mainly two periods during which the spillover potential increased materially: 2008Q4-2009Q2 and 2011Q3-2012Q3. These two periods correspond to i) the initial eruption of the global financial crisis following the Lehman Brothers failure and ii) the peak of the European sovereign debt crisis. Finally, the results show that, according to the measure of cross-sectoral spillover risk presented in this analysis, the shadow banking system stands out as the comparably most vulnerable part of the financial system.

CO0292: **Modeling systemic risk: Time-varying tail dependence when forecasting marginal expected shortfall**
Presenter: **Tobias Eckernkemper**, University of Cologne, Germany
A framework for modeling the Marginal Expected Shortfall has been recently proposed. It uses the conditional correlations of a DCC-GARCH model to drive dependencies between market's and institution's innovations in a factor model and explicitly allow for additional, non-linear types of dependence (like tail dependence). While this approach is very flexible and performs well in different applications, it vitally assumes that the non-linear dependence is constant over time, which was found to be not the case in a number of studies. Therefore, a copula-based alternative is proposed which embeds main elements of that approach but accounts for variations in linear as well as non-linear dependence structures. In particular, dependencies are generated by a dynamic mixed copula model combined with a previous Generalized Autoregressive Score model. In an empirical application on the Dow Jones U.S. Financial Index, it is confirmed that non-linear dependencies indeed change over time and that their impact on systemic risk is essential and captured by the proposed model.

CO376 Room Bloomsbury QUANTILE REGRESSION IN FINANCE AND ECONOMICS Chair: Massimiliano Caporin

CO0337: **Oil returns conditional quantiles and uncertainty indexes: Causality and forecasting implications**
Presenter: **Massimiliano Caporin**, University of Padova, Italy
Co-authors: Giovanni Bonaccolto, Rangan Gupta
Uncertainty indexes have been extensively analysed to determine their relation with macroeconomic and financial variables. We provide further evidences on the impact of uncertainty on oil price returns from a novel perspective, based on quantile regression. We first evaluate the presence of causality relations in the mean, in the variance and at quantiles by using recent contributions in the area. Given the presence of causality relation from uncertainty to oil, we exploit this spillover within a forecasting exercise. Contrasting results at different frequencies (monthly versus daily) and over different samples (by using rolling methods), we show that both the causality relation and the forecasting improvements change over time.

CO0431: **Asset allocation strategies based on penalized quantile regression**
Presenter: **Giovanni Bonaccolto**, University of Padova, Italy
Co-authors: Massimiliano Caporin, Sandra Paterlini
It is well known that quantile regression model minimizes the portfolio extreme risk, whenever the attention is placed on the estimation of the response variable left quantiles. We show that, by considering the entire support of the response variable conditional distribution, and not just its left tail, it is possible to optimize a set of different risk and performance indicators. In particular, we introduce a risk-adjusted profitability measure, useful in evaluating financial portfolios under a pessimistic perspective, since the reward contribution is net of the most favorable outcomes. Moreover, as we consider large portfolios, we also cope with the dimensionality issue by introducing an $l1$-norm penalty on the assets weights, which improves the out-of-sample performance of the allocation strategies.

CO0552: **A Quantile-Boosting Approach to Forecasting Gold Returns**
Presenter: **Marian Risse**, Helmut Schmidt University, Germany
Co-authors: Christian Pierdzioch, Sebastian Rohloff
We use a quantile-boosting approach to compute out-of-sample forecasts of gold returns. The approach accounts for model uncertainty and model instability, and it allows forecasts to be computed under asymmetric loss functions. Different asymmetric loss functions represent different types of investors (optimists versus pessimists). We document how the performance of a simple trading rule varies across investor types.

CC1028: **Estimating daily financial quantiles by means of high-frequency data: A scaling method approach**
Presenter: **Timo Dimitriadis**, University of Konstanz, Germany
Co-authors: Roxana Halbleib
We introduce a new method for estimating quantiles of daily financial returns by using high-frequency data. The method is based on a scaling relationship between quantiles of daily returns and quantiles of intraday returns stemming from the assumption that the returns follow an unifractal process. We provide empirical evidence that estimates and forecasts of daily quantiles derived from scaling-up relationships of high-frequency information are more accurate than those derived from classical approaches relying on daily information within a Value at Risk application.

CO0934: **Revisiting the public debt-growth relationship using threshold quantile regression**
Presenter: **Luis Filipe Martins**, ISCTE-IUL, Portugal
The soaring public debt-to-GDP ratios in several countries of OECD, especially after the recent financial crisis, is one of the main concerns for policy makers as it jeopardizes economic growth and social welfare. As a result, authorities adopt corresponding countermeasures to bring debt to sustainable levels (eg, the Greek rescue deal). A substantial body of this literature has been proposed to modeling and testing for a threshold effect in the link between debt and real GDP growth but apparently it has not reached a broad consensus. The well-known 90% debt level above which economic growth slows down and default risks rise surely depends on the economy under study and the database used but also on the chosen methodology. We estimate and test for the existence of threshold effects in the relationship between debt and growth by means of threshold quantile regression. Thus, we look at the whole conditional distribution of growth instead of the mean quadratic regression. Looking at country-specific cases, we use the updated Reinhart-Rogoff dataset and extend it by adding a few number of covariates. We present methods for estimation and testing that we believe to be new in the time-series context.

CO504 Room Torrington REGIME SWITCHING, FILTERING, AND PORTFOLIO OPTIMIZATION Chair: Joern Sass

CO0386: **Dividend maximization under changing economic environment and partial information**
Presenter: **Michaela Szoelgyenyi**, Vienna University of Economics and Business, Austria
The value of an insurance company can be defined as the maximal expected discounted dividend payments that can be paid out during the lifetime of the company. We follow this approach to solve the valuation problem of an insurance company. Extending classical contributions we study the optimization problem in a framework that allows for shifts of the economical environment, which is reasonable due to the usual long time horizon of the observation. Furthermore, we assume the current economical state to be unobservable. Specifically, we model the surplus of the insurance company as a diffusion process with an unobservable drift parameter that might shift. This results in a joint filtering and stochastic optimization problem. After applying filtering theory to overcome uncertainty, we are able to characterize the solution of the optimization problem as the unique viscosity solution to the associated Hamilton-Jacobi-Bellman equation. A numerical treatment of the problem leads to dividend strategies of threshold type, for which we finally prove admissibility.

CO0475: **Generalized dynamic factor models: An estimation approach via autoregressive processes**
Presenter: **Julia Reynolds**, Vienna Graduate School of Finance, Austria
Co-authors: Manfred Deistler, Leopold Soegner
Due to their unique ability to reconcile dynamic properties with approximate factor structures, generalized dynamic factor models (GDFMs) have become a popular tool with which to examine large datasets. We examine an estimation procedure for determining the number of dynamic factors in a GDFM framework in which the latent factors are modeled as autoregressive processes. The procedure uses static PCA to estimate the static factors, and autoregressive parameters are estimated using Yule-Walker equations. The number of dynamic factors is then estimated as the rank of the covariance matrix of the autoregressive system. We examine the robustness of the estimation procedure using simulated data, and finally apply it to a financial dataset in order to determine the number of dynamic factors driving a panel of liquidity measures.

CO0476: **The consistency of estimators of the dimension of factors: Box-counting, local and quantization estimators**
Presenter: **Klaus Poetzelberger**, WU Vienna, Austria
Establishing the dimension or an estimate of the dimension of a probability distribution or of the distribution of factors driving a stochastic dynamical system is often the first step for a statistical analysis or a more general decision problem. We present consistency results for estimators derived from the concepts of the box-counting dimension, the local dimension and the quantization dimension of the support of probability distributions.

All these estimators are typically highly biased. The estimation is simpler if the dimension to be estimated is known to be an integer. Instead of the mean squared error, the probability/proportion of false classification is the criterion for the performance of the estimator. Consistency results exist for the distributions of innovations of Ito processes and allow the construction of consistent estimators of the dimension of the factors, i.e. of the dimension of the Brownian motion driving the process. Possible applications are the optimization of portfolios with nonlinear dependencies.

CO0538: **A Bayesian methodology for systemic risk assessment in financial networks**
Presenter: **Luitgard Veraart**, London School of Economics, United Kingdom
Co-authors: Axel Gandy
A Bayesian methodology for systemic risk assessment in financial networks such as the interbank market is developed. Nodes represent participants in the network and weighted directed edges represent liabilities. Often, for every participant, only the total liabilities and total assets within this network are observable. However, systemic risk assessment needs the individual liabilities. We propose a model for the individual liabilities, which, following a Bayesian approach, we then condition on the observed total liabilities and assets. We construct a Gibbs sampler to generate samples from this conditional distribution. These samples can be used in stress testing, giving probabilities for the outcomes of interest. As one application we derive default probabilities of individual banks and discuss their sensitivity with respect to prior information included to model the network. An R-package implementing the methodology is provided.

CO0950: **Investment strategies within a regime switching model for asset returns**
Presenter: **Christina Erlwein-Sayer**, Fraunhofer Institute for Industrial Mathematics ITWM, Germany
Co-authors: Stefanie Grimm, Peter Ruckdeschel, Joern Sass, Tilman Sayer
The proposed asset allocation model allows for regime shifts in parameters of asset returns. In this hidden Markov model (HMM) framework, parameters of a discretized geometric Brownian motion are guided by a Markov chain in discrete time. The HMM is a flexible model setting in which switching market conditions or behavioural aspects of market participants can be captured. These changing conditions are introduced through switching volatilities and drifts into the asset price processes. We consider a multivariate observation process with and without correlation. The model parameters are estimated through a filter-based EM-algorithm. Adaptive filters for the Markov chain and related quantities are derived and used to find optimal parameter estimates of the model. The estimated densities of asset returns are utilized for investment strategies, the on-line parameter estimates obtained are used to make investment decisions within a regime-switching framework. A Markowitz-type investment strategy is examined and compared to pure and mixed asset allocation strategies. The obtained portfolios realistically handle regime-shifts in markets.

CO358 Room Gordon NUMERICAL METHODS AND ESTIMATION OF DSGE MODELS Chair: Martin M Andreasen

CO0411: **A theory of pruning**
Presenter: **Giovanni Lombardo**, BIS, Switzerland
Co-authors: Harald Uhlig
Often, numerical simulations for dynamic, stochastic models in economics are needed. Higher order methods can be attractive, but bear the danger of generating explosive solutions in originally stationary models. Previously, it has been proposed pruning to deal with this challenge for second order approximations. We provide a theory of pruning and formulas for pruning of any order. We relate it to results on perturbing dynamical systems.

CO0445: **Identification of DSGE models: The effect of higher-order approximation and pruning**
Presenter: **Willi Mutschler**, TU Dortmund, Germany
The aim is to show how to check rank criteria for a local identification of nonlinear DSGE models, given higher-order approximations and pruning. This approach imposes additional restrictions on (higher-order) moments and polyspectra, which can be used to identify parameters that are unidentified in a first-order approximation. The identification procedures are demonstrated by means of previous models. Both models are identifiable with a second-order approximation. Furthermore, analytical derivatives of unconditional moments, cumulants and corresponding polyspectra up to fourth order are derived for the pruned state-space.

CO0547: **Higher-order dynamics in asset-pricing models with recursive preferences**
Presenter: **Karl Schmedders**, University Zurich, Switzerland
Co-authors: Ole Wilms, Walter Pohl
An analysis is presented of the higher-order dynamics of key financial quantities in asset-pricing models with recursive preferences. For this purpose, we first describe a projection-based algorithm for solving such models. The method outperforms common methods like discretization and log-linearization in terms of efficiency and accuracy. Our algorithm allows us to document the presence of strong nonlinear effects in the modern long-run risks models which cannot be captured by the common methods. For example, for a prominent recent calibration of a popular long-run risks model, the log-linearization approach overstates the equity premium by 100 basis points or 22.5%. The increasing complexity of state-of-the-art asset-pricing models leads to complex nonlinear equilibrium functions with considerable curvature which in turn have sizable economic implications. Therefore, these models require numerical solution methods, such as the presented projection methods, that can adequately describe the higher-order equilibrium features.

CO0473: **Solving second and third-order approximations to DSGE models: A recursive Sylvester equation solution**
Presenter: **Andrew Binning**, Norges Bank, Norway
The matrix chain rules are derived for solving a second and a third-order approximation to a DSGE model that allow the use of a recursive Sylvester equation solution method. In particular we use previous solution algorithms to solve the generalised Sylvester equations. Because we use matrix algebra instead of tensor notation to find the system of equations, we are able to provide standalone Matlab routines that make it feasible to solve a medium scale DSGE model in a competitive time. We also provide Fortran code and Matlab/Fortran mex files for this method.

CO1266: **Short run consumption risks in the long run risk model**
Presenter: **Kasper Jorgensen**, Aarhus University, Denmark
Co-authors: Martin M Andreasen
We decompose consumption into (1) a stationary but persistent component, and (2) a stochastic trend with a small long-run predictable component. The representative agent has recursive preference and a flexible utility kernel. Our model can justify the equity premium, the risk free rate and a large part of the volatility of the market return, risk free rate and price-dividend ratio with relevant behavioral characteristics. In particular, intertemporal elasticity of substitution is less than one and risk aversion is less than ten.

CO360 Room Montague ESTIMATION AND INFERENCE IN LONG MEMORY PROCESSES Chair: Josu Arteche

CO0482: **Bias-free estimation of fractional integrated panel data models**
Presenter: **Carlos Velasco**, Universidad Carlos III de Madrid, Spain
Parameter estimation in a dynamic panel data model with individual fixed effects and long memory is considered. This model generalizes autoregressive specifications and can describe more general forms of nonstationarity beyond unit roots. Methods to remove fixed effects in this framework can introduce bias in estimates when the time series is short compared to the cross section dimension and can even be inconsistent due to an initial conditions problem. We explore new GMM estimates based on moment conditions defined on first difference data with readjusted initial conditions that can be used with minimal assumptions on the dimensions of the panel. The new estimates can achieve the same efficiency as the PMLE estimates based on the same differenced data when the number of moments increase with the time series dimension T and inference can be simpler and more robust than that based on ML estimation for a fixed T panel or without assumptions on the panel dimension.

CO0527: **Testing the fractionally-integrated hypothesis with $k-$step $M-$estimators**
Presenter: **Paulo Rodrigues**, Universidade Nova de Lisboa and Bank of Portugal, Portugal
Co-authors: Antonio Rubia, Matei Demetrescu
Normality is a common assumption in time series analysis and in econometrics in general. Characteristically, regression based tests for long-memory have been based on least squares (LS) estimation, achieving optimal properties only under the assumption of Gaussianity. In practice, however, many observable variables are driven by heavy-tailed distributions and often exhibit large and sudden movements similar to additive outliers; leading examples are financial time series. We devise tests for fractional integration based on iterated $M-$estimators, which are suitable for this non-normal environment. Large sample results as well as an in-depth Monte Carlo analysis of the performance of the tests are provided. It is shown that the tests display very interesting properties and finite sample performance.

CO0307: **Robust spectral estimators for long-memory processes**
Presenter: **Valderio Anselmo Reisen**, DEST-CCE-UFES, Brazil
The outlier effects on the estimation of a spectral estimator is discussed for long memory process under additive outliers and robust spectral estimators are proposed. Some asymptotic properties of the proposed robust methods are derived and Monte Carlo simulations investigate their empirical properties. Pollution series, such as PM (Particulate matter), SO2 (Sulfur dioxide), are the examples investigated to show the usefulness of the robust methods in real applications. These pollutants present, in general, observations with high levels of pollutant concentrations which may produce sample densities with heavy tails and these high levels can be identified as outliers which can destroy the statistical properties of sample functions such as the standard mean, covariance and periodogram.

CO1636: **Generalized fractional stochastic volatility models**
Presenter: **Shelton Peiris**, University of Sydney, Australia
Co-authors: Manabu Asai
In recent years fractionally differenced processes have received a great deal of attention due its flexibility in financial applications with long memory. The aim is to consider a generalized class of fractionally differenced processes generated by Gegenbauer polynomials, known as Gegenbauer Autoregressive Moving Average (GARMA) models. We extend this class to incorporate both the long memory and stochastic volatility (SV) components. The existence and uniqueness of second order solutions will be established. Under certain regularity conditions, it is shown that the estimation of parameters, identification and forecasting through the underlying process are not affected. Furthermore, various new results associated with the moments, kurtosis and minimum mean square error forecasts will be reported. A simulation study has been added and a potential application will be discussed to justify the usefulness of this new class of models in financial econometrics.

CO0369: **Optimal bandwidth selection in local Whittle estimation**
Presenter: **Josu Arteche**, University of the Basque Country, Spain
The local Whittle estimator is one of the most popular techniques for estimating the memory parameter in long memory series due to its simple implementation and nice asymptotic properties under mild conditions. However, its empirical performance depends heavily on the band of frequencies used in the estimation, known as bandwidth. Different choices may lead to different conclusions about, for example, the stationarity of the series or its mean reversion. Optimal bandwidth selection is thus of crucial importance for accurate estimation of the memory parameter, but few strategies for assuring this have been proposed to date, and their results in applied contexts are poor. A new strategy based on minimising a bootstrap approximation of the mean square error is proposed and its performance is shown to be convincing in an extensive Monte Carlo analysis and in applications to real series.

CO448 Room Jessel MACROECONOMIC UNCERTAINTY AND POLICY Chair: Wojciech Charemza

CO0589: **When information on forecast uncertainty improves the performance of a combined forecast**
Presenter: **Christian Conrad**, Heidelberg University, Germany
Forecast surveys such as the Survey of Professional Forecasters (SPF) provide cross sections of density forecasts for macroeconomic variables such as output growth, inflation or unemployment. Two widely employed measures of ex-ante uncertainty associated with such forecasts are the average across individual uncertainties on the one hand and the cross sectional dispersion of point forecasts ("disagreement") on the other hand. We compare the informative content of these statistics and ask under which circumstances they can be used to reduce the mean squared error of a combined forecast. We show that forecast precision can be enhanced if some forecasters' objectives are characterized by the asymmetric Linex-loss function. Then, average individual uncertainty provides the most predictive content. Moreover, we provide empirical evidence for these considerations based on data from the SPF.

CO0607: **Quantitative easing and tapering uncertainty: Evidence from Twitter**
Presenter: **Annette Meinusch**, Justus-Liebig University Giessen, Germany
The aim is to quantify the effects of changes in people's beliefs about the timing of the exit from Quantitative Easing ("tapering") on asset prices. To quantify beliefs of market participants, we use data from Twitter, the social media application, covering the entire Twitter volume on Federal Reserve tapering in 2013. Based on the time series of beliefs about an early tapering, we estimate a VAR model with appropriate sign restrictions on the impulse responses to identify a belief shock. The results show that shocks to tapering beliefs have strong and robust effects on interest rates, exchange rates and asset prices. We use social media data for the first time for analyzing monetary policy and add to the rapidly emerging literature on macroeconomic uncertainty shocks.

CO0727: **Understanding the role of uncertainty in the Euro area business cycle**
Presenter: **Geoff Kenny**, European Central Bank, Germany
Co-authors: Geoff Kenny
The aim is to examine the role of different indicators of uncertainty using the probability forecasts provided in the Euro area SPF. The properties of

different aggregate uncertainty measures are compared also with reference to the different proxies of uncertainty commonly used in the literature. We also use a VAR to compare the business cycle impact of uncertainty shocks derived from different variables (GDP, inflation, the unemployment rate) and from different forecasting horizons over which uncertainty is measured. The relevance of uncertainty shocks during the Great Recession and subsequent sovereign debt crisis is then assessed.

CO0831: **Measuring economic uncertainty using news-media textual data**
Presenter: **Peter Eckley**, Bank of England, United Kingdom
We develop a news-media textual measure of aggregate economic uncertainty - the fraction of Financial Times articles containing uncertainty-related keyphrases - for 1982-2014 at daily to annual frequencies. In our theory of measurement, uncertainty references arise from a latent propensity to express uncertainty, closely linked to the cognitive state involved in economic decision-making. We show a single newspaper contains enough articles for acceptable signal-to-noise ratios. Our empirical implementation improves on similar existing measures. First, we reveal extensive and irregular duplication of articles in the leading news database, and provide a simple de-duplication algorithm. Second, we boost the signal by 14% by adding the word uncertainties to uncertain and uncertainty, and show that further uncertainty-related keyphrases would constitute a second-order adjustment. Third, we demonstrate the importance of normalising article counts by total news volume and provide the first UK textual uncertainty measure to do so. We establish the plausibility of our uncertainty proxy through a detailed narrative analysis and comparative analysis with another popular uncertainty proxy, stock returns volatility. There is a strong relationship on average, but it breaks down periodically. We offer plausible explanations for this behaviour. Finally, we establish the absence of Granger causation between the measures, even down to daily frequency.

CO0876: **ECB footprints on the inflation forecast uncertainty**
Presenter: **Svetlana Makarova**, University College London, United Kingdom
The main aim is to evaluate the hypothesis that the monetary policy of the European Central Bank leads to a decrease in macroeconomic uncertainty across Euro area. Macroeconomic uncertainty has been measured by the root mean squared pseudo-ex-post errors of inflation forecasts net of ARCH-GARCH effects. A test has been proposed based on regressing the rate of growth of the uncertainty ratio (that is the ratio of the uncertainty measure free from the policy effects to the gross measure) on the initial conditions, in a way analogous to the panel data beta-convergence growth models. Results obtained with the use of monthly data for 16 countries for the period from January 1997 to November 2014 and with forecast horizons from 1 to 18 show evidence of the effectiveness of the ECB uncertainty reduction policy for the Euro countries.

CO516 Room Athlone RISK AND VOLATILITY MODELLING Chair: Giuseppe Storti

CO0615: **The effect of additive outliers on fractional unit root tests**
Presenter: **Arie Preminger**, Yezreel Valley College, Israel
Co-authors: Christian Hafner
It is well known that additive outliers that occur with a small probability have a bias effect on the asymptotic distribution of classical unit root statistics. It is shown that such outliers do not affect the asymptotic distribution in the case where the error term is fractionally integrated of order d, where $0 < d < 1/2$, while there is a bias for the case $1/2 < d < 1$. Convergence to the asymptotic distribution is slow, such that the bias effect of outliers may be important in finite samples, for which numerical evidence is provided. We then show that these results essentially do not change if the unknown d is replaced by a consistent estimator, which may have a slow rate of convergence. Such an estimator can be obtained by using an outlier correction procedure. We finally apply our results to a realized volatility series of the S&P 500 for which we find evidence against the unit root hypothesis, as opposed to a procedure which neglects outliers.

CO0776: **Whittle estimation of multivariate exponential volatility models with long memory**
Presenter: **Malvina Marchese**, University of Genoa, Italy
Co-authors: Paolo Zaffaroni
The strong consistency and the asymptotic normality of the Whittle estimate of the parameters in a class of multivariate exponential volatility processes are established. The class includes the multivariate Stochastic Volatility model with Leverage and the Constant Conditional Correlation model with EGARCH individual volatilities. Under the general (MEV) model the logarithm of the squared returns is decomposed into the sum of a signal vector-linear process and a white noise. We allow for correlation between the signal and the noise since it arises in the CCC model and in the MSV model as a consequence of leverage. We allow for a wide degree of persistence of shocks to the conditional variance, including both short and long memory parametrization of the signal process. We assess the small sample properties of the estimator by means of a Monte Carlo exercise. We present an empirical application and we discuss diagnostics to evaluate the appropriateness of the multivariate exponential specification.

CO0913: **NonparaSkew graphical models for systemic risk**
Presenter: **Giacomo Morelli**, Sapienza University of Rome, Italy
Co-authors: Mauro Bernardi, Lea Petrella
Recent methods to estimate sparse undirected financial graphs assume that the data follow a joint normal distribution. In practice, data strongly deviate from normality and traditional methodologies may fail to correctly identify interdependences. We propose a generalization of the nonpara-normal framework for graphical models by means of a skew copula. We study theoretical properties and derive statistical inference. We then apply the results to real data to assess systemic risk.

CO1256: **Realizing commodity correlations**
Presenter: **Niels Strange Hansen**, Aarhus University, Denmark
Co-authors: Kasper Vinther Olesen, Asger Lunde, Harry Vander Elst
We propose to use a Realized Beta GARCH model to exploit the potential of using high-frequency data in commodity markets for the modeling of correlations. The model produces accurate forecasts of pairwise correlations between commodities which can be used to construct a composite covariance matrix. We evaluate the attractiveness of this matrix in a portfolio context and compare it to models more commonly used in the industry. We demonstrate significant economic gains in a realistic setting including short selling constraints and transaction costs.

CO1332: **Robust clusterwise autoregressive conditional heteroskedasticity**
Presenter: **Giuseppe Storti**, University of Salerno, Italy
Co-authors: Pietro Coretto, Michele La Rocca
The aim of the work is twofold. First, we investigate the dishomogeneity of the cross-sectional distribution of realized stock volatilities, using model based clustering algorithms to tame the extreme tail behaviour typically characterizing this kind of data. Second, we show that this information can be profitably used for specifying parsimonious state-dependent models for volatility forecasting. Under this respect we propose novel GARCH-type model specifications whose parameters can vary through time as a function of the probability that, at a given time point, the stock belongs to a specific volatility cluster. Finally the empirical performance of the proposed models is assessed by means of an application to a panel of U.S. stocks traded on the NYSE.

CO286 Room G21A FINANCIAL TIME SERIES Chair: Melanie Schienle

CO1155: **Test for serial correlation in mean and variance of a sequence of time series objects**
Presenter: **Taewook Lee**, Hankuk University of Foreign Studies, Korea, South
Co-authors: Cheolwoo Park
With help from modern technology large-scale data storage becomes common in the Big Data era, and it motivates statisticians to analyze new types of data. Testing serial correlation in a sequence of sets of time series is considered, namely time series objects. An example is serial correlation of monthly stock returns when daily stock returns are observed. One could consider a representative or summarized value of each object to measure the serial correlation, but this approach would ignore the variation information available in the observed data. We develop a Ljung-Box type test for serial correlation in both mean and variance using the wild bootstrap sampling approach, which takes the variation within a time series object into account. We study the asymptotic property of the proposed bootstrap test and present its financial performance using simulated and real examples.

CO0939: **ARMA-GARCH models and kernel density estimation for short-term density forecasting of wave energy**
Presenter: **Jooyoung Jeon**, University of Bath, United Kingdom
Co-authors: James Taylor
Wave energy has great potential as a renewable source of electricity. With developments in technology, wave energy is likely to play an important role in the future mix of electricity generation. The short-term forecasting of wave energy is needed for the efficient operation of wave farms and power grids, as well as for energy trading. The intermittent nature of wave energy motivates the use of probabilistic forecasting. We evaluate the accuracy of probabilistic forecasts of wave energy flux from a variety of methods, including unconditional kernel density estimation, univariate and bivariate conditional kernel density estimations, univariate and bivariate autoregressive moving average generalised autoregressive conditional heteroskedasticity (ARMA-GARCH) models, and a regression- based method proposed in the literature. Our empirical analysis uses hourly data from the FINO1 research platform in the North Sea to evaluate density and point forecasts, up to 24 hours ahead, for the wave energy flux and for wave power data generated from wave height and wave period observations using a conversion matrix from the literature.

CO0923: **A Semiparametric Intraday GARCH-X Model**
Presenter: **Peter Malec**, University of Cambridge, United Kingdom
Co-authors: Melanie Schienle
We propose a multiplicative component model for intraday volatility. The model consists of a semiparametric and parametric component. The former captures the well-documented intraday seasonality of volatility as well as the impact of the state of the limit order book, utilizing a semiparametric additive structure. The parametric component accounts for short-run fluctuations by means of a unit GARCH specification. The model is estimated by a simple and easy to implement algorithm, consisting of smooth-backfitting and QML steps. Further, we provide an automatic data-driven procedure for bandwidth choice. We show the asymptotic properties of the estimator and document its finite sample performance in a comprehensive simulation study. Finally, our empirical application based on high-frequency data for NASDAQ equities investigates non-linearities in the relationship between the limit order book and subsequent return volatility and underlines the usefulness of including order book variables for out-of-sample forecasting performance.

CO1293: **Consistent estimation of the high dimensional efficient frontier**
Presenter: **Nestor Parolya**, Leibniz University Hannover, Germany
Co-authors: Nikolaus Hautsch, Taras Bodnar
We analyze the asymptotic behavior of the main characteristics of the mean-variance efficient frontier using the random matrix theory. Our particular interest covers the case when the dimension p and the sample size n tend to infinity simultaneously and their ratio p/n tends to a positive constant $c \in (0,1)$. We neither impose any distributional nor structural assumptions on the asset returns. For our theoretical framework we only need some regularity conditions and the existence of the 4th moments. It is shown that two out of three quantities of interest are biased and overestimated by their sample counterparts under the high-dimensional asymptotic regime. This is shown by finding the asymptotic deterministic equivalents of the sample plug-in estimators. Using them we construct the consistent estimators for the three characteristics of the efficient frontier. Furthermore, the asymptotic normality of the considered estimators for the parameters of the efficient frontier is proved. These theoretical results are also verified within an extensive simulation study. The derived estimator for the efficient frontier is a great alternative to the sample estimator for high-dimensional data. Finally, an empirical application is presented where we deal with the estimation of the efficient frontier from stocks traded on the London Stock Exchange.

CO1281: **Testing against asymmetry between tail dependence functions of financial data**
Presenter: **Carsten Bormann**, Karlsruhe Institute of Technology, Germany
Co-authors: Melanie Schienle
Financial investors often seek to minimize portfolio exposure to joint extreme events. In extreme value theory, multivariate extremes are modelled by tail dependence functions, e.g. the tail copula. We propose a new test for detecting differences between two tail copulas. We compare empirical tail copulas over a partition of the unit simplex with multiplier bootstrap $t-$tests, amounting to an entire set of individual $p-$values. Equality of tail copulas is rejected by controlling the per comparison error rate. The test locates sample regions where tail asymmetry is present and thus provides investors with a detailed picture of the tail asymmetry structure. As a byproduct, the test can also detect asymmetries within a tail copula. In Monte Carlo simulations, the new test outperforms existing competitors. For SP500 constituents of the past 25 years, we are able to identify 60% more tail asymmetries than current tests. Hence our test detects substantially more opportunities for hedging tail risks.

CO386 Room Senate INDIRECT INFERENCE AND RELATED METHODS Chair: Veronika Czellar

CO1169: **Misspecification of causal and noncausal orders in autoregressive processes**
Presenter: **Christian Gourieroux**, University of Toronto, Canada
Co-authors: Joann Jasiak
The aim is to study the consequences of estimating a past-dependent (causal) AR model from data generated by a stationary noncausal process with a future-dependent component. We show that the outcomes of that estimation depend on the noncausal persistence. When the noncausal persistence is strong, the (pseudo)-ML estimator of the misspecified causal model is consistent, and reveals the presence of a noncausal component as long as the sample is sufficiently non-Normal. When the noncausal persistence is weak, the (pseudo)-ML estimator is inconsistent and leads to a misleading conclusion that the fit of the past-dependent model is correct. The results are derived theoretically from examining the binding functions and illustrated by simulations of noncausal AR processes with errors that follow mixture distributions with varying proportions of normal and Cauchy variables.

CO0213: **Quantile-based inference for tempered stable distributions**
Presenter: **David Veredas**, Vlerick Business School, Belgium
Co-authors: Hassan Fallahgoul, Frank Fabozzi
A simple, fast, and accurate method for the estimation of numerous distributions that belong to the tempered stable class is introduced. The method is based on the Method of Simulated Quantiles and it consists of matching empirical and theoretical functions of quantiles that are informative about the parameters of interest. In the Monte Carlo study we show that MSQ is significantly faster than Maximum Likelihood and the estimates are almost as precise as under MLE. A Value-at-Risk and Expected Shortfall study for 13 years of daily data and for an array of market indexes world-wide shows that the tempered stable estimation with MSQ estimates provides reasonable risk assessments.

CO0884: **Estimating the role of limited participation in the joint behavior of asset prices and individual consumptions**
Presenter: **Francois Le Grand**, EMLyon Business School, France
Co-authors: Veronika Czellar
Following the literature highlighting the role of limited participation and heterogeneity in asset pricing, we propose an asset pricing model featuring both limited participation and heterogeneity. We estimate the model through indirect inference on individual US consumption data. In our model agents face a three-stage financial market participation risk. When the odds are bad, agents do not participate in the financial markets and trade neither bonds nor stocks. When the odds are good, agents participate in both markets. Finally, when the odds are intermediate, agents can trade bonds but no stocks. Depending on how they participate in financial markets, agents are endowed with different consumption processes, which potentially satisfy Euler conditions for bond and stock prices. We jointly estimate consumption, stock return and stock market participation processes. In particular, we propose a measure of limited-participation risks, which are the respective probabilities to be kicked out of the stock and the bond market.

CO1150: **Nonparametric mixture models and HMMS**
Presenter: **Judith Rousseau**, University Paris Dauphine, France
Co-authors: Elisabeth Gassiat, Elodie Vernet, Kerrie Mengersen
In the recent years some results have been obtained about the identifiability of mixture models - possibly dynamical - when the emission distributions are not specified. In particular, in the case of independent and identically distributed hidden states living on a finite state space, the parameters (emission distributions and weights of the mixture) are identifiable when each individual is associated to three independent observations. In the case of non independent hidden states, then as soon as the transition matrix is invertible, then the parameters are identifiable. We investigate estimation in these models and discuss some aspects of semi-parametric Bayesian estimation, including Bernstein von Mises theorems for the weights (or transition matrices) and estimation of the number of hidden states.

CO0245: **A reliable and testable alternative to long-run restrictions in structural VAR models**
Presenter: **Florian Pelgrin**, EDHEC Business School, France
Co-authors: Alain Guay
A new identification method for structural VAR models is proposed based on frequency interval restrictions. In doing so, we use a previous methodology, the generalization of GMM for a continuum of moment conditions, in the case of the asymptotic least squares method, and we thus propose a new estimator, namely the continuum asymptotic least squares estimator (C-ALS). Our new methodology allows to obtain consistent estimates of impulse responses and reliable confidence intervals in contrast to usual long-run restrictions. Moreover the imposed restrictions can be tested formally and it offers a data-driven procedure that can assess formally the relevance of the imposed identifying restrictions. Finally, we provide some new results using extensive Monte Carlo simulations and an application regarding the hours-productivity debate.

CP002 Room Macmillan Hall and Crush Hall POSTER SESSION Chair: Stella Hadjiantoni

CP0186: **Individual labour income, stock prices and who it may concern**
Presenter: **Jan Voelzke**, University of Muenster, Germany
A panel model is estimated, which describes the relationship of individual labour income and stock prices in Germany. We identify groups of individuals that cluster concerning the model parameters that describe firstly the individual labour income dynamics and secondly the relationship between the individual labour income and financial markets. There are several literatures that are connected to the investigated topic, e.g.: Pricing of Human Capital Contracts, Hedging of individual income risk, Portfolio Optimization or Asset Pricing. For the empirical analysis, the German Socio-Economic Panel (SOEP), which is a wide-ranging representative longitudinal study of private households, is used. The German financial market is represented by the DAX30. Methodically a Bayesian model-based non-Gaussian panel data approach is used. A group of individuals with a high cluster assignment probability is found. The characteristics of this group, whose individuals share the same autoregressive dynamics and a common, relatively high dependence on financial markets are investigated further. It can be shown that this group has a statistical significantly different partition of major occupational groups, leading to several implications for the aforementioned literatures.

CP1159: **Time-varying transition probabilities for Markov switching copula models**
Presenter: **Anna Czapkiewicz**, AGH University of Science and Technology, Poland
Co-authors: Pawel Jamer, Joanna Landmesser
Markov switching Copula-Garch model with time varying probabilities for the transitions is proposed to describe the relationships between markets from G6 group. Different copula functions are taken into consideration in this model. It is assumed that the probability of switching from one regime to the other depend on the behavior of some economic indicators such as an unemployment rate, CPI, long-term interest rate, industrial production. The logistic function for the transition probabilities is used to maps the transition variables into the unit interval. The selection of the transition variables is done in two steps. Firstly, the null hypothesis of a standard Markov switching model against a time-varying transition probability Markovian model is tested. Secondly, it is verified that the influence of macroeconomic data is statistically significant. The EM algorithm adopted to estimate unknown parameters of discussed model is presented.

CP1505: **Robustified sequential estimation of GARCH processes**
Presenter: **Radek Hendrych**, Charles University in Prague, Czech Republic
The generalized autoregressive conditional heteroscedasticity (GARCH) process is a particular modelling scheme, which is capable of forecasting the current level of volatility of financial time series. This classic benchmark model is designed to track changes in volatility of financial returns by using an exact function of past squared measurements. Recently, different on-line (sequential) estimation methods suitable for this class of processes have been introduced in the literature. They undoubtedly represent attractive alternatives to the standard off-line calibration procedures with many practical applications. It might be truly advantageous to adopt numerically effective estimation techniques that can calibrate and control these models sequentially (in real time). However, abnormal observations may occur in data. They may be caused by many reasons, e.g. by additive errors or measurement failures. Exceptional data points will influence the model estimation considerably if no specific action is taken. The aim is to introduce and analyse various possible modifications that can make the sequential estimation schemes more robust. Particularly, extensive

Monte Carlo experiments are performed in order to compare and evaluate all introduced approaches. A real data example is also considered and discussed.

CP1518: **Intraday long-term linkages on European stock markets**
Presenter: **Tomasz Wojtowicz**, AGH University of Science and Technology, Poland
We investigate long-term linkages between intraday prices, volatility and trading volume on stock exchanges in Frankfurt, Vienna and Warsaw. These are diverse markets. Two of them are developed, while the last one is an emerging market. They also differ in size. We study long memory properties and fractional cointegration of 5-minute data of individual firms as well as market indices. The presented analysis describes information flow on the markets and between them. Particularly, existence of common long-term in volatility and trading volume is important in the context of Mixture Distribution Hypothesis. We apply different nonparametric estimation methods of long memory and fractional integration and cointegration. Hence, we also examine usefulness of these estimation methods to intraday data.

CP1618: **A new forecast combination approach using large time-varying VARs applied to a seven-country data set**
Presenter: **Till Weigt**, University of Muenster, Germany
We propose a new forecast combination approach. In considering k alternative forecast models of the same target variable, it is well-known that the information contained in the past forecast errors of any specific model cannot be used separately to systematically improve future forecasts of that model. We can, however, use the contemporaneous correlation between past forecast errors of all models to systematically improve the future forecasts of every model. Technically speaking, we use a VAR model to simultaneously regress the current forecast model of each model on the past forecast errors of models involved. The resulting forecasts of this VAR model represent our revised future forecast errors, which we then subtract from the genuine future forecasts. This novel methodology belongs to the group of so-called forecast combination methods. In contrast to conventional combination methods, which calculate optimal combination weights to generate one superior forecast, our method improves each forecast on its own by modelling the forecast error. Combining many linear forecast models we forecast output growth in a seven-country quarterly economic data set. In order to handle the resulting large VAR we apply recently developed shrinkage techniques for time-varying VARs.

CP1693: **Measuring systemic risk in the European banking system: A copula approach**
Presenter: **Xinghua Xia**, University of Leicester, United Kingdom
We propose a new way for three widely used systemic risk estimates, which are the CoVaR (Conditional Value-at-Risk), MES (Marginal Expected Shortfall), and SRISK (Systemic Risk Measure), based on copula functions. Our Copula-based methodology provides simple explicit expressions for a broad range of copula families. In this regard, we offer a dynamic and more flexible approach for systemic risk modelling. We focus on a portfolio of large European banks and estimate CoVaR, MES, SRISK measures. We measure the institution's contribution to systemic risk using all three main systemic risk measures including CoVaR, MES and SRISK. And we also investigate whether there are common system factors explaining institutions contribution to systemic risk.

CP1700: **Measurement of common risk factors: A panel quantile regression models for returns and volatility**
Presenter: **Frantisek Cech**, Charles University in Prague, Czech Republic
Co-authors: Jozef Barunik
We propose to measure the common risk factors using panel quantile regression models for returns and volatility. By exploring the fact that volatility crosses all quantiles of the return distribution, and employing panel quantile regressions, we focus on the commonalities in the quantiles of the returns in a selected portfolio. Results of our research are important for correct identification of the sources of systemic risk, and will be particularly attractive for relatively high dimensional applications for dimensionality reduction.

CP1772: **A wavelet-based time-varying co-integration model**
Presenter: **Lubos Hanus**, Charles University in Prague, Czech Republic
Co-authors: Lukas Vacha
We focus on locally stationary models exploiting wavelet analysis to fitting time varying autoregressive models. We develop a wavelet time-varying vector error correction model to unravel short and long-term dynamics of economic systems. First, we critically revise the current literature of time-varying models, as these methods are often plagued by the problem of non-stationary time series or ignoring cointegration. The proposed model overcomes the non-stationarity issue by assuming only locally stationary time-series. We provide large-scale simulations in order to make clear how a particular setup should perform. Using both simulations and an economic example we will encounter the model with several alternatives such as those with Bayesian interference or Markov-switching models, for instance, to provide the comparison and its capabilities.

CP1562: **The stock price effect of Apple keynotes**
Presenter: **Ethan Petersen**, Rose-Hulman Institute of Technology, United States
We analyze the volatility of Apple's stock beginning January 3, 2005 up to October 9, 2014, then focus on a range from 30 days prior to each product announcement until 30 days after. Product announcements are filtered; announcements whose 60 day range is devoid of other events are separated. This filtration is chosen to isolate, and study, a potential cross-effect. Concerning Apple keynotes, there are two significant dates: the day the invitations to the event are received and the day of the event itself. As such, the statistical analysis is conducted for both invite-centered and event-centered time frames. A comparison to the VIX is made to determine if the trend is simply following the market or deviating. Regardless of the filtration, we find that there is a clear deviation from the market. Comparing these data sets, there are significantly different trends: isolated events have a constantly decreasing, erratic trend in volatility but an increasing, linear trend is observed for clustered events. According to the Efficient Market Hypothesis, we would expect a change when new information is publicly known and the results support this claim.

CP1648: **The effect of regulation on Bitcoin**
Presenter: **Connor Kispert**, Rose-Hulman Institute of Technology, United States
One of the most notable qualities of Bitcoin is its decentralized nature and disconnect with the banking and regulatory institutions. Despite this, governments in various countries have taken an opposing stance to the cryptocurrency and the unregulated freedoms it may provide. In analyzing the surging growth and subsequent crash in the Bitcoin market, it was found that the turning point of the Bitcoin market was coincident to the introduction of a major government regulation on cryptocurrency, one of the first of many such regulatory events occurring throughout the turbulent decline in Bitcoins value since December of 2013. This leads to the question of just how much of an effect any single country's regulatory stance can have on the Bitcoin market. The aim is to explore the relationship between the event of new government regulation and the trends in the value of Bitcoin by providing statistical support of the real effect of regulation. This is expanded upon with a discussion of potential causation and implications of these effects.

CC1629: **Nonlinear interdependencies or contagions phenomenons between the main European stock market indices**
Presenter: **Rachida Hennani**, Montpellier - Lameta, France
The recent crises that shook the European's financial markets suggest interdependencies' phenomenon which may intensify in times of crisis. It is then more appropriate to use the term of contagion. This difference in terminology translates different realities that affect the choice of economic policies. To detect these phenomena on the European's financial markets, we use an original dynamic model which is a combination of the noisy

Mackey-Glass model of Kyrtsou and Terraza with dynamical or constant Conditional Correlations GARCH errors. This model allows taking into account two types of transmission: a fundamentals-based contagion modelled by the Mackey-Glass equation and a shift contagion due to the behaviours of investors and modelled by the variance equation. We highlight that it exists a fundamentals-based contagion between French and German indices, from the French index to Greek index, from German and French indices to Italian index and from German index to Portuguese index. We identify a shift contagion between all indices, except Hellenic index.

CC1204: **Intraday asset risk pricing: An integer autoregressive model**
Presenter: **Oren Tapiero**, Concordia University, Canada
Intraday and high frequency time series are mostly defined by a non-continuous prices process. An integer-based ARMA model is introduced and found to be a better predictor of prices and their mean squared and absolute errors than continuous time estimators (such as GARCH or multiplicative error models). Using intraday data on transactions on the E-Mini S&P500 at time intervals of five and ten minutes we provide a forecasting model for price variations, which performs better than a number of other models.

EI012 Room CLO B01 SPECIAL SESSION ON OPTIMAL EXPERIMENTAL DESIGN Chair: Alexander Donev

EI0190: **Experimental designs attuned for time trends**
Presenter: **Alexander Donev**, University of Manchester, United Kingdom
Co-authors: Camilla Sammut-Powell
Experimental observations are frequently collected sequentially. As a result trends in the mean and the variance of the response can be caused by lurking variables. Often the order in which the observations are collected is randomised to minimise the impact. However, when the nature of the trends can be anticipated, it is better to construct experimental designs that allow for the results to be adjusted for the trends. We propose a general methodology for designing such experiments based on the framework of generalized linear models. Designs completely eliminating the impact of the trends are rarely possible but minimising the effect by choosing appropriate designs proves rewarding when estimation of the model parameters is of interest.

EI1205: **Experiments with sets of treatment combinations**
Presenter: **Anthony Atkinson**, London School of Economics, United Kingdom
The motivation is an experiment in deep-brain therapy in which each patient receives a set of eight treatment combinations and provides a response to each. The experimental region contains sixteen distinct sets of eight treatments. A design question is, which of the sixteen sets should be used and in what proportions? Since the linear model for the results contains only six parameters, it is unlikely that all sixteen points in the design region should be included in the experiment. The same structure can occur in response surface designs where each choice of an experimental setting provides k responses at k different settings of the explanatory variables. The application of D-optimality to such problems is described. The theory for experiments involving a design region of single measurements at single points provides interesting guidance on the structure of the optimum design. Formulae for incrementation of the information matrix by the addition of the k observations elucidate the properties of the optimum designs, which do not necessarily involve equal replication, even when the optimum design for a model with p parameters has p support points.

EI1257: **Cross-sectional versus longitudinal design for repeated measures: A comparison**
Presenter: **Rainer Schwabe**, Otto-von-Guericke University Magdeburg, Germany
In experimental situations there is often a substantial variability of the experimental units. This typically arises in bio-sciences but may also appear in engineering experiments caused by varying quality conditions of the material. Since experimental units generally stem from a larger entity, it is commonly assumed that the impact of the experimental units is properly described by random coefficients. Then the resulting observations will be correlated within each experimental unit when repeated measurements are taken. If the random effects are associated with the experimental conditions, then they cause additionally heteroscedasticity of the single observations which has strong impact on the performance of an experimental design. Optimal designs are derived in a longitudinal setup, where experimental conditions may vary within observational units and are the same for all units, and in a cross-sectional setup, where the experimental settings remain fixed within each observational unit but may vary across units. The optimal designs obtained can have quite different structures, and their relative efficiencies can become quite low. Therefore the experimental setup has to be chosen properly to fit the experimental situation.

EI1372: **Optimal designs for comparing curves**
Presenter: **Holger Dette**, Ruhr-Universitaet Bochum, Germany
We consider the optimal design problem for a comparison of two regression curves, which is used to establish the similarity between the dose response relationships of two groups. An optimal pair of designs minimizes the width of the confidence band for the difference between the two regression functions. Optimal design theory (equivalence theorems, efficiency bounds) is developed for this non standard design problem and for some commonly used dose response models optimal designs are found explicitly. The results are illustrated in several examples modeling dose response relationships. It is demonstrated that the optimal pair of designs for the comparison of the regression curves is not the pair of the optimal designs for the individual models. In particular it is shown that the use of the optimal designs proposed, instead of commonly used "non-optimal" designs, yields a reduction of the width of the confidence band by more than 50%.

EO246 Room MAL 421 MULTIVARIATE SURVIVAL DATA WITH LATENT VARIABLES Chair: Taeryon Choi

EO0165: **Robust estimation for general transformation models with random effects**
Presenter: **Yuanyuan Lin**, The Chinese University of Hong Kong, China
The semiparametric transformation models with random effects are useful in analyzing dependent data, for example, recurrent data and clustered data. With the error and random effect distributions specified, it has been previously proved that the nonparametric maximum likelihood estimators (NPMLE) are semiparametric efficient. We consider a more general class of transformation models with random effects, under which an unknown monotonic transformation of the response is linearly related to the covariates and the random effects with unspecified error and random effect distributions. This class of models is broad enough to include many popular models and allows various random effect distributions. We propose an estimator based on the maximum rank correlation, which does not reply on any further model assumption except the symmetry of the random effect distribution. The consistency and asymptotic normality of the proposed estimator is established. A random weighting resampling scheme is employed for inference. Moreover, the proposed method can be easily extended to handle censored data and clustered data. Numerical studies demonstrate that the proposed method performs well in practical situations. An application is illustrated with the Framingham cholesterol data.

EO0344: **Analysis of proportional mean residual life model with latent variables**
Presenter: **Jingheng Cai**, Sun Yat-sen University, China
Co-authors: Xinyuan Song
End-stage renal disease (ESRD) is one of the most serious diabetes complications. Numerous studies have been devoted to revealing the risk factors of the onset time of ESRD in diabetes literature. A proportional mean residual life (MRL) model with latent variables is proposed to assess

the effects of observed and latent risk factors on MRL function of ESRD in a cohort of Chinese type 2 diabetic patients. The proposed model generalizes conventional proportional MRL model to accommodate latent variables and right censored data. A factor analysis model is employed to characterize latent risk factors via multiple observed variables. A borrow-strength estimation procedure, which incorporates the expectation-maximization algorithm and the corrected estimating equation approach is further developed. The asymptotic properties of the proposed estimators are established. Simulation shows that the performance of the proposed methodology is satisfactory. The application to the study of type 2 diabetes reveals insights into the prevention of ESRD.

EO0370: **A Bayesian analysis of generalized transformation latent variable models with heterogeneous censored data**
Presenter: **Junhao Pan**, Sun Yat-sen University, China
Co-authors: Jingheng Cai, Xinyuan Song

Latent variable models are commonly adopted in the situations where multiple outcomes are collected to measure the quantity of interest indirectly. A generalized latent variable approach is proposed to analyze mixed types and heterogeneity of outcome, especially with heterogeneous censored data. The basic model features a mixture of transformation latent variable models. It also employs a multinomial logit model for assessing the influence of some covariates on the class membership probability within the mixture model framework. A Bayesian approach was developed to estimate the unknown parameters and to choose the appropriate number of classes. A simulation study was reported to demonstrate the performance of the proposed approach, and a real data analysis was also provided to illustrate the practical value of the proposed model and methodology.

EO0452: **A joint modeling approach for analyzing multivariate current status failure time data with latent risk factors**
Presenter: **Chunjie Wang**, Changchun University of Technology, China

The regression analysis of multivariate current status data based on additive hazards model is investigated. We propose a joint model to analyze multivariate case interval censored failure time data with latent risk factors. The joint model comprises a confirmatory factor analysis model and an additive hazards model, wherein the former characterizes each latent risk factor on the basis of multiple correlated observed variables and the latter reveals the potential risk factors of failure times of interest. We develop a hybrid procedure that combines the EM algorithm and the multidimensional pseudo-score estimating equations approach to perform parameter estimation. The asymptotic properties of the parameter estimators are established. Simulation studies demonstrate that the proposed method performs satisfactorily. An application to a study of diabetic complications is presented.

EO0604: **A latent variable regression model with differential covariates effect**
Presenter: **Zheyu Wang**, Johns Hopkins University, United States
Co-authors: Xiao-Hua Zhou

The continued efforts to evaluate biomarkers' predictive abilities and identify optimal biomarker combinations are often challenged by the absence of a gold standard, i.e. the true disease status. This is because the motivation for using biomarkers is often that ascertaining the gold standard is too costly, too invasive, or too late for any treatment intervention to be effective. As a result, it may be impractical or unethical to obtain validated information regarding the gold standard. Latent variable models can be utilized in this situation to describe the unobserved gold standard and to assess biomarker performance. Nevertheless, examining covariate effects is necessary in biomarker study due to their crucial role in facilitating accurate evaluation. A latent profile model is considered that allows for flexible biomarker distributions and incorporation of previous knowledge about risk factors while simultaneously permitting researchers to model paticipants' characteristics that putatively affect biomarker levels, and therefore provides information needed to develop more personalized diagnostic procedures. Additionally, the proposed method presents a potential strategy for biomarker combination when gold standard information is unavailable, as it derives a composite risk score for the underlying disease status.

EO278 Room MAL B29 INFERENCE FOR STOCHASTIC PROCESSES Chair: Michael Wiper

EO0192: **Closing the loop: Computational vs. analytical results in continuous-time modeling of infectious diseases**
Presenter: **Carles Breto**, The University of Michigan, United States

The advent of cheap computer power promises to continue to catalyze empirical findings of scientific interest. Such findings sometimes come from data analyses based on models for which analytical results may not be readily available. However, such analytical results complete the circle: empirical findings are more likely to be valued by other scientists once they understand the model used to obtain them and, for such understanding, deriving model analytical properties has proved fundamental. An example of such a loop will be presented, focusing on plug-and-play computational statistical algorithms applied to the analysis of infectious disease data based on continuous-time stochastic processes of the SIR family, commonplace in ecology and epidemiology. Plug-and-play approaches allow the modeler to analyze data by simply writing computer code to simulate models. Empirical results will be presented based novel Markov chain models where the transition rates of the chain are subject to continuous-time white noise. These empirical findings led us to studying the properties of those Markov chains in the unconventional white-noise random environment. Finally, some of these properties will be presented, which reveal an unexpected change in the fundamental nature of the initial Markov chains. Such a fundamental change shows that with great computational power comes great analytical responsibility.

EO0426: **The conditional predictive $p-$value in ABC**
Presenter: **Stefano Cabras**, University Carlos III of Madrid, Spain
Co-authors: Maria Eugenia Castellanos Nueda, Oliver Ratmann

The approach to evaluate the goodness of fit of statistical models is via calibrated $p-$values that are uniformly distributed in $[0,1]$ under the true model. While such $p-$values are available for very simple models, these are prohibitively expensive to calculate for complex ones. We show that, even for models whose likelihood is not available in a closed form expression, asymptotically calibrated $p-$values can be efficiently obtained as a by-product of Approximate Bayesian Computations. Such models often arise in stochastic processes with many latent variables. Specifically, given a set of summary statistics and test statistics, we are able to derive the conditional predictive distribution of the test statistic given the summary statistic. After the exposition of the theory, we illustrate the technique in some examples.

EO0496: **Bayesian analysis of the stationary MAP2**
Presenter: **Pepa Ramirez Cobo**, Universidad de Cadiz, Spain
Co-authors: Rosa Lillo, Michael Wiper

A method for carrying out Bayesian estimation for the two-state stationary Markovian arrival process (MAP2) is described, which has been proposed as a versatile model in a number of contexts. The approach is illustrated on both simulated and real data sets, where the performance of the MAP2 is compared against that of the well-known MMPP2. As an extension of the method, we estimate the queue length and virtual waiting time distributions of a stationary MAP2/G/1 queueing system, a matrix generalization of the M/G/1 queue that allows for dependent inter-arrival times. Our procedure is illustrated with applications in internet traffic analysis.

EO0571: **The role of summary statistics in ABC methodology for controlled branching processes**
Presenter: **Ines Maria del Puerto Garcia**, University of Extremadura, Spain
Co-authors: Miguel Gonzalez Velasco, Carmen Minuesa Abril
Controlled branching processes are stochastic growth population models in which the number of individuals with reproductive capacity in each generation is controlled by a random control function. The purpose is to examine the Approximate Bayesian Computation methods in the context of the controlled branching processes. This methodology provides a satisfactory approach to intractable likelihood problems, approximating the posterior distribution of the parameters without explicit likelihood calculations. The search of appropriate summary statistics, with possible dimension reductions in order to fight the curse of dimensionality, is tackled. We illustrate the accuracy of the proposed methods by the way of a simulated example developed with the statistical software R.

EO0574: **Minimum disparity estimators for the offspring parameters of controlled branching processes**
Presenter: **Miguel Gonzalez Velasco**, University of Extremadura, Spain
Co-authors: Carmen Minuesa Abril, Ines Maria del Puerto Garcia
In the context of discrete-time stochastic processes, one of the most interesting generalizations of the Bienayme-Galton-Watson process is the controlled branching process (CBP). This model is characterized by the fact that the number of individuals with reproductive capacity in each generation is controlled by a random control function. Furthermore, an important issue in the inference theory of branching processes is the development of robust estimation methodologies. In connection with this question, we consider the minimum disparity estimators of the underlying offspring parameters of a CBP. We assume that the offspring distribution belongs to a general parametric family. First, we obtain these estimators considering that the entire family tree up to a certain generation can be observed. After that we consider two incomplete data schemes: that given by the total number of individuals and progenitors in each generation, and the one given only by the population sizes. We examine the asymptotic and robustness properties of the estimators proposed. Measures for robustness qualities against gross errors are also studied. The results are illustrated by simulated examples developed by the statistical software R.

EO216 Room MAL 402 CHANGE POINT ANALYSIS Chair: Zuzana Praskova

EO0205: **Detecting non-simultaneous changes**
Presenter: **Daniela Jaruskova**, Czech Technical University, Czech Republic
Sequentially in time, a sequence of independent random vectors is observed. Using hypotheses testing the aim of statistical inference is to decide whether there is a change in the means of the observed vectors. One usually assumes that if the mean vector changes, it changes only once and the change occurs in all components at the same time. We suppose that every coordinate may change only once but the change in different coordinates does not necessarily occur simultaneously. Supposing that the observed vectors are normally distributed with the independent components a test statistic equivalent to likelihood ratio may be applied. Then, approximate critical values may be obtained using tail behavior of the limit variable being a maximum of a χ^2 field. In case the coordinates are correlated (with a known correlation matrix) we may transform them into uncorrelated components. However, a test statistic equivalent to likelihood ratio reflects the fact that the transformed coordinates may change more times. Under the assumption of no change, an asymptotic distribution of the proposed test statistic is again given by distribution of maxima of a χ^2 random field so that high-level exceedance probability of a non-homogeneous χ^2 field may be applied to get approximate asymptotic critical values.

EO0233: **Two-sample change-point analysis**
Presenter: **Zdenek Hlavka**, Charles University in Prague, Czech Republic
Co-authors: Marie Huskova
Applications of change-point analysis are numerous: many authors studied change-points in mean, variance, regression coefficients, or changes occurring in time series. In contrast to the mainstream of existing literature concerning mostly one-sample problems, we are interested in the investigation of differences observed between two independent random samples. One possible area of application are gender specific differences observed between growth curves. As an illustration, we describe a simple two-sample gradual change analysis and show that the two-sample change-point approach leads to more reasonable and more powerful tests than application of standard two-sample t-test in each age category.

EO0859: **Dependent wild bootstrap and its applications in change point monitoring**
Presenter: **Zuzana Praskova**, Charles University in Prague, Czech Republic
Dependent wild bootstrap is a resampling procedure for dependent data that has been developed as an alternative to existing block-bootstrap methods with the aim to mimic the dependency structure of the analyzed data not only in the blocks but in the whole sample. The method can be used also for irregularly observed data. We use this method to approximate critical values of a test statistic monitoring a change in linear regression models with dependent regressors and errors and demonstrate the performance of the method numerically.

EO0904: **Change points in dependence structures of weak and strong dependent high-dimensional time series**
Presenter: **Ansgar Steland**, University Aachen, Germany
For high-dimensional vector time series of dimension $d = d_n$ depending on the sample size n, the case that d is large compared to n or is even larger than n is of particular interest. We are interest to construct change-point statistics for such high-dimensional time series, in order to investigate the dependence structure for changes. The proposed test can also be used to study changes in linear projections of high-dimensional data. Within a high-dimensional time series model that allows for full covariance matrices, we propose novel large sample approximations for bilinear forms of the sample variance-covariance matrix, in terms of strong approximations by Brownian motions. The results cover weakly as well as many long-range dependent linear processes and are valid for a large class of projection vectors that arise, naturally or by construction, in many statistical problems extensively studied for high-dimensional vector time series. Among those key applications are sparse financial portfolio optimization and sparse principal component analysis. Our results are also directly applicable to the problem of shrinkage estimation. The large sample approximations finally allow us to propose a high-dimensional change-point analysis, in order to test for the presence of a change-point in the dependence structure.

EO0935: **Testing for instability in covariance structures**
Presenter: **Lorenzo Trapani**, Cass Business School, United Kingdom
We propose a test for the stability over time of the covariance matrix of multivariate time series. The analysis is extended to the eigensystem to as certain changes due to instability in the eigenvalues and/or eigenvectors. Using strong Invariance Principles and Law of Large Numbers, we normalise the CUSUM-type statistics to calculate their supremum over the whole sample. The power properties of the test versus local alternatives and alternatives close to the beginning/end of sample are investigated theoretically and via simulation. We extend our theory to test for the stability of the covariance matrix of a multivariate regression model. The testing procedures are illustrated through two applications: we study the stability of the principal components of the term structure of 18 US interest rates; and we investigate the stability of the covariance matrix of the error term of a Vector Auto Regression applied to exchange rates.

EO609 Room MAL B30 MIXTURE MODELS FOR NON-VECTORIAL DATA Chair: Charles Bouveyron

EO0247: **The discriminative functional mixture model for a comparative analysis of bike sharing systems**
Presenter: **Julien Jacques**, University Lyon II, France
Co-authors: Charles Bouveyron, Etienne Come

Bike sharing systems (BSSs) have become a means of sustainable intermodal transport and are now proposed in many cities worldwide. Several European BSSs are analyzed and compared to identify common operating patterns in BSSs and to propose practical solutions to avoid potential issues. Our approach relies on the identification of common patterns between and within systems. To this end, a model-based clustering method, called FunFEM, for time series (or more generally functional data) is developed. It is based on a functional mixture model that allows the clustering of the data in a discriminative functional subspace. This model presents the advantage in this context to be parsimonious and to allow the visualization of the clustered systems. The application of FunFEM to BSS data from JCDecaux and the Transport for London Initiative allows us to identify 10 general patterns, including pathological ones, and to propose practical improvement strategies based on the system comparison. The visualization of the clustered data within the discriminative subspace turns out to be particularly informative regarding the system efficiency. The proposed methodology is implemented in a package for the R software, named funFEM, which is available on the CRAN.

EO0340: **Joint extraction of topics and sentiments: An application to probabilistic topic modeling**
Presenter: **Julien Velcin**, Univ Lyon, France

In data mining, topic modeling and sentiment analysis (or opinion mining) are two popular tasks that deal with textual data. Those tasks are usually treated separately, even though they are complementary: sentiments usually target topics and topics can be the basis of subjective positions. We will present an attempt to jointly extract topics and sentiments by extending the classical probabilistic LDA model. Based on two case studies experiments, we will show that our model named TTS is more fitted to capture the overall dynamics of the opinions expressed on short messages.

EO0392: **Robust non-normal mixture of experts**
Presenter: **Faicel Chamroukhi**, University of Toulon, France

Mixture of Experts (MoE) is a popular framework for modeling heterogeneous data. For regression data, MoE usually use normal experts. However, it is known that the normal distribution may not be able to accommodate observations with heavy tails or outliers. We propose new robust non-normal mixture of experts (NNMoE) which can deal with these issues regarding heavy-tailed and possibly noisy data. We present the robust t MoE and skew t MoE, respectively named TMoE and STMoE. We develop dedicated expectation-maximization (EM) and expectation conditional maximization (ECM) algorithms to estimate the parameters of the proposed models by monotonically maximizing the observed data log-likelihood. We describe how they can be used in prediction and in model-based clustering of regression data. We also present applications on simulated and real-world regression data.

EO0508: **Analyzing spatio-temporal behaviors of bike sharing systems through clustering and topic modelling: Some applications**
Presenter: **Etienne Come**, IFSTTAR, France

The analysis of Bike Sharing System (BSS) data is investigated from different angles. First, we present a set of spatio-temporal visualizations derived from basic stock statistics that already highlight some of the big challenges encountered in operating BSS and some links between cities structure and Bike Sharing Activities. Secondly, we present more advanced statistical analysis techniques that we used to analyze behaviors of bike-sharing systems. The first method that we present is based on a specific model-based clustering algorithm that we developed to cluster the stations of a BSS network using Origin/Destination (OD) data. After an introduction of the generative model underlying the clustering we detail the results obtained on the large scale system of Paris. The analysis of the results shows that the discovered clustering is closely related to the city functions (transportation, leisure, employment) which can be helpful for a variety of applications, including urban planning and the choice of business location. Crossing the results of the model with sociological and economic data is carried out to show the close relationships between these two aspects. Eventually, we present some experiments around using topic models (such as the Latent Dirichlet Allocation) to study the dynamics of Origins/Destinations of BSS and also present some results obtained with such an approach on the Paris network.

EO1299: **A statistical test for anomaly detection using the reconstruction error of the Kernel PCA**
Presenter: **Chloe Friguet**, Bretagne-Sud University, France
Co-authors: Laetitia Chapel

A non-parametric statistical test that allows the detection of anomalies given a set of (possibly high dimensional) sample points drawn from a nominal probability distribution is presented. Its test statistic is based on the distance between a query point mapped in a feature space and its projection on the eigen-structure of the kernel matrix computed on the sample points. The statistical test is shown to be uniformly most powerful for a given false alarm level α when the alternative density is uniform over the support of the null distribution. The computational performances of the procedure are assessed as the algorithm can be computed in $O(n^3+n^2)$ and testing a query point only involves matrix vector products. Our method is tested on both artificial and benchmarked real data sets and demonstrates good performances regarding both type-I and type-II errors w.r.t. competing methods.

EO102 Room MAL B35 MULTIVARIATE EXTREMES Chair: Michael Falk

EO0285: **Exact simulation of max-stable processes**
Presenter: **Clement Dombry**, Universite de Franche Comte, France
Co-authors: Sebastian Engelke, Marco Oesting

Max-stable processes play an important role as models for spatial extreme events. Their complex structure as the pointwise maximum over an infinite number of random functions makes simulation highly nontrivial. Algorithms based on finite approximations that are used in practice are often not exact and computationally inefficient. We will present two algorithms for exact simulation of a max-stable process at a finite number of locations. The first algorithm generalizes the approach by Dieter and Mikosch for Brown-Resnick processes and it is based on simulation from the spectral measure. The second algorithm relies on the idea to simulate only the extremal functions, that is, those functions in the construction of a max-stable process that effectively contribute to the pointwise maximum. We study the complexity of both algorithms and prove that the second procedure is always more efficient. Moreover, we provide closed expressions for their implementation that cover the most popular models for max-stable processes and extreme value copulas. For simulation on dense grids, an adaptive design of the second algorithm is proposed.

EO0600: **Estimation of marginal expected shortfall under asymptotic independence**
Presenter: **Vicky Fasen**, Karlsruhe Institute of Technology, Germany
Co-authors: Bikramjit Das

Systemic risk attempts to measure the amount of risk contagion in large networks in finance, banking, insurance, and many other large scale systems. Suppose $Z=(Z_1,Z_2)$ denotes risk pertaining to two components of some system. In the context of systemic risk we can think of Z_1 as the risk of a stock we are interested in and Z_2 that of the entire market. It is often observed that the joint risk vector $Z=(Z_1,Z_2)$ is asymptotically independent, meaning that it is highly unlikely that both risks are large together. We study the behavior of $E(Z_1|Z_2>t)$ which is the expected

shortfall of one risk given that the other risk is high under assumptions of multivariate and hidden regular variation. This quantity called Marginal Expected Shortfall (MES) is an important component in the study of systemic risk. Under asymptotic independence of the variables a naive estimate for MES is $E(Z_1)$. We present an extrapolation method to estimate MES which provides sharper estimates and show that our estimator is consistent. The estimates are verified in both simulated and real data sets.

EO0851: **Modelling the extremal dependence by Bernstein polynomials**
Presenter: **Giulia Marcon**, Bocconi Unversity, Italy
Modern-day analyses of extremes enquire the estimation of multiple episodes and their dependence. An accurate estimation of extremal dependence is crucial in order to not underestimate the return period of extremes. Under mild conditions, the dependence among multivariate block maxima should be modelled by a non-parametric structure that has to satisfy a few constraints. This motivates the introduction of Bernstein polynomials as a particular representation of the extremal dependence structures. Their shape-preserving property provides an easy handling of functions subject to specific constraints. The Bernstein polynomials are involved in a non-parametric inferential method for estimating the Pickands dependence function and a simulating strategy that uses the estimated dependence structure for simulating multiple extreme episodes. This methodology can be applied in many fields and involving different representations of the extremal dependence.

EO1057: **Tests for neighborhoods of multivariate and functional generalized Pareto distributions**
Presenter: **Stefan Aulbach**, University of Wuerzburg, Germany
Co-authors: Michael Falk
In multivariate and functional extreme value theory (where "functional" means processes in $C[0,1]$), the margins of a distribution function F are typically transformed such that the result is in the domain of attraction of a max-stable distribution (MSD) with standard Frechet margins. However, if the copula C of F is considered instead -i.e. the margins are transformed to the uniform distribution on $(0,1)$- then the corresponding MSD, if any, has standard negative exponential margins and its stable tail dependence function turns out to be a norm, called D-norm. The characterization of the terms "domain of attraction" and "generalized Pareto distribution" (GPD) by means of D-norms leads to so-called δ-neighborhoods of a GPD, which essentially collect those distributions with a polynomial rate of convergence towards the underlying MSD. We deal with several tests for such δ-neighborhoods, which are compared by means of a simulation study. Since these tests rely on some properly chosen degrees of freedom, their derivation is emphasized and a graphical tool -similar to mean excess plots- is introduced.

EO1247: **Estimating systematic risk under extremely adverse market conditions**
Presenter: **Chen Zhou**, Bank of The Netherlands, Netherlands
The problem of estimating a linear model between two heavy-tailed variables is considered when the explanatory variable has an extremely low (or high) value. We propose an estimator for the model coefficient by exploiting the tail dependence between the two variables and prove its asymptotic properties. Simulations show that our estimation method yields a lower mean squared error than regressions conditional on tail observations. In an empirical application we illustrate the better performance of our approach relative to the conditional regression approach in projecting the losses of industry-specific stock portfolios in the event of a market crash.

EO108 Room MAL B36 OUTLIERS AND ROBUSTNESS IN TIME SERIES Chair: Roland Fried

EO0289: **Depth estimation for autoregressive models with explosion**
Presenter: **Christoph Kustosz**, TU Dortmund University, Germany
Co-authors: Anne Leucht, Martin Wendler, Christine Mueller
The analysis of autoregressive models under mild explosivity spoils the application of standard methods for estimation and testing based on ordinary least squares due to bias problems. Further ordinary least squares methods are affected by nonstandard error distributions and outliers. We propose alternative statistics based on simplicial depth, which reduce the bias close to the unit root and allow parameter tests for autoregressive models. Due to the simple form of our statitsics, the concept can be applied to various models, as nonlinear autoregression or growth models. To avoid some of the computational obstacles in the calculation of our main statistic, we propose simplifications to reduce the computational costs and to allow an application for moderate and large samples. The resulting methods are robust to outliers and jumps in growth processes. The methods finally are applied to real data from crack growth experiments to derive S-N curves describing the relation of load and lifetime.

EO0499: **Detecting scale changes using pairwise differences**
Presenter: **Daniel Vogel**, University of Aberdeen, United Kingdom
Co-authors: Carina Gerstenberger, Martin Wendler
The mean of all pairwise differences is commonly referred to as Gini's mean difference. It has also been previously proposed to use the 1/4th sample quantile of all pairwise differences and call the resulting estimator Q_n. Both estimators are popular scale estimators that combine very good statistical properties with an intriguing conceptual simplicity. We consider these estimators in the context of change-point analysis. We review their efficiency and robustness properties and then construct CUSUM-type change-point test statistics based on Gini's mean difference and various sample quantiles of the pairwise differences. We use recent results on the asymptotics of U-statistics and U-quantiles for dependent data to derive critical values. The behavior of the tests is examined by means of numerical simulations, demonstrating their general superiority over the classical second-moments-based CUSUM test for detecting scale changes.

EO0869: **Classification of structural changes**
Presenter: **Tobias Voigt**, TU Dortmund, Germany
Co-authors: Roland Fried
In the recent past there have been many advances in structural change detection in time series. Tests for different types of structural changes on stationary and non-stationary time series have been developed, which include tests on changes in the mean and in the variance-covariance structure of a time series. Most of these tests were developed for a specific type of time series and usually have more or less strong requirements to be applicable. A problem arises in applications in which one cannot make assumptions on the structure of a time series and/or when it is not clear which kind of structural change one is searching for. In such cases it is not clear which test should be used. We try to develop a methodology to search for structural changes in situations, in which only few assumptions can be made about a time series and in which several types of structural changes are possible. For this we incorporate ideas from classification and machine learning.

EO1042: **Detecting patterns in time series with extreme events**
Presenter: **Svenja Fischer**, Ruhr-University Bochum, Germany
Co-authors: Andreas Schumann, Alexander Schnurr
When dealing with floods one is not only interested in estimating specific quantiles, but especially for publicness a simple method to calculate alert phases for all different gauges is needed. A problem when considering simple empirical thresholds in this case is the presence of extreme events in very short time series, which lead to very high alert steps. We want to present a robustified Camp-Meidell-type method to calculate these alert phases based on annualities. Using these one cannot only gain the above-mentioned classification but can also detect certain clusters of extreme

events. Nevertheless, the correlation coefficients do not indicate such strong dependency. A useful tool in this case are the so called ordinal patterns, which only take into account the order of small and large observations, not their exact value. We show that they are able to detect coincidences in the patterns and therefore measure another type of dependence than the empirical correlation coefficient does. By using the special form of only ordinal patterns they are also robust against extreme events.

EO0764: **Measuring and combining different aspects of goodness of fit of dynamical extreme value models in hydrology**
Presenter: **Peter Ruckdeschel**, University of Oldenburg, Germany
Co-authors: Sascha Desmettre, Bernhard Spangl, Andreas Maendle
Different modeling approaches have been introduced previously to simultaneously capture time dependence and extreme value behaviour of river discharge data, i.e., a state-space model approach, a shot-noise approach, a Gaussian copula approach, and a GLM approach, and we confined ourselves to qualitative modeling aspects. We now assess the goodness of fit (GoF) of these modeling approaches in a quantitative way, and focus on the following aspects separately: GoF for the (one-dimensional) marginals, for the induced interarrival time distribution of subsequent exceedances of a high threshold, for the bivariate copula of two subsequent observations to capture the general one-step dependence structure, and a weighted version of the latter, focussing on the tail dependence structure for extremely large observations. Each of these GoF measures may be used at its own right, but at the same time, can also either be combined to give a (weighted) overall GoF measure or can be treated like a multiple test, judged against the family-wise error rate. We present simulation based critical values for the respective tests and first asymptotic distributional results and report the respective results when applied to hydrological data from a gauge at Donauwoerth.

EO232 Room MAL B20 ADVANCES IN SURVIVAL AND RELIABILITY Chair: Mariangela Zenga

EO0377: **Fuzzy semi-Markov migration processes within a fuzzy stochastic market in credit risk**
Presenter: **Panagiotis Vassiliou**, Aristotle University of Thessaloniki, Greece
The general problem is the asymptotic behavior of the survival probabilities of a non-homogeneous semi-Markov chain with fuzzy and absorbing states within a fuzzy stochastic environment. The stochastic environment within which the non-homogeneous semi-Markov chain operates is modeled as a non-homogeneous semi-Markov chain with fuzzy states. We choose to present the results within the framework of an important application for the worldwide economy that of credit risk. We provide a recursive relation, from which it is possible to calculate the survival probabilities as functions of the basic parameters, the maximum likelihood estimators of which are also discussed. We study the asymptotic behavior of the survival probabilities in the fuzzy states, within the fuzzy stochastic environment, and we provide under easily met conditions their convergence in closed analytic form.

EO0682: **Technical tools for the use of extended gamma processes in reliability**
Presenter: **Sophie Mercier**, Universite de Pau et des Pays de l Adour, France
Co-authors: Zeina Al Masry, Ghislain Verdier
Standard gamma processes are widely used to model the phenomena of cumulative degradation. However, a notable restriction of a standard gamma process is that its variance-to-mean ratio is constant over time. To overcome this restriction (among others), the use of an extended gamma process is proposed, which is defined as a stochastic integral with respect to a standard gamma process. Contrary to its standard version, the scale parameter of an Extended Gamma Process (EGP) may vary over time. This allows for more flexibility for modelling purpose. However, there is a cost and the use of an EGP presents technical difficulties which have to be dealt with, before EGPs to be used in a practical reliability context: as a first step, the probability distribution of an EGP is not known in full form and it has to be numerically assessed; as a second step, no exact simulation procedure is available, so that approximate simulation procedures have to be developed. We deal with these two issues: an approximate EGP with a piecewise constant scale function is first constructed, which allows both to compute the cumulative distribution function of a general EGP at a known precision and to simulate approximate paths. The quality of the approximation is studied and compared to other possible methods from the literature, both from a theoretical and numerical point of view.

EO1171: **A longitudinal model with individual repeated measures as predictors of a Coxian phase-type survival distribution**
Presenter: **Conor Donnelly**, Queens University Belfast, United Kingdom
Co-authors: Lisa McCrink, Adele Marshall, Hannah Mitchell
Previous research has found that there exists a relationship between an individual's survival time and the trajectory of some related, repeatedly observed, time-varying covariate. For instance, the dynamic nature of the haemoglobin (Hb) levels within individuals suffering from chronic kidney disease has a strong association with the individuals' time until death. The repeated observations on individuals longitudinal response are analysed by employing a linear mixed effects model, utilised to generate not only a population-average trajectory for the longitudinal response over time but, more importantly, individual-specific trajectories. The individual deviations from the population-average trajectory, represented by the latent random effects, are subsequently incorporated within a Coxian phase-type distribution so as to investigate their effect on disease progression and thus on the survival of renal patients. The Coxian phase-type distribution is a special type of Markov model, shown previously to accurately represent the time until an event occurs, with the capability to estimate the rate of deterioration of individuals through sequential, unobserved states of the disease before the event is realised. By linking these longitudinal and survival processes, greater insight is offered into how end-stage renal disease progresses and the effect of changing Hb levels on this progression.

EO0919: **Influence of the dependence patterns in the reliability of the BMAP**
Presenter: **Rosa Lillo**, Universidad Carlos III de Madrid, Spain
Co-authors: Pepa Ramirez Cobo, Joanna Rodriguez
It is well known that the Batch Markovian Arrival Process (BMAP) permits dependent inter-event times and batch sizes. The characterization of the dependence structure related to this model is analyzed for the general stationary $BMAP_m(k)$, with event occurrences up to size k. In the case of two states, it is proved that both auto-correlation functions (inter-event times and batch sizes) decrease geometrically as the time lag increases. More rich patterns can be found when more than two states are considered in the embedded Markov process. It is also shown how the dependence associated to the model affects the quantities and distributions that describe the reliability of the process.

EO0994: **A discrete homogeneous Markov model with multiple absorbing states**
Presenter: **Juan Eloy Ruiz-Castro**, University of Granada, Spain
Co-authors: Mariangela Zenga
Markov processes are usual in the survival literature to analyse the behavior of an illness that evolves over time. Most of studies consider only one absorbent state in survival analysis but sometimes it could be too restrictive because of different absorbing causes could be of interest. A discrete homogeneous Markov model with an indeterminate number of absorbing states is developed in an algorithmic and matrix form. Several absorption states are introduced in a general way, and a methodology that allows us to express the quantities in a well-structured form is applied. The model and the likelihood function for a multi-censored data set are built to obtain the parameter estimations and some interesting measures, such as defective survival functions and mean sojourn times are worked out by considering matrix blocks. The results have been implemented computationally and they have been applied to study the evolution of some biomedical/social processes according to different group risks.

EO228 Room MAL B33 HIGH-DIMENSIONAL PEAKS OVER THRESHOLDS METHODS Chair: Holger Rootzen

EO0420: **The outlook for likelihood-based inference on extreme value distributions in high dimensions**
Presenter: **Jenny Wadsworth**, Lancaster University, United Kingdom

In recent years, multivariate and spatial peaks-over-threshold likelihoods have emerged as efficient methods for inference on the tails of high-dimensional distributions. Mathematically these likelihoods are more tractable than alternatives based on max-stable distributions as they do not involve differentiation of an exponential term. Intuitively, it is the extra information gained from observing directly the extremal events, rather than componentwise maxima, that permits the simpler form of the likelihood. However, this information comes at a price, as the configuration of simultaneously "extreme" events at a moderately high threshold may not be the same as at a very high threshold; if this is the case then bias may be induced. The focus is on likelihoods for high-dimensional inference, potential bias and how it scales with dimension, and examine the outlook for bias reduction.

EO1131: **Applications of the tail process for extremal inference**
Presenter: **Anja Janssen**, University of Copenhagen, Denmark
Co-authors: Holger Drees

The extremal behavior of a d-dimensional stationary regularly varying time series $(Y_t)_{t\in\mathbb{Z}}$ can be described by its index of regular variation $\alpha > 0$ and the distribution of the so-called spectral tail process $(\Theta_t)_{t\in\mathbb{Z}}$, which satisfy $\mathcal{L}((Y_{-n},\ldots,Y_m)/y \mid \|Y_0\| > y) \Rightarrow \mathcal{L}((Y\cdot\theta_{-n},\ldots,Y\cdot\theta_m)),\quad y\to\infty$,for a Pareto($\alpha$)-distributed random variable Y independent of $(\Theta_t)_{t\in\mathbb{Z}}$. We will discuss how the specific features of this process (in particular the so-called "time change formula") can be used to construct alternative estimators of extremal characteristics of the process $(Y_t)_{t\in\mathbb{Z}}$, for example of the expression $P(Y_t/\|Y_t\| \in A \; mid \|Y_t\| > y)$ as $y\to\infty$ for Borel subsets A of the d-dimensional unit sphere.

EO1052: **Bayesian inference for multivariate extreme value distributions**
Presenter: **Sebastian Engelke**, Ecole Polytechnique Federale de Lausanne, Switzerland
Co-authors: Marco Oesting, Clement Dombry

The density of parametric multivariate extreme value distributions involves a huge number of terms in higher dimensions. In many applications maximum likelihood estimation of model parameters using full likelihoods is thus infeasible. Composite likelihood methods based on bivariate densities only have therefore been widely applied. Recently, a Bayesian hierarchical model framework has been introduced by using the partition of occurrence of the maxima as a latent variable to fit a Brown-Resnick process to temperature extremes. This enables the use of full likelihoods for the estimation of model parameters by Markov Chain Monte Carlo methods where the updating step for the partition is done by the Gibbs sampler. We explore the Bayesian hierarchical model framework for general max-stable distributions. Applying this approach for popular models in extreme value statistics, we show how the estimation accuracy can be substantially improved compared to composite likelihood methods. Further, the Bayesian framework allows for new hypothesis tests for multivariate extremes.

EO0933: **Estimating tail dependence parameters by least-squares fitting of extremal coefficients**
Presenter: **John Einmahl**, Tilburg University, Netherlands
Co-authors: Anna Kiriliouk, Johan Segers

Consider a random sample in the max-domain of attraction of a multivariate extreme value distribution, and assume that the stable tail dependence function of the attractor belongs to a parametric model. We propose a novel estimator constructed as the minimizer of the distance between a vector of parametric extremal coefficients and the vector of their estimators. This leads to a consistent and asymptotically normal estimator. In addition, the asymptotic variance of the estimator can be minimized by replacing the Euclidian distance by a quadratic form based on a weight matrix estimated from the data. Its computation is very fast since the extremal coefficients have a simple analytical expression for most popular parametric models. Hence the procedure can be used in very high dimensions. Moreover, this estimator is particularly flexible because it is not limited to pairwise inference. We will demonstrate its good performance and propose possible applications.

EO0903: **Estimating tail dependence parameters by least-squares fitting of extremal coefficients II**
Presenter: **Anna Kiriliouk**, Universite Catholique de Louvain, Belgium
Co-authors: Johan Segers, John Einmahl

Consider a random sample in the max-domain of attraction of a multivariate extreme value distribution, and assume that the stable tail dependence function of the attractor belongs to a parametric model. We propose a novel estimator constructed as the minimizer of the distance between a vector of parametric extremal coefficients and the vector of their estimators. This leads to a consistent and asymptotically normal estimator. In addition, the asymptotic variance of the estimator can be minimized by replacing the Euclidian distance by a quadratic form based on a weight matrix estimated from the data. Its computation is very fast since the extremal coefficients have a simple analytical expression for most popular parametric models. Hence the procedure can be used in very high dimensions. Moreover, this estimator is particularly flexible because it is not limited to pairwise inference. We demonstrate its good performance and propose possible applications.

EO274 Room MAL 541 QUANTILE REGRESSION IN HIGH DIMENSION Chair: Jianhui Zhou

EO0429: **Regularized quantile regression for quantitative genetic traits**
Presenter: **Linglong Kong**, University of Alberta, Canada

Genetic studies often involve quantitative traits, such as blood pressure, lipids level, or gene expression. To identify the genetic features that influence these traits can help uncover the etiology of diseases. Traditional methods for analyzing genetic quantitative traits are based on least square estimates, which only consider the conditional mean of the response variable. Quantile regression method instead considers the conditional quantiles of the response variable and is able to describe the underlying regression structure in a more complete manner. At the same time, genetic studies often involve high-dimensional genomic features that further complicate the analysis. We introduce a regularized quantile regression method that is able to characterize the underlying genetic structure in a more integrative manner, and is able to account for the potential genetic heterogeneity. We investigate the theoretical property of our method, and examine its performance through a series of simulation studies. A real dataset is analyzed to demonstrate the usefulness of the proposed method.

EO0673: **Inference for single-index quantile regression models with profile optimization**
Presenter: **Shujie Ma**, University of California-Riverside, United States

Single index models offer greater flexibility in data analysis than linear models but retain some of the desirable properties such as the interpretability of the coefficients. We consider a pseudo-profile likelihood approach to estimation and testing for single-index quantile regression models. We establish the asymptotic normality of the index coefficient estimator as well as the optimal convergence rate of the nonparametric function estimation. Moreover, we propose a score test for the index coefficient based on the gradient of the pseudo-profile likelihood, and employ a penalized procedure to perform consistent model selection and model estimation simultaneously. We also use Monte Carlo studies to support our asymptotic results, and use an empirical example to illustrate the proposed method.

EO0701: **Sparse trimmed regression**
Presenter: **Emre Barut**, George Washington University, United States

Robust techniques for high dimensional estimation is a key topic in real-world applications. We propose a robust method for high-dimensional regression with trimmed loss functions. We establish connections between the proposed method and Bayesian robust regression. We discuss theoretical results characterizing the approach, and develop a custom optimization approach to efficiently fit the penalized high dimensional model. The extensions of the algorithm for generalized linear models is presented, along with numerical experiments that highlight its relevance.

EO0743: **Functional quantile regression**
Presenter: **Ana-Maria Staicu**, North Carolina State University, United States

The aim is to discuss estimation and statistical inference in functional quantile regression in which the response is scalar and the covariate is a functional (curve). The proposed methods are evaluated theoretically and numerically in finite samples, as well as compared with competitive approaches in simulations. Our proposed approach is applied to a bike sharing datastudy to gain better understanding of the hourly bike rentals for casual users.

EO1233: **Partially linear additive quantile regression in ultra-high dimension**
Presenter: **Lan Wang**, University of Minnesota, United States

We consider flexible semiparametric quantile regression model for analyzing high dimensional heterogeneous data. By considering different conditional quantiles, we may obtain a more complete picture of the conditional distribution of a response variable given high dimensional covariates. The sparsity level is allowed to be different at different quantile levels. We approximate the nonlinear components using B-spline basis functions. We first study estimation under this model when the nonzero components are known in advance and the number of covariates in the linear part diverges. We then investigate a non-convex penalized estimator for simultaneous variable selection and estimation. We derive its oracle property for a general class of non-convex penalty functions in the presence of ultra-high dimensional covariates under relaxed conditions. To tackle the challenges of nonsmooth loss function, non-convex penalty function and the presence of nonlinear components, we combine a recently developed convex-differencing method with modern empirical process techniques. We also discuss how the method for a single quantile of interest can be extended to simultaneous variable selection and estimation at multiple quantiles.

EO254 Room MAL 540 HEALTH ECONOMICS Chair: Michael Talias

EO0439: **Bias-corrected performance measurement of health care units in the presence of targeted variables**
Presenter: **Panagiotis Zervopoulos**, Bursa Orhangazi University, Turkey

Noise in data is not uncommon in real-world cases, although it is commonly omitted from performance measurement studies. We develop a stochastic DEA-based methodology to measure performance when the endogenous (e.g. efficiency) and exogenous variables (e.g. perspectives of patients satisfaction), which are incorporated in the assessment, are inversely related. This methodology identifies benchmark units that are not only highly performing but are also assigned scores for their exogenous variables, which are at least equal to user-defined critical values. The measurements we obtain are bias corrected as we apply a multi-parametric method for bias correction, which results in better estimates than the bootstrap-based methods.

EO0477: **Efficient Gaussian process regression to calculate the expected value of partial perfect information in health economics**
Presenter: **Gianluca Baio**, University College London, United Kingdom
Co-authors: Ioanna Manolopoulou, Anna Heath

The Expected Value of Perfect Partial Information (EVPPI) is a decision-theoretic measure of the cost of uncertainty in decision making, used principally in health economic decision making. Despite having optimal properties in terms of quantifying the value of decision uncertainty, the EVPPI is rarely used in practise. This is due to the prohibitive computational time required to estimate the EVPPI via Monte Carlo simulations. However, a recent development has demonstrated that the EVPPI can be estimated by non parametric regression methods, which have significantly decreased the computation time required to approximate the EVPPI. Under certain circumstances, high-dimensional Gaussian Process regression is suggested, but this can still be prohibitively expensive. Applying fast computation methods developed in spatial statistics using Integrated Nested Laplace Approximations (INLA) and projecting from our high-dimensional input space allows us to decrease the computation time for fitting these high-dimensional Gaussian Processes from around 13 minutes to 10 seconds. We demonstrate that the EVPPI calculated using this new method for Gaussian Process regression is in line with the standard Gaussian Process regression method and that despite the methodological complexity of this new method, R functions are available in the package R-INLA to implement it simply and efficiently.

EO0739: **Predictive model assessment for mapping between measures of health outcomes**
Presenter: **Caterina Conigliani**, Universita Roma Tre, Italy
Co-authors: Andrea Tancredi, Andrea Manca

Patient-reported outcomes measures (PROMs), which include those evaluating health-related quality of life, are used extensively in health services research and to inform policy making; in general, PROMs can be distinguished as preference-based and non-preference-based measures, depending on whether the valuation method that is used to derive their index score is consistent with economic theory. Unfortunately, it is quite common for clinical studies to employ different PROMs, thus limiting the comparability of the evidence base that they contribute to. For this reason, national agencies as the National Institute for Health and Care Excellence for England and Wales, while have identified in EuroQoL-5D (EQ-5D) the PROM of choice, are accepting the use of a mapping approach to predict EQ-5D from other PROMs when EQ-5D data have not been collected. We consider the problem of directly predicting EQ-5D responses from 'Short form 12' by means of multivariate ordered probit models, and carry out the analysis within a Bayesian framework. In particular, we address the key problem of choosing an appropriate summary measure of agreement between predicted and actual results when analysing PROMs, and focus on scoring rules, with particular attention devoted to the case of discrete ordered data and to the importance of propriety.

EP0713: **Decomposing efficiency of Greek health system to input-output invariant components under technology heterogeneity**
Presenter: **Kostas Kounetas**, University of Patras, Greece

Adopting a two stages methodological framework the aim is to investigate the productive performance of hospitals and health centers in Greece in 2011, taking into account possible heterogeneity. Based on the metafrontier notion, we re-introduce a methodology previously proposed, which allows the identification of technology gaps among different health structures and their decomposition into input- and output-invariant components using a bootstrap approach. Empirical findings indicate the existence of significant and rather increasing technology heterogeneity within Greek health system and the regions that it belongs, with the health centers to lay away from the Greek health metafrontier. Furthermore, the decomposition results suggests specific prefectures to be output driving for both hospitals and health centers drawing the attention to catching up policies focusing on input and/or output scale adjustments. Finally, a hierarchical modeling approach used to examine the factors that affect technology gaps reveal the statistical importance of prefecture GDP and population as the health structures years of operation.

EC1612: **In search of realistic QALY models for optimal decisions under uncertainty**
Presenter: **Michael Talias**, Open University of Cyprus, Cyprus
We review the statistical structure of the QALY models in the literature of Health Economics and compare the usefulness of various statistical approaches in the decision making process.

EO042 Room CLO 102 RECENT ADVANCES ON FUNCTIONAL DATA ANALYSIS AND APPLICATIONS Chair: M Carmen Aguilera-Morillo

EO0537: **The two-sample problem for functional data**
Presenter: **Joaquin Ortega**, Centro de Investigacion en Matematicas, Mexico
Co-authors: Roberto Barcenas, Adolfo Quiroz
A frequent problem in the analysis of functional data is that of deciding if two samples of functions come from the same population. We present a family of statistics for this problem. These statistics are quadratic forms associated to dot products of functions in the sample with a finite number of adequately chosen functions. The convergence of the statistics proposed to a Gaussian limit is proved under metric entropy and regularity conditions for the functional data. This family of statistics includes as particular cases some statistics based on functional principal components, previously proposed in the literature. As applications of this family of statistics some problems associated to the analysis of sea waves as functional data are considered: the effect of the amount of energy present in the sea surface, as measured through the spectral density of the waves, has on the shape of waves, and the analysis of asymmetries of real waves as compared to waves simulated from Gaussian models.

EO0702: **Multinomial functional regression with application to lameness detection for horses**
Presenter: **Helle Sorensen**, University of Copenhagen, Denmark
The motivation arises from a dataset consisting of 85 acceleration signals collected from trotting horses that are either healthy or have an induced lameness on one of the four limbs. Our aim is to develop a method that uses such a signal for detection of lameness and identification of the lame limb. This is a supervised classification problem with five groups and functions (acceleration curves) as predictors. We propose to use a multinomial functional regression model. We combine the discrete wavelet transform and LASSO penalization for estimation of the model and use the fitted model to predict the class membership for new curves.

EO0711: **Using P-splines for variable selection in functional regression**
Presenter: **M Carmen Aguilera-Morillo**, Universidad Carlos III de Madrid, Spain
Co-authors: Rosa Lillo, Juan Romo
The focus is on the selection of variables in functional linear models (FLM) with functional response and scalar covariates. Specifically, a computationally efficient version of functional LASSO is proposed in terms of the basis representation of the functional response variable, since it is well known that LASSO is a powerful method to select variables and shrink parameters in linear models. If the original data are functions with noise, it is advisable to use P-splines in order to improve the estimation of the linear model and the corresponding forecasts. In addition, it is shown that P-splines can also be a useful tool in variable selection. In order to find the best method to carry out both variable selection and estimation of the parameters in the function-on-scalar regression models, functional LASSO, ordinary least squares and P-spline penalty are combined providing interesting results.

EO1153: **Multivariate functional clustering: A comparison on remote sensing Landsat data**
Presenter: **Maria Francesca Carfora**, Istituto per le Applicazioni del Calcolo CNR, Italy
Co-authors: Anestis Antoniadis, Umberto Amato, Stefano Pignatti, Simone Pascucci
Functional clustering is currently a very active research topic. In particular, several curve-based clustering methods have been proposed in the recent literature and some of them are specifically designed to face multivariate data. We compare the performance of these methods on a real dataset, composed by time sequences of multispectral Landsat images of cultivated land. Data have been collected in the framework of the FP7 founded project ERMES (dedicated to the rice sector to support authorities and farmers) with the objective of discriminating between different cultivations of rice from their constant and seasonal patterns. In this context, the aim is to discuss the implementation issues of the considered methods on real data and to highlight the respective advantages and drawbacks. A special attention has been paid to the common problems of selecting the appropriate number of clusters and of choosing a suitable basis of functions for representing the data curves.

EO1040: **Generalized functional linear models with derivative component**
Presenter: **Besnik Pumo**, Agrocampus Ouest - Angers campus, France
Functional data analysis refers to the area of statistics where the data are observed in form of curves which realizations come from an underlying random process varying over a continuum. This new field of statistics became really popular at the end of the nineties. We focus our attention to two particular models for functional data: regression models (with scalar response) and autoregressive functional models. The functional linear regression (FLR) is given by $Y = r(X) + e$ where r is a linear operator and X is the explanatory (functional) variable. We consider the FLRD (FLR with derivative component) which is defined by $Y = a(X) + b(X) + e$ where X denotes the derivative of X and a, b are linear operators defined on suitable spaces. Autoregressive process in a Hilbert space H (ARH) is a generalization of the multivariate AR process. We consider the ARHD (ARH model with derivative component) and present some statistical results. Finally we consider the GFLM models with derivative component.

EO338 Room MAL 415 STATISTICAL BOOSTING Chair: Andreas Mayr

EO0624: **On the choice and influence of the number of boosting steps**
Presenter: **Heidi Seibold**, University of Zurich, Switzerland
Co-authors: Anne-Laure Boulesteix, Riccardo De Bin
In biomedical research, boosting-based regression approaches have gained much attention in the last decade. Their intrinsic variable selection procedure and their ability to shrink the estimates of the regression coefficients toward 0 make these techniques appropriate to fit prediction models in case of high-dimensional predictors e.g. gene expression data. Their prediction performance, however, highly depends on specific tuning parameters, in particular on the number of boosting iterations to perform. This crucial parameter is usually selected by a K-fold cross-validation procedure, and, consequently, its value may highly depend on a completely random component, namely the considered partition into K folds. We empirically study how much this randomness affects the results of the boosting techniques, in terms of selected predictors and prediction ability of the related models. We take advantage of four publicly available data sets related to four different diseases. In these studies the goal is to predict survival end-points when a large number of continuous candidate predictors are available. We focus on two well known boosting approaches implemented in the R-packages CoxBoost and mboost which cope with this situation. Finally, we empirically show how the variability in selected predictors and prediction ability of the model is reduced by averaging over several repetitions of cross-validation in the selection of the tuning parameters.

EO0175: **Likelihood-based and model-based boosting for Cox models: A comparison with focus on the handling of mandatory variables**
Presenter: **Riccardo De Bin**, University of Munich, Germany

Despite the limitations imposed by the proportional hazards assumption, the Cox model is probably the most popular statistical tool in analyzing survival data, thanks to its flexibility and ease of interpretation. For this reason, novel statistical/machine learning techniques are usually adapted to fit it. This is the case with boosting. The popularity of boosting has been further driven by the availability of user-friendly software such as the R packages mboost and CoxBoost, both of which allow the implementation of component-wise boosting in conjunction with the Cox model. The component-wise boosting, in particular, is a version of boosting in which the explanatory variables are treated separately, allowing the handling of high-dimensional data. Despite the common underlying boosting principles, these two packages use different techniques: mboost is an adaption of model-based boosting, while CoxBoost adapts likelihood-based boosting. After a comparison between these two boosting techniques, we examine the solutions proposed in the corresponding R packages to handle mandatory variables, i.e. variables that - for any number of reasons - must be included in the model. We explore the possibility of extending solutions currently only implemented in one package to the other. We illustrate the usefulness of these extensions through the application to two real data examples.

EO0154: **Boosting multi-state models**
Presenter: **Thomas Kneib**, University of Goettingen, Germany
Co-authors: Holger Reulen

Multi-state models describe the evolution of discrete phenomena (such as the health state of individuals) in continuous time with the goal of analysing the transition intensities between the different states. A common model specification relies on a Cox-type structure with unspecified hazard rate and multiplicative effects of the covariates. We present a functional gradient descent boosting approach that allows us to implement model choice and variable selection in an automated fashion in multi-state models. The approach relies on a stratified Cox model representation that has the particular advantage to allow for the inclusion of cross-transition effects, i.e. effects that are common to multiple transition types. The automatic variable selection framework offered by functional gradient descent boosting now also allows to automatically detect which effects can be combined across the transition types.

EO1017: **Gradient boosting techniques for AUC-based performance criteria**
Presenter: **Matthias Schmid**, University of Bonn, Germany

The area under the receiver operating characteristic curve (AUC) is an often used performance measure to evaluate prediction rules for binary outcome variables. Extensions of the AUC criterion include the partial area under the AUC curve (PAUC), which is given by the AUC within a restricted interval of false positive rates, and the concordance index for survival data (C index), which summarizes the AUC values for a time-to-event outcome at a series of time points. The aim is to present an overview of methods to optimize AUC-based performance criteria within the model-based boosting framework. Using simulated and real data, it is shown how AUC-optimal combinations of predictor variables can be derived via gradient boosting techniques.

EO0911: **Boosting distributional regression models for multivariate responses**
Presenter: **Andreas Mayr**, Friedrich-Alexander-University Erlangen-Nuremberg, Germany
Co-authors: Nadja Klein

Over the last few years, statistical modelling approaches that go beyond the classical regression of the conditional mean have gained more and more attention. One of the most popular model classes in this context are generalized additive models for location, scale and shape (GAMLSS). The main idea of GAMLSS is that each parameter of the conditional distribution - not only the expected value - is modelled by its own additive predictor. We extend this approach towards multivariate responses and present a statistical boosting algorithm that is able to estimate the unknown quantities of these complex models in potentially high-dimensional settings by circling through the different parameter and outcome dimensions.

EO186 Room MAL 414 AUTOMATIC BANDWIDTH SELECTION FOR KERNEL ESTIMATORS Chair: Lola Martinez-Miranda

EO0626: **Calibrating non- and semiparametric inference: Tests and confidence intervals**
Presenter: **Stefan Sperlich**, Univserity of Geneva, Switzerland

The selection of bandwidth is typically focused on minimizing ISE, MISE, AMSE or similar criteria to obtain an optimal estimate of the function or parameter of interest. Much less attention has been paid to automatic bandwidth selection when constructing confidence intervals or bands. This is also true for all other smoothers and their smoothing parameter choice. Even worse, while non- and semiparametric specification is still a very popular topic, the calibration of these tests has typically been ignored. At most, some papers have discussed data-adaptive bandwidth choices to optimize the power of the test. As a consequence, right in the area where most (at least theoretical) research in nonparametric methods is presently happening, its usefulness remains doubtful. We try to highlight the problem and potential ways to solve them. This is combined with a survey of existing selection methods.

EO0259: **Pilot pre-smoothing without reference distributions**
Presenter: **Jose E Chacon**, Universidad de Extremadura, Spain

Cross-validation bandwidth selectors have the desirable property of not relying on a further pilot bandwidth choice, but exhibit a large variability. Pre-smoothing the data greatly helps to reduce this variability, but the price to pay is that a pilot bandwidth must be chosen. Usually the problem becomes cyclic, so that at some stage a reference distribution is used as an initial pilot bandwidth guess. We present a new bandwidth selection methodology that performs some pilot pre-smoothing, although no pilot bandwidth needs to be chosen. Hence, it reduces the variability over cross-validation, but it does not involve a new bandwidth selection problem nor the use of a reference distribution.

EO0535: **Adaptive bandwidth selection with cross validation for locally stationary processes**
Presenter: **Stefan Richter**, Heidelberg University, Germany
Co-authors: Rainer Dahlhaus

Locally stationary processes behave in an approximately stationary way over short periods of time. There exists a subclass of such processes which allows a decomposition into a stationary time series model depending on a parameter, and a deterministic parameter curve which encodes the time dependence. A prominent example of such a process is the tvARMA process. We assume that the stationary time series model is known, but the parameter curves are not. For estimation of these curves, nonparametric kernel-type maximum likelihood estimates (depending on a smoothing parameter) have been proposed. To the best of our knowledge, the theoretical behavior of a data adaptive bandwidth choice method for such estimates has not been considered in the literature. We propose an adaptive bandwidth choice via cross validation. We prove that this procedure is asymptotically optimal with respect to a Kullback-Leibler-type distance measure under mild assumptions on the unknown parameter curve. The performance of the method is also studied in a simulation.

EO1110: **A bandwidth selector for nonparametric quantile regression**
Presenter: **Mercedes Conde-Amboage**, University of Santiago de Compostela, Spain
Co-authors: Cesar Sanchez-Sellero
In the framework of quantile regression, local linear smoothing techniques have been studied by several authors. The problem of bandwidth selection was addressed in the literature by the usual approaches, such as cross-validation or plug-in methods. Most of the plug-in methods rely on restrictive assumptions on the quantile regression model in relation to the mean regression, or on parametric assumptions. We present a plug-in bandwidth selector for nonparametric quantile regression, that is defined from a completely nonparametric approach. To this end, the curvature of the quantile regression function and the integrated sparsity (inverse of the conditional density) are both nonparametrically estimated. The new bandwidth selector is shown to work well in different scenarios, particularly when the conditions commonly assumed in the literature are not satisfied.

EO0853: **Do-validating local linear hazards**
Presenter: **Lola Martinez-Miranda**, University of Granada, Spain
Co-authors: Maria Luz Gamiz, Enno Mammen, Jens Perch Nielsen
The theory and practice of local linear kernel hazard estimation is brought together. Bandwidth selection is analysed, including the Do-validation method that is shown to have good practical and theoretical properties. Insights are provided into the choice of the weighting function in the local linear minimization and it is pointed out that classical weighting sometimes lacks stability. A new semiparametric hazard estimator transforming the survival data before smoothing is introduced and shown to have good practical properties. The methods are illustrated using mortality data and simulation experiments show their finite sample properties. Finally the recently published R-package named DOvalidation is presented. It provides original functions that implement the methods as well as the datasets used in the illustrations.

EO054 Room CLO 203 MULTIVARIATE ANALYSIS Chair: Apostolos Batsidis

EO0655: **Empirical divergence estimates in moment condition models: Robustness properties**
Presenter: **Aida Toma**, Bucharest Academy of Economic Studies, Romania
Co-authors: Amor Keziou
Estimation methods through divergence minimization for moment condition models with unknown parameter are presented. The approach uses a dual representation of divergences under constraints and the explicit forms of the divergence projections. The estimators are asymptotically best in the sense of Hansen and yields a smallest asymptotic covariance matrix. We analyze the local robustness properties through the influence function approach. It is found that the estimators have bounded influence function if and only if the function inducing the natural orthogonality conditions of the model is bounded. This is the case of some other classical estimators. Examples include least square estimators, GMM estimators, the exponential tilting estimator, as well as linear and nonlinear instrumental variables estimators. Since many available econometric models are based on an unbounded orthogonality function, we propose a unified setting for constructing minimum dual divergence estimators with robustness properties by using truncated versions of the original moment conditions.

EO0798: **Self-consistency tests for bivariate distributions**
Presenter: **Jochen Einbeck**, Durham University, United Kingdom
Co-authors: Simos Meintanis
It is well known that a two-dimensional spherically symmetric distribution X is self-consistent w.r.t. to the circle $E||X||$, that is, each point on that circle is the expectation of all observations that project onto that point. This fact motivates the use of the self-consistency property in order to test for spherical symmetry. We construct an appropriate test statistic based on empirical characteristic functions, which turns out to have an appealing closed-form representation. Critical values of the test statistic are obtained empirically. The nominal level attainment of the test is verified in simulation, and the test power under different alternatives is studied. The extendibility of this concept to higher dimensions, or other test problems involving self-consistency (such as: is a given straight line a principal component?) is briefly touched upon.

EO1059: **Efficient dependence modelling of random vectors given extreme components**
Presenter: **Ioannis Papastathopoulos**, University of Edinburgh, United Kingdom
Multivariate extreme value theory is a highly active area of research and its methods and applications constitute a very useful framework for risk modelling and statistical estimation of rare events. However, when modelling data in several dimensions, one is typically confronted with the curse of dimensionality. It is widely recognised that the construction of efficient statistical models and techniques that overcome this problem is imperative. Conditional independence constitutes one of the most fundamental tools and concepts in this direction. We prove that conditional independence is naturally preserved in limit laws of random vectors given extreme components and discuss a broad range of theoretical and practical features including model generation, and statistical estimation and prediction of extreme quantiles.

EO1041: **Robust inference in generalized linear models with missing responses**
Presenter: **Isabel Maria Rodrigues**, CEMAT and University of Lisbon, Portugal
Co-authors: Ana Maria Bianco, Graciela Boente
When dealing with situations in which the responses are discrete or show some type of asymmetry, the linear model may be not appropriate to establish the relation between the responses and the covariates. Generalized linear models serve this purpose, since they allow one to model the mean of the responses through a link function linearly on the covariates. However, sometimes in practice response variables may be missing and some atypical observations may arise. To overcome this situation, we introduce robust procedures to estimate the regression parameter, under a generalized linear model, when missing data possibly occur in the responses. The robustness of the proposed procedures is studied through the influence function. Besides, outlier detection rules are defined using the influence function. Using that the robust estimators are asymptotically normally distributed we construct a robust Wald-type procedure. A simulation study allows us to compare the behaviour of the classical and robust procedures, under different contamination schemes. Applications to real data sets enable to investigate the sensitivity of the Wald-type test $p-$value to the missing scheme and to the presence of outliers.

EO1343: **MaxSkew: An R package for skewness-based projection pursuit**
Presenter: **Cinzia Franceschini**, Urbino University, Italy
Co-authors: Nicola Loperfido
Projection Pursuit is a powerful multivariate method whose widespread use has been hampered by computational difficulties. We deal with them when linear projections maximizing skewness are sought. We propose an algorithm which is implemented in the R package MaxSkew. Statistical applications include cluster analysis, independent component analysis, outlier detection and normality testing.

EO256 Room MAL B34 ROBUST STATISTICS IN R Chair: Valentin Todorov

EO0875: **Robust approaches for three-way compositional data**
Presenter: **Michele Gallo**, University of Naples Orientale, Italy
Co-authors: Maria Anna Di Palma, Valentin Todorov
Robust alternatives for CP and Tucker3, which allow to explore multi-way data taking into account the underlying complex structure, have been proposed in the statistical literature. These algorithms modify in some way the typically used alternating least squares (ALS) procedure, in order to cope with possibly present in the data outlying observations. Furthermore these approaches have been adapted to work on compositional data. Starting from these proposals several algorithms available in R packages will be compared on simulations and examples in order to verify strengths and weaknesses of each of them.

EO1021: **On a more adaptive use of the tclust package**
Presenter: **Luis Angel Garcia-Escudero**, Universidad de Valladolid, Spain
Co-authors: Alfonso Gordaliza, Agustin Mayo-Iscar
The tclust package is a flexible R package for doing robust clustering based on trimming. This high flexibility allows to deal with non-necessarily spherical clusters and cope with different amounts/types of contamination. However, due to this high flexibility, the use of this package in real data applications is not completely straightforward and it requires the specification of some tuning parameters (the number of clusters, the trimming proportion and a constant constraining the relative clusters shapes and sizes). Although a fully automatic way to choose simultaneously all these parameters is not feasible (as happens with any other clustering method), some graphical/numerical tools exist which are aimed at helping the user in this task. With this in mind, we will exemplify the use of some new data-driven tools that can be successfully applied when the tclust user does not want to impose a very particular structure in the obtained clusters. Additionally, these methods allow the tclust package to be initialized with preventive/not very risky parameters configurations that can be later adapted to the particular data set at hand.

EC0845: **Outlier detection in complex survey data including semi-continuous components and missing values**
Presenter: **Matthias Templ**, Vienna University of Technology, Austria
Co-authors: Peter Filzmoser, Olivier Dupriez
Poverty and inequality are measured based on household consumption, income expenditure data, or household income data. Outliers may especially introduce large variances of indicators, but also measurement errors may lead to biased estimates. Especially, indicators such as the Gini coefficient or the Quintile Share Ratio (QSR) are highly sensitive to outliers if non-robust estimation is applied. Adapting and implementing a collection of techniques for detecting outliers and for fixing them by imputation is considered. Important issues concerning the data and outlier detection methods are the number of missing values in each data set as well as structural zeros in income or consumption components. A main focus lies in the understanding of more than ten different outlier detection and imputation methods and their influence of the estimated Gini coefficient as well as a recommendation on which of the outlier detection methods should be preferred for household expenditure data. Additionally to the finding of a simulation study, outlier detection and imputation are carried out on six real-world consumption data and the results are presented.

EO1331: **Least trimmed squares estimators for functional principal component analysis**
Presenter: **Holger Cevallos Valdiviezo**, Ghent University, Belgium
Co-authors: Stefan Van Aelst, Matias Salibian-Barrera
Classical functional principal component analysis can yield erroneous approximations in presence of outliers. To reduce the influence of atypical data we propose two methods based on trimming: a multivariate least trimmed squares (LTS) estimator and its componentwise variant. The multivariate LTS minimizes the multivariate scale corresponding to $h-$subsets of curves while the componentwise version uses univariate LTS scale estimators. Consider a general setup in which observations are realizations of a random element on a separable Hilbert space $\mathcal{H}$. For a fixed dimension q, we aim to robustly estimate the q dimensional linear space in $\mathcal{H}$ that gives the best approximation to the functional data. Our estimators use smoothing to first represent irregularly spaced curves in a high-dimensional space and then calculate the LTS solution on these multivariate data. The solution of the multivariate data is subsequently mapped back onto $\mathcal{H}$. Poorly fitted observations can therefore be flagged as outliers. A simulation study and real applications show that our estimators yield competitive results in identifying outliers and approximating regular data when compared with other existing methods.

EO1379: **The mrfDepth package for multivariate, regression and functional depth**
Presenter: **Mia Hubert**, KU Leuven, Belgium
Co-authors: Peter Rousseeuw, Pieter Segaert, Kaveh Vakili
The concept of depth first appeared in the context of multivariate data with the halfspace depth. It is a way to order the data from the centre outwards, and can, e.g. be used to visualise the distribution of bivariate data by means of the bagplot. Several other notions of multivariate depth have also been introduced including the simplicial depth, the projection depth and the skew-adjusted projected depth based on the adjusted outlyingness. The idea of depth has been extended to the regression setting and more recently to functional data. We have developed an R-package "mrfDepth" which contains implementations of many of these depth-based estimators as well as several graphical representations of the data. User-friendly R-code is combined with C++ implementations of several fast algorithms for the halfspace depth, the Stahel-Donoho outlyingness, the adjusted outlyingness, the depth median and depth contours. Also corresponding Matlab functions are available as part of the LIBRA toolbox.

EO607 Room CLO 306 STATISTICAL ASPECTS OF TOPOLOGICAL DATA ANALYSIS Chair: Bertrand Michel

EO1008: **Random topology: A survey**
Presenter: **Anthea Monod**, Duke University, United States
We give an overview of research topics stemming from applied and random topology, a relatively new field resulting from the application of point-set and algebraic topological methods to data that arise from random processes. The use of topological methods to analyze data is particularly applicable to information technology and networks (e.g. analyzing internet traffic and communications systems), physics (e.g. analyzing cosmological data), biology (e.g. brain imaging and studying gene regulation), and engineering (e.g. studying coordinated robotics). From the perspective of statistics, such methods are applicable to data mining and image processing. We outline the methodology of topological data analysis, and discuss recent results on the algebraic topology of random fields and complexes. We then present some problems that are currently being studied. In particular, we underline some contributions of applied topology to resolving statistical problems, and how statistical methodology can contribute to resolving applied topological problems.

EO0819: **Statistical learning for shape analysis via persistence diagrams**
Presenter: **Mathieu Carriere**, INRIA Saclay, France
Co-authors: Steve Oudot, Maks Ovsjanikov
A novel way is presented to use topological tools, the so-called persistence diagrams, in machine learning. More precisely, a construction is introduced that maps these objects to finite dimensional normed vector spaces while preserving the stability properties they enjoy. The construction

is flexible in the sense that the dimension of the target space can be reduced at will while preserving the stability guarantees. Furthermore, it allows to use all classical kernel methods on the persistence diagrams directly. Then, results in two applications coming from shape analysis are shown: 3d shape matching and shape segmentation, via the use of kernel Support Vector Machines.

EO1032: **Statistical inference with Wasserstein distances**
Presenter: **Max Sommerfeld**, Georg-August-University Goettingen, Germany
Co-authors: Axel Munk
The Wasserstein distance, one of the most fundamental probability metrics, recently has also been recognized as a central tool in statistical applications. It has been demonstrated to elicit important structure from complex data, where conventional methods fail. Despite this growing interest, tools for rigorous statistical inference remain scarce. This motivates us to investigate extensions of the asymptotic theory for Wasserstein distances of empirical measures and their statistical applications.

EO0896: **Topological consistency via kernel estimation**
Presenter: **Omer Bobrowski**, Duke University, United States
Let $X_1, X_2, ..., X_n$ be iid random variables in $d-$dimensional Euclidean space, generated by an unknown probability density function f. The level sets of the density function are of a considerable interest in many areas of statistics. Previous work focused on recovering the level sets of f, minimizing local measures such as the Hausdorff distance or the symmetric distance. We focus on global features related to the topology of level sets. In particular, we are interested in the homology of these sets. Briefly, homology is an algebraic structure describing the topology of a set in terms of connected components and holes. The main difficulty in recovering the homology is that even small perturbations to the estimated density function can generate a very large error. We discuss these problems, and present an estimator that overcomes these difficulties. We then show that this estimator is consistent, and discuss possible applications for clustering and topological manifold learning.

EO0617: **Asymptotic theory for density ridges**
Presenter: **Yen-Chi Chen**, Carnegie Mellon University, United States
The large sample theory of estimators for density modes is well-understood. We consider density ridges, which are a higher-dimensional extension of modes. Modes correspond to zero-dimensional, local high-density regions in point clouds. Density ridges correspond to $s-$dimensional, local high-density regions in point clouds. We derive geometric properties for density ridges and study asymptotic behavior for ridge estimators. We propose a valid procedure for constructing a confidence set for density ridges using the bootstrap.

EO202 Room CLO 101 APPLICATIONS OF FUNCTIONAL DATA ANALYSIS Chair: Alicia Nieto-Reyes

EO1460: **A functional deformation model with Wasserstein metrics**
Presenter: **Jean-Michel Loubes**, University of Toulouse, France
Co-authors: Eustasio del Barrio
We propose a study of a distribution registration model for general deformation functions. In this framework, we provide estimators of the deformations as well as a goodness of fit test of the model. For this, we consider a criterion which studies the Frechet mean (or barycenter) of the warped distributions whose study enables to make inference on the model. In particular we obtain the asymptotic distribution and a bootstrap procedure for the Wasserstein variation.

EO1120: **Predictive domain-selection for brain image data**
Presenter: **Ah Yeon Park**, University of Cambridge, United Kingdom
Co-authors: John Aston
Alzheimers Disease (AD) affects almost 50% of those over the age of 85 and is the sixth leading cause of death in the US. Unfortunately, there is not a single biological marker for AD yet available. To increase its diagnostic power, many studies have focused on understanding the association between cognitive impairment and brain image scans. We consider the baseline PET (Positron Emission Tomography) scans (160,160,96) as a predictor to predict the measure of cognitive performance, MMSE (Mini-Mental State Exam). We propose an algorithm that selects sub-domains of brain with most predictive-power for MMSE. We view the entire 3D image as a single functional input. Our algorithm involves two estimation steps. First, we segment 3D brain images into several parts, depending on its correlation structure. The second is the selection of the optimal combination of brain segments, based on a cross-validated error criterion. Our sequential algorithm extends the type of possible combinations of segments retained at each step, while keeping the computational burden in a reasonable range. The algorithm is terminated when the difference between two successive mean squared error of Y falls below a given level. Our new approach facilitates interpretation as the estimator of the coefficient function is exactly zero over regions where no relationship to MMSE is present.

EO0976: **Mixture modeling for longitudinal data**
Presenter: **Annie Qu**, University of Illinois at Urbana-Champaign, United States
We propose an unbiased estimating equation approach for a two-component mixture model with correlated response data. We adapt the mixture-of-experts model and a generalized linear model for component distribution and mixing proportion, respectively. The new approach only requires marginal distributions of both component densities and latent variables. We utilize serial correlations from subjects' subgroup memberships, which improves estimation efficiency and classification accuracy, and show that estimation consistency does not depend on the choice of the working correlation matrix. The proposed estimating equation is solved by an Expectation-Estimating-Equation (EEE) algorithm. In the E-step of the EEE algorithm, we propose a joint imputation based on the conditional linear property for the multivariate Bernoulli distribution. In addition, we establish asymptotic properties for the proposed estimators and the convergence property using the EEE algorithm. Our method is compared to an existing competitive mixture model approach in both simulation studies and an election data application. Supplementary material is available online.

EO0423: **Shape analysis and interpoint distance distributions**
Presenter: **Antonio Cuevas**, Autonomous University of Madrid, Spain
Co-authors: Jose Berrendero, Beatriz Pateiro-Lopez
A shape can be informally defined as an equivalence class made of plane figures (i.e. compact sets in the Euclidean plane) which only differ by a change of scale plus an isometry. The main interest is discussing to what extent the shapes can be identified, for practical purposes, using the so-called interpoint distance distribution, that is, the distribution of the distance between two points randomly chosen on the figure. To this end, we define a suitable metric in the space of shapes. We prove that the transformation which takes every shape to the corresponding interpoint distance density is Lipschitz continuous (with respect to both the L1 and the Wasserstein metric in the space of density functions). We also prove an identification result for shapes in terms of the interpoint distance distribution. We discuss a real data example, concerning fishes families classification based on otolith images. Some relevant connections with Functional Data Analysis are pointed out.

EO1165: **Analysis of traffic data using models for time series and functional data**
Presenter: **Paulo Canas Rodrigues**, University of Tampere, Finland
Co-authors: Paula Syrjarinne, Jyrki Nummenmaa
Traffic congestion is a big problem for many major cities. Besides governmental and organizational measures, the analysis of traffic data can improve the well-being of people by reducing the amount of time spent in traffic jams. We will use models for functional and time series data to find patterns in traffic along the day, the week and the year. The data under analysis is bus movement historical data of the city of Tampere, Finland. It includes the location of the bus every second, which was then transformed in the time taken by a given bus in his route every 10 meters, to allow the treatment of different trips as comparable series.

EO078 Room MAL G15 INFERENCE FOR LARGE MATRICES WITH MODERN APPLICATIONS Chair: Igor Pruenster

EO1653: **Estimating network edge probabilities by neighborhood smoothing**
Presenter: **Liza Levina**, University of Michigan, United States
The problem of estimating probabilities of network edges from the observed adjacency matrix has important applications to predicting missing links and network denoising. It has usually been addressed by estimating the graphon, a function that determines the matrix of edge probabilities, but is ill-defined without strong assumptions on the network structure. We propose a novel computationally efficient method based on neighborhood smoothing to estimate the matrix of edge probabilities directly, without making the strong structural assumptions graphon estimation requires. The neighborhood smoothing method requires little tuning, has a competitive mean-squared error rate, and outperforms many benchmark methods on link prediction in both simulated and real networks.

EO0731: **Robust covariance matrix estimation via matrix depth**
Presenter: **Zhao Ren**, University of Pittsburgh, United States
Co-authors: Mengjie Chen, Chao Gao
Covariance matrix estimation is one of the most important problems in statistics. To deal with modern complex data sets, not only we need estimation procedures to take advantage of the structural assumptions of the covariance matrix, it is also important to design methods that are resistant to arbitrary source of outliers. We define a new concept called matrix depth and propose to maximize the empirical matrix depth function to obtain a robust covariance matrix estimator. The proposed estimator is shown to achieve minimax optimal rate under Huber's ε-contamination model for estimating covariance/scatter matrices with various structures such as bandedness and sparsity. Competitive numerical results are presented for both simulated and real data examples.

EO1199: **Statistical inference on the structures of large precision matrices with dependent data**
Presenter: **Yumou Qiu**, University of Nebraska Lincoln, United States
Co-authors: Wen Zhou, Jinyuan Chang
Motivated by the widely used graphic models in modern statistics, we consider to test for the overall conditional dependence structure for the graphic model. Due to the well-known relationship between conditional dependence and precision matrix under normality, equivalently, we test for some specific structures of the corresponding precision matrix. We propose a maximum type test statistic that maximizes the standardized estimates of the entries of the precision matrix. The test is adaptive to the large p small n situations and time dependent observations, which are commonly seen in the areas of brain imaging, spatial temporal statistics, genetic studies and so on. The Gaussian approximation of the proposed statistic is derived. Based on this, we construct a test for the structure of the precision matrix via the multiplier bootstrap established on the moving blocks of the observations. Numerical simulations and case studies on FMRI and Markov Random Fields are conducted to confirm and demonstrate the proposed testing approaches.

EO0828: **Tests for covariance structures with high-dimensional repeated measurements**
Presenter: **Ping-Shou Zhong**, Michigan State University, United States
Co-authors: Wei Lan, Peter Song, Chih-Ling Tsai
In regression analysis with repeated measurements, such as longitudinal data and panel data, structured covariance matrices characterized by a small number of parameters have been widely used and play an important role in parameter estimation and statistical inference. To assess the adequacy of a specified covariance structure, one often adopts the classical likelihood-ratio test when the dimension of the repeated measurements (p) is smaller than the sample size (n). However, this assessment becomes quite challenging when p is bigger than n, since the classical likelihood-ratio test is no longer applicable. The aim is to propose an adjusted goodness-of-fit test to examine a broad range of covariance structures under the scenario of "large p, small n". The analytical examples are presented to illustrate the effectiveness of adjustment for assessing the goodness-of-fit of covariance. In addition, large sample properties of the proposed test are established. Moreover, simulation studies and a real data example are provided to demonstrate the finite sample performance and the practical utility of the test.

EO1272: **Joint limiting laws and a new high-dimensional independence test**
Presenter: **Lingzhou Xue**, Penn State University, United States
Testing independence is of significant interest in many areas of high-dimensional statistical analysis. Existing methods either employ the quadratic-type statistics to test against the dense alternative where the covariance contains a lot of small nonzero entries, or utilize the extreme-value-type statistics to test against the sparse alternative where the covariance has a few nonzero entries. However, both methods do not have good power against the general alternative (either dense or sparse), which is more realistic in practice. In order to resolve this issue, we propose a new high-dimensional independence test, which combines both advantages of the extreme-value-type statistic and the quadratic-type statistic. We study the joint limiting laws of the new test statistics in the high-dimensional setting where the dimension can be much larger than the sample size. Furthermore we derive the convergence rate of the limiting distribution. Numerical performance of the new statistic is also examined.

EG015 Room CLO 204 CONTRIBUTIONS IN TIME SERIES AND TIME-VARYING COEFFICIENTS Chair: Jens-Peter Kreiss

EC1500: **Joint estimation and model order selection for one dimensional ARMA models via convex optimization**
Presenter: **Stephane Chretien**, NPL, United Kingdom
Co-authors: Tianwen Wei, Basad Al-sarray
The problem of estimating ARMA models is computationally interesting due to the nonconcavity of the log-likelihood function. Recent results were based on the convex minimization. Joint model selection using penalization by a convex norm, e.g. the nuclear norm of a certain matrix related to the state space formulation was extensively studied from a computational viewpoint. The goal is to present the first theoretical study of a nuclear norm penalization based variant of the subspace method under the assumption of a Gaussian noise process.

EC1499: **Shape testing in varying coefficient models**
Presenter: **Mohamed Ahkim**, Universiteit Antwerpen, Belgium
Co-authors: Anneleen Verhasselt, Irene Gijbels
Linear regression models are often too rigid for regression analysis. We consider varying coefficient models, which are an extension of the classical linear regression models in the sense that the regression coefficients are replaced by functions in certain variables (often time t). Varying coefficient models have been popular in longitudinal data and panel data studies, and have been applied in fields such as finance and health sciences. We estimate the coefficient functions by splines. An important question in a varying coefficient model is whether a coefficient function is monotonic or convex. We develop consistent testing procedures for monotonicity and convexity. Moreover, we provide procedures to test simultaneously the shapes of certain coefficient functions in a varying coefficient model. The tests use constrained and unconstrained regression splines. The performances of the proposed tests are illustrated on simulated data.

EC1710: **Diagnostic testing for nonstationary fractional cointegration**
Presenter: **Keiko Yamaguchi**, University of Hyogo, Japan
Fractional cointegration generalizes the traditional $I(1)/I(0)$ cointegration model. It is important to know the existence of the cointegarion, or the cointegration rank as well as the estimation of cointegrating relations. We extend the Hausman specification test in a previous work to accommodate both stationary and nonstationary fractionally integrated processes. The test statistic is conceptually and computationally simple, and robust to possible inequality between memory parameters of observables. The extension is accomplished by applying the fully extended local Whittle (FELW) estimator. The previous one also proposed the nonstationary version test used by tapering. We compare our test with the previous one via Monte Carlo simulations, and find that our test statistic peforms better.

EC1769: **Measuring cloud workload burstiness**
Presenter: **Sara Sjostedt de Luna**, Umea University, Sweden
Workload burstiness and spikes are among the main reasons for service disruptions and decrease in the Quality- of-Service (QoS) of online-services. It is a hurdle that complicates autonomic resource management of data centers. We review the state-of-the-art in online identification of workload spikes and quantifying burstiness. The applicability of some of the proposed techniques is examined for cloud systems where different workloads are co-hosted on the same platform. We discuss Sample Entropy (SampEn), a measure used in biomedical signal analysis, as a potential measure for burstiness. A modification to the original measure is introduced to make it more suitable for cloud workloads.

EC0380: **Distribution-free estimation in multivariate state space models with a stationary VAR(1) state equation**
Presenter: **Marco Costa**, University of Aveiro, Portugal
Co-authors: A Manuela Goncalves
The distribution-free estimators have as main advantage the fact that they not need any assumptions on data distribution. Usually, these estimators are less efficient than those of maximum likelihood. However, in many applications, we do not know the data distribution. On the other hand, the maximum likelihood estimation needs hard computational calculations and they do have not an explicit expression. We suggest distribution-free estimators based on the generalized method of moments to the parameters of multivariate state space models with time-varying coefficients with a stationary VAR(1) state equation. These estimators do not assume any data distribution and they are an alternative to the maximum likelihood estimators. In some applications, for instance, in environmental or actuarial data, the Gaussian distribution is inadequate or it is not verifiable. So, distribution-free estimators are an important alternative when the normality is not suitable or they can be used as initial estimates in the iterative procedures to obtain the ML estimates. The proposed distribution-free estimators have an explicit expression and they can be easily computed from data. Additionally, we established sufficient conditions to their consistency.

EG013 Room MAL 539 CONTRIBUTIONS ON DEPENDENCE MODELLING Chair: Jun Yan

EC1736: **Compound random measures**
Presenter: **Fabrizio Leisen**, University of Kent, United Kingdom
Co-authors: Jim Griffin
Many nonparametric priors have been proposed for related distributions and have found a wide range of applications in statistics and machine learning. We describe a new class of dependent random measures which we call compound random measures. These priors can be constructed with gamma, stable and generalized gamma process marginals and their dependence can be characterized using both the Levy copula and correlation function. Normalized version of these random measures can be used as dependent priors for related distributions and inference can be made using a slice sampling algorithm or a Polya Urn approach.

EC1529: **Evading the curse of dimensionality in nonparametric density estimation with simplified vines**
Presenter: **Thomas Nagler**, Technische Universitaet Muenchen, Germany
Co-authors: Claudia Czado
Nonparametric density estimators in more than a few dimensions suffer a great deal from the well-known curse of dimensionality: convergence slows down as dimension increases. A result will be presented that shows that the curse of dimensionality can be avoided by assuming a simplified vine copula model for the dependence between variables. In such models, the conditional dependencies are not affected by the values of the conditioning variables. Under mild and general assumptions, corresponding estimators are consistent and the speed of convergence is independent of dimension. Simulation experiments illustrate the large gain in accuracy - even when the true density does not belong to the class of simplified vines.

EC1493: **Model distances for vine copulas in high dimensions with application to testing the simplifying assumption**
Presenter: **Matthias Killiches**, Technische Universitaet Muenchen, Germany
Co-authors: Daniel Kraus, Claudia Czado
Vine copulas are a flexible class of dependence models consisting of bivariate building blocks and have proven to be particularly useful in high dimensions. Classical model distance measures include multivariate integration and thus suffer from the curse of dimensionality. We provide numerically tractable methods to measure the distance between two vine copulas even in high dimensions. For this purpose, we develop a distance measure based on the Kullback-Leibler distance to reduce numerical calculations by focusing only on crucial spots. For inference and model selection of vines one usually makes the simplifying assumption that the copulas of conditional distributions are independent of their conditioning variables. We present a hypothesis test for this simplifying assumption based on parametric bootstrapping and our distance measure and empirically show the test to have a high power.

EC1382: **Comparison of statistical tests in the paired and independent survey sampling case**
Presenter: **Daniel Gaigall**, Leibniz Universitaet Hannover, Germany
Let $n \in \mathbb{N}$ and $(X_1, Y_1), \ldots, (X_{2n}, Y_{2n})$ be independent and identically distributed random variables with values in a product set with distribution $\mathcal{L}(X_1, Y_1)$. In many practical situations one can choose between the experimental design of a paired survey sampling with observations of the

independent and identically distributed random variables $(X_1,Y_1),\ldots,(X_n,Y_n)$ and the experimental design of an independent survey sampling with observations of the two independent blocks of independent and identically distributed random variables $X_1,\ldots,X_n$ and $Y_{n+1},\ldots,Y_{2n}$. It is obvious and important to pose the question of the better design. Assume that the underlying statistical problem is the testing problem of equality of the marginal distributions, $H:\ \mathcal{L}(X_1)=\mathcal{L}(Y_1),\ K:\ \mathcal{L}(X_1)\neq\mathcal{L}(Y_1)$. Depending on the kind of experiment, different statistical tests apply. The comparison of such tests will be done by using asymptotic relative efficiencies of statistical tests. Of course, this efficiencies depend on the distribution $\mathcal{L}(X_1,Y_1)$, in particular on the dependence structure between X_1 and Y_1. The focus is on the practical most important case of multivariate normal distribution. In this context, new statistical tests for verifying specific properties of the covariance matrix are presented.

EC1705: **Some alternative bivariate Kumaraswamy models**

Presenter: **Indranil Ghosh**, University of North Carolina Wilmington USA, United States

We discuss various strategies for constructing bivariate Kumaraswamy distributions. As alternatives to a previous bivariate model, four different models are introduced utilizing a conditional specification approach, a conditional survival function approach, an Arnold-Ng bivariate beta distribution construction approach, and a copula based construction approach. Distributional properties for such bivariate distributions are investigated. Parameter estimation strategies for the models are discussed, as are the consequences of fitting two of the models to a particular data set involving hemoglobin content in blood samples before and after treatment.

Sunday 13.12.2015 14:30 - 16:10 Parallel Session I – CFE-CMStatistics

CI020 Room Beveridge Hall SPECIAL SESSION ON ADVANCES IN DYNAMIC FACTOR ANALYSIS Chair: Christopher Otrok

CI0441: **Time-varying spillovers**
Presenter: **Christopher Otrok**, University of Missouri and FRB St Louis, United States
We develop a Bayesian Dynamic Factor model to measure cross-country spillovers in macroeconomic aggregates. The model is able to distinguish common shocks across countries from spillovers of country specific shocks. The key innovation is to allow for time variation in the spillover channels. Spillovers may be different in expansions and contractions. They also may vary with specific events, such as financial crises. We apply the model to European countries to measure the evolution of spillovers over time.

CI0446: **Common and country specific economic uncertainty**
Presenter: **Haroon Mumtaz**, Queen Mary University of London, United Kingdom
A factor model with stochastic volatility is used to decompose the time-varying variance of Macroeconomic and Financial variables into contributions from country-specific uncertainty and uncertainty common to all countries. We find that the common component plays an important role in driving the volatility of nominal and financial variables. The co-movement in volatility of real and financial variables has increased over time with the common component becoming more important over the last decade.

CI1824: **Forecasting fed funds target changes**
Presenter: **Michael Owyang**, Federal Reserve Bank of St Louis, United States
Most interest rate rules are continuous functions of deviations of output from its potential and expected inflation from its target. In practice, central banks move the target rate in discrete increments and base their decisions on a wide-range of data. We estimate a dynamic ordered probit model of movements in the federal funds rate. In our model, monetary policy reacts to a large dataset that is summarized by small set of dynamic factors. We then use the model for out-of-sample forecasting and evaluate these forecasts using methods unique to problems of classification.

CO536 Room Bedford FINANCIAL CONDITIONS INDICES Chair: Garry Young

CO0267: **A financial conditions index using targeted data reduction**
Presenter: **Garry Young**, Bank of England, United Kingdom
Co-authors: Simon Price, George Kapetanios
Financial conditions indices (FCIs) aim to summarise the state of financial markets. We construct two types of measure: a principal component of a medium sized set of relevant financial indicators and an alternative that takes information from a large set of macroeconomic variables weighted by the joint covariance with subsets of financial indicators, using multiple partial least squares (MPLS). Our approach aims to weight latent factors from a macroeconomic data set using information from financial variables. Both are useful for forecasting monthly GDP, but the MPLS based FCIs are superior to that based on the PC.

CO0544: **Financial stress regimes and the macroeconomy**
Presenter: **Ana Galvao**, University of Warwick, United Kingdom
Co-authors: Michael Owyang
The aim is to identify financial stress regimes using a model that explicitly links financial variables with the macroeconomy. The financial stress regimes are identified using a large unbalanced panel of financial variables with an embedded method for variable selection and, empirically, are strongly correlated with NBER recessions. The empirical results on the selection of financial variables support the use of credit spreads to identify asymmetries in the responses of economic activity and prices to financial shocks. We use a novel factor-augmented vector autoregressive model with smooth regime changes (FASTVAR). The unobserved financial factor is jointly estimated with the parameters of a logistic function that describes the probabilities of the financial stress regime over time.

CO0932: **A high frequency measure of U.S. GDP with application to financial conditions indexes**
Presenter: **Scott Brave**, Federal Reserve Bank of Chicago, United States
Co-authors: Andrew Butters
We use the data underlying the Chicago Fed National Activity and National Financial Conditions indexes to construct a weekly measure of U.S. GDP growth. A mixed-frequency dynamic factor model is posed which links cyclical movements in GDP growth to a small set of latent weekly real and financial activity factors. Collapsed dynamic factor methods are used to estimate the latent factors in addition to a latent time-varying mean (or trend) for GDP growth. A real-time out-of-sample forecasting exercise is then conducted to evaluate the ability of the model to forecast near-term GDP growth relative to prominent surveys of professional forecasters.

CO1220: **Financial conditions in the Euro area: A narrative of the crisis and its consequences for the real economy**
Presenter: **Hiona Balfoussia**, Bank of Greece, Greece
Co-authors: Heather Gibson
We construct and present financial conditions indices (FCIs) for the Euro area, for the period 2003 onwards, using a wide range of prices, quantities, spreads and survey data, grounded in the theoretical literature. The FCIs fit in well with a narrative of financial conditions since the creation of the monetary union. FCIs for individual Euro area countries are also provided, with a view to comparing financial conditions in core and periphery countries. There is evidence of significant divergence both before and during the crisis, which becomes less pronounced when monetary policy variables are included in the FCI. However, the impact of monetary policy on financial conditions appears not to be entirely symmetric across the Euro area. We subsequently explore the relationship between financial conditions and real economic activity in the Euro area as a whole and for Greece in particular, by estimating the potential impact of the TLTROs on aspects of economic activity. Our results suggest that financial conditions do have a significant effect on economic activity and thus the TLTROs, to the extent that they are designed to improve financial conditions, will provide a boost to the real economy. A further extension explores the link between financial conditions and firm investment, and finds it significant.

CO480 Room SH349 ADVANCES IN FINANCIAL FORECASTING Chair: Ekaterini Panopoulou

CO0357: **Forecasting market returns: Bagging or combining**
Presenter: **Andrew Vivian**, Loughborough University, United Kingdom
Co-authors: Steven Jordan, Mark Wohar
The aim is to provide evidence on applying the bagging method to forecast stock returns out-of-sample for the G7 and a broad set of Asian countries for which there is little prior evidence. We focus on using the recently developed bagging method that explicitly addresses model uncertainty and parameter uncertainty. We are amongst the first to apply the bagging method to market return predictability and amongst the first to examine if bagging can generate economic gains. We find that, when portfolio weight restrictions are applied, bagging generally improves forecast accuracy and generates economic gains relative to the benchmark; bagging also performs well compared to forecast combinations in this setting. We also provide new evidence that the results for bagging cannot be fully explained by data mining concerns. Finally, we report that bagging generates economic gains in G-7 countries and overall these gains are highest for countries with high trade openness and high FDI. The potentially substantial economic gains could well be operational given the existence of index funds for most of these countries.

CO0986: **A Bayesian non-parametric multiple quantile model for forecasting the asset return distribution**
Presenter: **Evangelia Mitrodima**, London School of Economics, United Kingdom
Co-authors: Jim Griffin, Jaideep Oberoi
We perform a Bayesian analysis using Markov Chain Monte Carlo (MCMC) methods and in particular, an Adaptive Metropolis Hastings algorithm to jointly model selected quantiles of the asset return distribution. Bayesian methodology is widely used in the literature concerning quantile regression, for improved estimation based on the Regression Quantile (RQ) criterion, by employing the Asymmetric Laplace likelihood. The Asymmetric Laplace distribution is a skew distribution, which offers a possible mathematical link between the minimization of the RQ criterion and the maximum likelihood theory. However, this does not address the underlying time-varying interdependence between individual quantiles in a natural way. An alternative is to use a non-parametric setting, as this seems more desirable under the Bayesian framework. To this end, we consider a histogram generated by the joint quantile model to approximate the density of asset returns with posterior inference for the parameters.

CO1118: **Quantile forecast combinations in realised volatility prediction**
Presenter: **Ekaterini Panopoulou**, University of Kent, United Kingdom
Co-authors: Ioannis Vrontos, Spyridon Vrontos, Loukia Meligkotsidou
Whether it is possible to improve realised volatility forecasts by conditioning on macroeconomic and financial variables is tested. We employ complete subset combinations of both linear and quantile forecasts in order to construct robust and accurate stock market volatility predictions. Our findings suggest that the complete subset approach delivers statistically significant out-of-sample forecasts relative to the autoregressive benchmark and traditional combination schemes. A recursive algorithm that selects, in real time, the best complete subset for each predictive regression quantile succeeds in identifying the best subset in a time- and quantile-varying manner.

CO426 Room Holden COMMON FEATURES IN MACROECONOMICS AND FINANCE Chair: Joao Victor Issler

CC0365: **Quantile factor models**
Presenter: **Jesus Gonzalo**, University Carlos III de Madrid, Spain
Co-authors: Juan jose Dolado, Liang Chen
A novel concept is proposed: quantile factor models, where a few unobserved common factors affect all parts of the distributions of many observed variables. A simple two step procedure is proposed to estimate the common factors and the quantile factor loadings. Uniform consistency and weak convergence results for the entire quantile factor loading processes are obtained. Based on these results, we show how to make inference of general forms in a location-scale shift factor model. Simulation results confirm the good performance of our estimators in small to moderate sample sizes.

CO1436: **Local unit root and inflationary inertia in Brazil**
Presenter: **Osmani Guillen**, Ibmec and BCB, Brazil
Co-authors: Wagner Gaglianone
The purpose is to study the persistence of Brazilian inflation using quantile regression techniques. To characterize the inflation dynamics we employ a Quantile Autoregression model. In this model, the autoregressive coefficient may assume different values, allowing testing the asymmetry hypothesis for the inflation dynamics. Furthermore, the model allows investigating the existence of a local unit root behavior. In other words, the model enables to identify locally unsustainable dynamics, but still compatible with global stationarity. In addition, the model can be reformulated in a more conventional coefficient notation, in order to reveal the periods of local nonstationarity. Another advantage of this technique is the estimation method, which does not require knowledge of the innovation process distribution, making the approach robust against poorly specified models. An empirical exercise with Brazilian inflation data and its components illustrates the methodology. As expected, the behavior of inflation dynamics is not uniform across different conditional quantiles. In particular, the results can be summarized as follows: (i) the dynamics is stationary for most quantiles of the sample period; (ii) the process is nonstationary in the upper tail of the conditional distribution; and (iii) the periods associated with local unsustainable dynamics can be related to those of increased risk aversion and higher inflation expectations.

CO1381: **Risk assessment of the Brazilian FX rate**
Presenter: **Wagner Gaglianone**, Central Bank of Brazil, Brazil
Co-authors: Jaqueline Marins
We construct several multi-step-ahead density forecasts for the foreign exchange (FX) rate based on statistical, financial data and economic-driven approaches. The objective is to go beyond the standard conditional mean investigation of the FX rate and (for instance) allow for asymmetric responses of covariates (e.g. financial data or economic fundamentals) in respect to exchange rate movements. We also provide a toolkit to evaluate out-of-sample density forecasts and select models for risk analysis purposes. An empirical exercise for the Brazilian FX rate is provided. Overall, the results suggest that no single model properly accounts for the entire density in all considered forecast horizons. Nonetheless, the GARCH model as well as the option-implied approach seem to be more suitable for short-run purposes (until three months), whereas the survey-based and some economic-driven models appear to be more adequate for longer horizons (such as one year).

CO1437: **Consumption-wealth ratio and expected stock returns: Evidence from panel data on G7 countries**
Presenter: **Joao Victor Issler**, Getulio Vargas Foundation, Brazil
Co-authors: Andressa Monteiro de Castro
Using a recent theoretical framework, we perform an empirical investigation on how widespread is the predictability of a modified consumption-wealth ratio – CAY – once we consider the set of G7 countries. The G7 countries represent more than 64% of net global wealth and 46% of global GDP at market exchange rates. We evaluate the forecasting performance of CAY using a panel-data approach, since applying cointegration and other time-series techniques is now standard practice in the panel-data literature. Hence, we generalize Lettau and Ludvigson's tests for a panel of important countries. We employ macroeconomic and financial quarterly data for the group of G7 countries, forming an unbalanced panel. For most

countries, data is available from the early 1990s until 2014Q1, but for the U.S. economy it is available from 1981Q1 through 2014Q1. Results of an exhaustive empirical investigation are overwhelmingly in favor of the predictive power of CAY in forecasting future stock returns and excess returns.

CO384 Room Senate MULTIVARIATE METHODS FOR ECONOMIC AND FINANCIAL TIME SERIES Chair: Gianluca Cubadda

CO0465: **Real time mixed frequency VARs: Nowcasting, backcasting and Granger causality**
Presenter: **Alain Hecq**, Maastricht University, Netherlands
Co-authors: Thomas Goetz, Lenard Lieb
A previous mixed-frequency VAR is extended to the inclusion of a noncausal part. This allows us to consider the future of the variables and to focus more closely on data releases. We propose a method based on the LAD estimator for estimating causal-noncausal VARs for mixed-frequency series. Test statistics are computed using a bootstrap approach. For the European Union, a system with the quarterly growth rate of the real gross domestic product and three monthly growth rates of the industrial production index, reveals that a pure noncausal VAR is preferred by the data to the usual reduced form causal VAR. This leads to what we call backcasting causality instead of the well know concept of Granger causality. Consequently, our modeling implies a new way to forecast time series in real time.

CO0582: **Industrial development in the Italian regions, 1861-1913: New evidence**
Presenter: **Stefano Fachin**, Rome Sapienza, Italy
Co-authors: Francesca Di Iorio, Carlo Ciccarelli
The aim is to tackle the problem of applying the approximate factor model to spatial data using as a case study manufacturing industrial value added in the Italian regions from 1861 to 1913 from a recently released dataset at industry level. The application of the factor model to spatial data raises two questions usually not addressed: (i) the analysis of the loadings, as their distribution for the different factors over the spatial units may reveal important features of the data; (ii) modelling the errors, as the possibile presence of spatial dependence may be also very important. With respect to the first issue we develop a boostrap test for the hypothesis that the loadings are equal to a known matrix, reporting encouraging results of some simulation experiments. The second point is explored estimating spatial autoregressive panel models. Given the analogy with FAVARs, we propose to label the two-step strategy entailed by estimation of spatial error models on the de-factored residuals as FASEM (Factor Augmented Spatial Error Model).

CO0891: **Adjustments of the effects of measurement errors using instrumental variables and mixed-models in cointegration analysis**
Presenter: **Hanwoom Hong**, Seoul National University, Korea, South
Co-authors: Sung Ahn, Sinsup Cho
The effects of measurement errors on the reduced-rank estimator and the cointegrating test of error correction models had been studied. It was shown that the asymptotic bias is present because of endogeneity caused by the measurement errors. Two methods are suggested to deal with the endogeneity. One method employs instrument variables and the other introduces a moving average term in the error correction model. It is investigated that asymptotic properties of the reduced rank estimators based on these two methods and is found that these estimators are no longer asymptotically biased. The asymptotic distribution of the likelihood ratio test for the cointegrating ranks based on these methods is also obtained. Finally, small sample properties of the estimators and the test through a Monte Carlo simulation study is investigated.

CO1686: **Volatility spillovers with multivariate stochastic volatility models**
Presenter: **Yuliya Shapovalova**, Maastricht University, Netherlands
Co-authors: Michael Eichler
Co-movements in financial time series suggest presence of volatility spillover effects among financial markets. Understanding fundamentals behind this phenomena is important for portfolio managers and policy makers. Currently in the literature GARCH-type models is the dominating approach for detecting volatility spillovers. Inference is often based on notions of causality in mean and variance. We aim to analyse volatility spillovers using a more natural approach for volatility modeling: multivariate stochastic volatility models (MSVM). The structure of MSVM allows to test for causality in volatility processes directly, and in contrast to GARCH models causality in variance and causality in volatility do not coincide in this framework. However, due to presence of latent volatility processes estimation of this class of models is a difficult task, it has been shown that standard methods of estimation such as quasi-maximum likelihood and GMM do not perform sufficiently well. We use alternative methods of estimation: particle filters and variational Bayes. The notion of Granger-causality is used to test for causal links in volatility processes and Dirac spike and slab priors for model selection which gives us a fully Bayesian approach. Finally, based on our estimation and model selection we build graphical models for further graphical inference of causal structure.

CO380 Room Court BAYESIAN NONLINEAR ECONOMETRICS Chair: Roberto Casarin

CO0519: **Nonparametric conditional Beta**
Presenter: **John Maheu**, McMaster University, Canada
Co-authors: Azam Shamsi
The effect of the market return on the value of systematic risk using a semiparametric multivariate GARCH model is investigated. We nonparametrically estimate the dynamic conditional beta without any restrictive assumption on the joint density of the data. This model captures movements in systematic risk over time, and we find that the time-varying beta of a stock nonlinearly depends on the contemporaneous value of excess market returns. The model is extended to allow nonlinear dependence in Fama-French factors. In general, in highly volatile markets, beta is almost constant, while in stable markets, the beta coefficient can be highly and asymmetrically dependent on the value of the market excess return.

CO0675: **Bayesian panel Markov-switching model with mixed data sampling**
Presenter: **Massimiliano Marcellino**, Bocconi University, Italy
Co-authors: Roberto Casarin, Claudia Foroni, Francesco Ravazzolo
A Bayesian panel Markov-switching model with mixed data sampling (MIDAS) is proposed. We follow the unrestricted MIDAS approach, provide a Markov-chain Monte Carlo (MCMC) procedure for posterior approximation.

CO0842: **Bayesian nonparametric sparse seemingly unrelated regression model**
Presenter: **Luca Rossini**, Ca Foscari University of Venice, Italy
Co-authors: Roberto Casarin, Monica Billio
Seemingly unrelated regression (SUR) models are used in studying the interactions among economic variables of interest. In a high dimensional setting and when applied to large panel of time series, these models have a large number of parameters to be estimated and suffer of inferential problems. We propose a Bayesian nonparametric hierarchical model for multivariate time series in order to avoid the overparametrization and overfitting issues and to allow for shrinkage toward multiple prior means with unknown location, scale and shape parameters. We propose a two-stage hierarchical prior distribution. The first stage of the hierarchy consists in a lasso conditionally independent prior distribution of the Normal-

Gamma family for the SUR coefficients. The second stage is given by a random mixture distribution for the Normal-Gamma hyperparameters, which allows for parameter parsimony through two components. The first one is a random Dirac point-mass distribution, which induces sparsity in the SUR coefficients; the second is a Dirichlet process prior, which allows for clustering of the SUR coefficients. We provide a Gibbs sampler for posterior approximations based on introduction of auxiliary variables. Some simulated examples show the efficiency of the proposed methods. We study the effectiveness of our model and inference approach with an application to macroeconomics.

CO1211: **Bayesian estimation of multimodal density features applied to DNA and economic data**
Presenter: **Herman van Dijk**, Erasmus University Rotterdam, Netherlands
Co-authors: Nalan Basturk

Important theoretical and practical issues that involve distribution functions that have multimodal densities occur in several scientific fields: bioinformatics, finance, and international economic growth. A Bayesian approach is proposed to estimate shape and other features of a multimodal density and the uncertainty around these values. The method is applicable for continuous and discrete data distributions. For continuous multimodal data, we show that estimates based on mixtures of normal densities with an unknown number of components provide a straightforward method to evaluate density features. For discrete such data, we propose a mixture of shifted Poisson densities with an unknown number of components. Mixture density estimates are obtained using simulation-based Bayesian inference with density features treated as random variables. Highest posterior intervals around features are automatically obtained without any extra computational effort. For discrete data a novel version of a Reversible Jump Markov Chain Monte Carlo (RJMCMC) method is developed which is an adapted version of Green's method. Instead of applying the more restrictive approach of choosing a particular number of mixture components using information criteria as in an earlier work, our method allows for an unknown number of components.

CO396 Room Gordon WAVELET METHODS IN ECONOMICS Chair: Marco Gallegati

CO0569: **Firm ownership and provincial CO2-emissions in China**
Presenter: **Fredrik NG Andersson**, Lund University, Sweden
Co-authors: Sonja Opper

Within only three decades China has emerged from one of the worlds poorest agricultural economies to a major manufacturing economy taking up the largest share of global carbon dioxide emissions. The Chinese economy capitalist transformation has been gradual and the present economy is a hybrid relying on private production a sizable state-owned sector. A large literature has shown that private firms are economically more efficient than state-owned firms, but whether private firms also are more carbon efficiency is another question. We test whether firm ownership affects carbon dioxide emissions using a provincial data covering the period from 1992 to 2010. In our estimations we separate between short-run and long-run effects using wavelet analysis and a band spectrum regression estimator. Our results offer two important policy implications: First, most of the short-run volatility in emission is due to non-economic factors and policy makers should focus on the long-run where they have a greater possibility to affect emissions growth. Second, over the long-term a larger share of private firms reduces the growth rate in carbon dioxide emissions and continuing structural reforms lifting remaining barriers of private firm production will be crucial to contain or even reduce national emissions.

CO0796: **Wavelet-based factor pricing and measurement of macroeconomic risks**
Presenter: **Joanna Bruzda**, Nicolaus Copernicus University, Poland

In the paper we suggest the use of analytic wavelets to measure the exposure to different risk factors within multifactor asset pricing models with macroeconomic sources of risks. Our wavelet-based betas are computed as wavelet partial gain coefficients, which incorporate the influence of any lead-lag effects on the values of these risk measures and enable the investor to track changes in systematic risk across wavelet scales. This proposal is illustrated with empirical wavelet APT models for industry portfolios on the US and European stock markets, showing that the modified beta coefficients can substantially change the assessment of macroeconomic risks and demonstrating how they influence the pricing of different risk factors.

CO1004: **Long waves in external imbalances, credit growth and asset prices: An historical perspective on global financial crisis**
Presenter: **Marco Gallegati**, Polytechnic University of Marche, Italy

The recent global financial crisis has favored a renewed interest in the long run view of macroeconomic history and in Minsky's contributions on the role the instability of the financial structure for the economy. We use a long-term historical dataset to investigate the timing relationships among long wave patterns in external imbalances, credit growth and asset prices. We document evidence of a recurring sequencing pattern ahead of global crisis periods where credit booms are preceded by growing external imbalances, and financial crisis occur at the low end of the contraction phase of asset price returns. The main implications of our results are the need to pay attention to widening global imbalances and to regulate and reform the financial sector by limiting capital markets movements.

CO1794: **The relationship of simple sum and Divisia monetary aggregates with real GDP and inflation: A wavelet analysis for US**
Presenter: **Michael Scharnagl**, Deutsche Bundesbank, Germany
Co-authors: Martin Mandler

We apply wavelet analysis to compare the relationship between simple sum and Divisa monetary aggregates with real GDP and CPI inflation for the US using data from 1967 to 2013. Wavelet analysis allows to account for variations in the relationships both across the frequency spectrum and across time. While we find evidence for a weaker comovement of Divisia compared to simple sum monetary aggregates with real GDP the relationship between money growth and inflation is estimated to be much tighter between Divisia monetary aggregates and CPI inflation than for simple sum aggregates, in particular at lower frequencies. Furthermore, for the Divisia indices for broader monetary aggregates (M2, M2M, MZM) we estimate a stable lead before CPI inflation of about four to five years.

CO500 Room Torrington VOLATILITY MODELS Chair: Eduardo Rossi

CO0652: **A fractionally co-integrated VAR for predicting equity and index returns**
Presenter: **Marwan Izzeldin**, Lancaster University Management School, United Kingdom
Co-authors: Xingzhi Yao

A fractionally co-integrated VAR (FCVAR) model is adopted to predict stock/index returns. The model setup allows for the co-integrating combinations of $I(d)$ processes which result in an $I(d-b)$ process, and thus has a wider appeal relative to the simplified case of $I(d)$ to $I(0)$ currently employed. The empirical analysis is based on the high frequency data for selected stocks and several index indicators (VIX, SPY and S&P 500) observed over the period of (2003-2013). The results show that predictions implied by the FCVAR model are superior to those by the conventional VAR as measured by the R-square constructed from the impulse response functions. The strength of the co-fractional relations between implied-realised volatility as well as the level of return predictability is sensitive to market regimes (pre/post crisis), structural breaks and stock/index degree of activity. The findings highlight the advantages of adopting a fractionally co-integrated framework in predicting returns, especially over long horizons, and the relevance of accounting for regime change and breaks in such a framework.

CO0996: **Inference from high-frequency data: A subsampling approach**
Presenter: **Kim Christensen**, Aarhus University, Denmark
Co-authors: Mark Podolskij, Bezirgen Veliyev, Nopporn Thamrongrat
We show how to estimate the asymptotic (conditional) covariance matrix, which appears in many central limit theorems in high-frequency estimation of asset return volatility. We provide a recipe for the estimation of this matrix by subsampling, which is based on using local stretches of high-frequency data to create rescaled miniature copies of the original statistic and then studying the sampling variation of these. We show that our estimator is consistent both in frictionless markets and models with additive microstructure noise. We derive a rate of convergence for it and are also able to determine the optimal rate of its tuning parameters (e.g., the number of subsamples). Subsampling does not require an extra set of estimators to do inference, rendering it trivial to implement. As a variance-covariance matrix estimator, it has the attractive feature that it is positive semi-definite by construction. Moreover, the subsampler is to some extent automatic, as it does not exploit knowledge about the structure of the asymptotic variance. It therefore tends to adapt to the problem at hand and be robust against misspecification of the noise process. As such, we facilitate the assessment of the sampling errors inherent in high-frequency estimation of volatility. We highlight the finite sample properties of the subsampler in a Monte Carlo simulation, while an empirical study demonstrates its use in practice to draw feasible inference about integrated variance.

CO0995: **Retrieving risk-neutral densities embedded in VIX options: A non-structural approach**
Presenter: **Paolo Santucci de Magistris**, Aarhus University, Denmark
Co-authors: Francesco Violante, Andrea Barletta
We propose a non-structural option pricing method that allows to recover the risk-neutral density of the expected volatility embedded in the call and puts VIX options. The proposed non-structural method directly recovers the risk-neutral density of the VIX future from the option prices by means of a finite orthogonal polynomial expansion around a kernel density. The polynomial expansion yields a description of the risk-neutral density without the need to specify the dynamics of volatility under the risk-neutral measure. The methodology can be thought of as an extension of the Hermite polynomial expansion of Madan and Milne to the case in which the underlying asset is a volatility and therefore is defined on the positive support. Furthermore, the method only imposes mild regularity conditions on the shape of the risk-neutral density. We prove that the polynomials are highly non-linear functions of the moments of the VIX futures.We propose a simple and robust way to estimate the coefficients of the polynomial expansion that involves an OLS regression. To control for multicollinearity/redundancy in the terms of the polynomial expansion, we rely on an orthogonalization of the regressors matrix by means of principal components. Examples based on artificial and real data are provided to support the proposed technique in a large variety of cases.

CO1258: **Measuring financial contagion: A multivariate stochastic volatility approach**
Presenter: **Xuguang Li**, Lancaster University, United Kingdom
Co-authors: Mike Tsionas, Marwan Izzeldin
We examine contagion effects resulting from the US subprime crisis on a sample of EU countries (UK, Switzerland, Netherlands, Germany and France) using a Multivariate Stochastic Volatility (MSV) framework augmented with implied volatilities. The MSV framework is estimated using Bayesian techniques and provides gains over the widely adopted DCC-GARCH framework which tends to underestimate the degree of contagion effects. Moreover, augmenting the MSV framework with implied volatilities further increases model fit. Compared with the original MSV framework, we find that the contagion effect becomes more significant. Therefore, implied volatility information is useful for detecting financial contagion or double checking some cases of market interdependence.

CO476 Room Jessel EMPIRICAL MACRO-FINANCE Chair: Cesare Robotti

CO0757: **Monetary and financial stability policy: Unconventional monetary policy, leverage and financial stress**
Presenter: **Kirstin Hubrich**, Federal Reserve Board, United States
The aim is to address the following questions: i. What is driving regime changes between normal times and high stress/crisis times? ii. Should the central bank lean against the wind by including financial stability into its reaction function? iii. Effectiveness of financial stability policy and the interaction of monetary and financial stability policy. We develop a structural vector autoregressive (SVAR) model with regime switching, extending a previous analysis in several dimensions. We develop a methodological extension of a Markov-Switching Structural VAR framework. First, we allow for the transition probabilities to be dependent on the state of the economy, and thereby to be time-varying. Second, we also facilitate more general, non-recursive structural identification restrictions. This new framework allows us to address the economic questions of high relevance for monetary and financial stability policy, in particular regarding the effectiveness of those policies in different states of the economy.

CO0827: **A likelihood-free reverse sampler of the posterior distribution**
Presenter: **Jean-Jacques Forneron**, Economics, United States
Co-authors: Serena Ng
The aim is to consider properties of an optimization based sampler for targeting the posterior distribution when the likelihood is intractable. It uses auxiliary statistics to summarize information in the data and does not directly evaluate the likelihood associated with the specified parametric model. Our reverse sampler approximates the desired posterior distribution by first solving a sequence of simulated minimum distance problems. The solutions are then re-weighted by an importance ratio that depends on the prior and the volume of the Jacobian matrix. By a change of variable argument, the output are draws from the desired posterior distribution. Optimization always results in acceptable draws. Hence when the minimum distance problem is not too difficult to solve, combining importance sampling with optimization can much faster than the method of Approximate Bayesian Computation. We estimate a model with occasionally binding constraints.

CO0861: **Two-pass cross-sectional regressions with individual stocks**
Presenter: **Cesare Robotti**, Imperial College London, United Kingdom
Co-authors: Valentina Raponi, Paolo Zaffaroni
A limiting theory for estimating and testing linear asset-pricing models with a large number of assets and a fixed time-series sample size is presented. The focus is on the modified ordinary least squares estimator of the ex-post risk premia. We derive the asymptotic distribution of this estimator and show how its limiting variance can be consistently estimated. In addition, we characterize the asymptotic distribution of a cross-sectional test of the fundamental asset-pricing relation. Finally, we show how our results can be extended to deal with an unbalanced panel. The practical relevance of our findings is demonstrated using Monte Carlo simulations and an empirical application to asset-pricing models with traded risk factors. Our analysis suggests that only the market, size, and value factors are priced consistently in the cross-section of NYSE-AMEX-NASDAQ individual stock returns. Overall, we cannot reject the null of zero risk premia for the profitability and investment factors.

CO0826: **Interest rate uncertainty and economic fluctuations**
Presenter: **Jing Cynthia Wu**, University of Chicago, United States
Co-authors: Drew Creal
Uncertainty associated with the monetary policy transmission mechanism is a key driving force of business cycles. To investigate this link, we

propose a new term structure model that allows the volatility of the yield curve to interact with macroeconomic indicators. The data favors a model with two volatility factors that capture short-term and long-term interest rate uncertainty. Increases in either of them lead higher unemployment rates, but they interact with inflation in opposite directions.

CO522 Room Montague TOPICS IN MULTIPLE TIME SERIES ANALYSIS Chair: Carsten Trenkler

CO0800: **Projection estimators for structural impulse responses**
Presenter: **Ralf Bruggemann**, University of Konstanz, Germany
Co-authors: Joerg Breitung

We discuss asymptotic inference for projection estimators of impulse responses in SVARs with short and long-run identifying restrictions. First, we provide the asymptotic distribution of a projection estimator that fully accounts for estimation uncertainty in recursive systems. Within our framework, the standard errors of the impulse responses are consistently estimated as the usual OLS standard errors and no HAC-correction is required. We also suggest a simple two-step procedure for estimation of impulse responses from SVARs with long-run identifying restrictions. Inference is based on an IV/GMM approach that takes the two-step nature of the procedure properly into account. Simulation evidence suggests that our projection methods (direct estimators) are more robust against model misspecification than traditional (iterated) estimators are.

CO1016: **Empirical characteristic functions-based estimation and distance correlation for locally stationary processes**
Presenter: **Carsten Jentsch**, University of Mannheim, Germany
Co-authors: Marco Meyer, Anne Leucht, Carina Beering

We propose a kernel-type estimator for the local characteristic function (local CF) of locally stationary processes. Under weak moment conditions, we prove joint asymptotic normality for local empirical characteristic functions (local ECF). Precisely, for processes having a (two-sided) time-varying MA(∞) representation, we establish a central limit theorem under the assumption of finite absolute first moments of the process. Additionally, we prove process convergence of the local ECF. We apply our asymptotic results to parameter estimation of time-varying distributions. Furthermore, by extending the notion of distance correlation to locally stationary processes, we are able to provide asymptotic theory for local empirical distance correlations. Finally, we provide a simulation study on minimum distance estimation for α-stable distributions and illustrate the pairwise dependence structure over time of log returns of German stock prices via local empirical distance correlations.

CO0755: **Specification of ARMA: Models with the adaptive lasso**
Presenter: **Christian Kascha**, University of Zurich, Switzerland

The properties of the adaptive lasso as a method for specifying and estimating autoregressive moving-average (ARMA) models are studied. We develop an estimation method based on the adaptive lasso and previous algorithms. We study the small sample properties of the method with a Monte Carlo simulation. We also investigate the precision of the resulting forecasts.

CO0817: **Forecasting VARs, model selection, and shrinkage**
Presenter: **Carsten Trenkler**, Universitaet Mannheim, Germany
Co-authors: Christian Kascha

The aim is to provide an empirical comparison of various selection and penalized regression approaches for forecasting with vector autoregressive systems. In particular, we investigate the effect of the system size as well as the effect of various prior specification choices on the relative and overall forecasting performance of the methods. The data set is a typical macroeconomic quarterly data set for the US. We find that these specification choices are crucial for most methods. Conditional on certain choices, the variation across different approaches is relatively small. There are only a few methods which are not competitive under any scenario. For single series, we find that increasing the system size can be helpful -depending on the employed shrinkage method.

CO368 Room Chancellor's Hall BOOTSTRAP METHODS FOR TIME SERIES Chair: Peter Boswijk

CO0892: **Adaptive information-based methods for determining the co-integration rank in heteroskedastic VAR models**
Presenter: **Luca De Angelis**, University of Bologna, Italy
Co-authors: Giuseppe Cavaliere, Robert Taylor, Anders Rahbek

Standard methods for determining the co-integration rank of vector autoregressive (VAR) systems of variables integrated of order one are affected by the presence of heteroskedasticity with sequential procedures based on Johansen's (pseudo-) likelihood ratio test being significantly over-sized in finite samples and even asymptotically. Notable solutions to this problem are the wild bootstrap applied to the traditional likelihood ratio test or an information criterion such as standard BIC. However, although asymptotically valid, these methods may show low power in small samples as they do not exploit the potential efficiency gains provided by the adaptation with respect to the volatility process. Therefore, adaptive methods where the covariance matrix is estimated nonparametrically can be particularly useful in the determination of the co-integration rank in VAR models driven by heteroskedastic innovations as they exploit the power gain potential, especially in the presence of nonstationary unconditional volatility. We show that adaptive information criteria are weakly consistent provided the usual conditions on the penalty term hold and display better finite sample results in many situations.

CO1230: **Wild bootstrap methods for heteroskedastic times series with leverage**
Presenter: **Peter Boswijk**, University of Amsterdam, Netherlands

The purpose is to study the use of various implementations of the bootstrap for inference on the mean of a time series with conditional variance displaying statistical leverage effects. These imply a nonlinear dependence in the time series, which is not reproduced by the wild bootstrap (nor by the i.i.d. bootstrap). Therefore, we consider various dependent versions of the wild bootstrap, and compare their effectiveness with parametric GARCH-based bootstrap methods. If the time series displays stationary GARCH with leverage effects, then standard inference is asymptotically valid, but suitable bootstrap implementations can be expected to yield improved finite-sample inference. On the other hand, if the heteroskedasticity is characterised by near-integrated GARCH or stochastic volatility, then standard inference methods are distorted, even asymptotically. We also study the effectiveness of the bootstrap in such cases.

CO1259: **Tests of the co-integration rank in VAR models in the presence of a possible break in trend at an unknown point**
Presenter: **Robert Taylor**, University of Essex, United Kingdom
Co-authors: Steve Leybourne, Dave Harvey

We consider the problem of testing for the co-integration rank of a vector autoregressive process in the case where a trend break may be present in the data. Un-modelled trend breaks can result in tests which are incorrectly sized under the null hypothesis and inconsistent under the alternative hypothesis. Extant procedures have attempted to solve this problem, but require to assume either that the trend break date is known or that any trend break cannot occur under the co-integration rank null hypothesis being tested. These procedures also assume the autoregressive lag length is known. Moreover in each of these strands there is also a presumption in calculating the tests that a trend break is known to have happened. This can lead to a substantial loss in finite sample power in the case where a trend break does not in fact occur. Using information criteria based methods to

select both the autoregressive lag order and to choose between the trend break and no trend break models, using a consistent estimate of the break fraction in the context of the former, we develop a number of procedures which deliver asymptotically correctly sized and consistent tests of the co-integration rank. By selecting the no break model when no trend break is present, these procedures also avoid the potentially large power losses associated with the extant procedures in such cases.

CO1368: **Bootstrap applied to ARCH models**
Presenter: **Anders Rahbek**, University of Copenhagen, Denmark
Co-authors: Giuseppe Cavaliere, Heino Bohn Nielsen
We investigate the properties of applying the bootstrap for inference in conditional volatility models. Emphasis will be put on non-standard inference results and establishing validity of the bootstrap under non-standard conditions.

CO356 Room G21A MODELLING AND FORECASTING TEMPORAL DATA Chair: Alessandra Amendola

CO1176: **Markov switching autoregressive forecasts: Symmetric or asymmetric loss functions**
Presenter: **Fabio Forte**, University of Salerno, Italy
Co-authors: Marcella Niglio
In nonlinear time series domain the forecast accuracy has been largely debated. It has been shown that, in some cases, the complex structure of nonlinear models does not guarantee more accurate predictions with respect to those obtained from the linear models of the ARMA class. The reasons can be differently attributed. Among them: the selected model is not able to catch the features of the series under analysis; the loss function selected to generate the predictors needs to be revised. We focus the attention on the second problem. In more detail, we investigate the forecast ability of some nonlinear models with switching structure. After showing how their predictors can be obtained by using symmetric and asymmetric loss functions (of the LinEx class) we investigate and discuss if (and when) the introduction of asymmetric loss functions can be of help to increase the forecast accuracy.

CO1712: **Mixture memory models for realized volatility**
Presenter: **Giovanni De Luca**, University of Naples Parthenope, Italy
Co-authors: Giampiero Gallo
The dynamics of financial volatility shows a behavior characterized by alternating periods of turbulence and relative quiet. We suggest to model it as a mixture memory model where time-varying mixing weights are a function of some forcing variable capable of sudden changes. In choosing a mixture approach we rely on previous evidence on the presence of a short and a long memory component in the observed series. We apply our model to the main Spanish and Italian stock index (IBEX and FTSEMIB, respectively). For the selection of the most appropriate forcing variable, we have analyzed the spread between the sovereign national and German bond rates as well as some variables measuring the German volatility. The results show a good performance in sample, pointing to the fact that fixed weights may be a limitation to an accurate description of volatility behavior.

CO1047: **Realized wavelet-based estimation of integrated covariance and co-jumps in the presence of noise**
Presenter: **Lukas Vacha**, FSV Charles University in Prague, Czech Republic
Co-authors: Jozef Barunik
Wavelet-based estimator of integrated covariation is proposed. Basing our estimator in two-scale covariance framework, we are able to utilize all available data and get unbiased estimator in the presence of noise as well. In addition, we focus on detection of co-jumps, and our estimator is able to distinguish between idiosyncratic jumps and co-jumps. The estimator is tested in a small sample numerical study, and is compared to other popular integrated covariation estimators under different simulation settings with changing noise as well as jump and co-jump levels. The results reveal that our wavelet-based estimator is able to estimate the realized measures with the greatest precision. Another notable contribution lies in the application of the presented theory. Our time-frequency estimators not only produce more efficient estimates, but also decompose the integrated covariation into arbitrarily chosen investment horizons.

CO1223: **Robust portfolio sorts**
Presenter: **Walter Distaso**, Imperial College London, United Kingdom
Co-authors: Valentina Corradi, Daniela Alifano
We revisit the ubiquitous practice (by both academics and practitioners) of creating portfolios by sorting financial assets according to a given variable. The sorting is usually done brute force and ignores the estimation error present in the measurement of the sorting variable. We propose a simple procedure to control for this and show that ignoring this error may produce a substantial classification error (i.e. for example stocks are included in the first decile portfolio when they should belong to the second decile one).

CO586 Room Woburn VOLATILITY MODELS AND THEIR APPLICATIONS Chair: Yasuhiro Omori

CO0977: **Volatility and quantile forecasts by realized stochastic volatility models with generalized hyperbolic distribution**
Presenter: **Makoto Takahashi**, Osaka University, Japan
The predictive performance of a previous realized stochastic volatility model, which incorporates the asymmetric stochastic volatility model with the realized volatility, is investigated. Considering well-known characteristics of financial returns, heavy tail and negative skewness, the model is extended by employing a wider class distribution, the generalized hyperbolic skew Student's $t-$distribution, for financial returns. With the Bayesian estimation scheme via Markov chain Monte Carlo method, the model enables us to estimate the parameters in the return distribution and in the model jointly. It also makes it possible to forecast volatility and return quantiles by sampling from their posterior distributions jointly. The model is applied to quantile forecasts of financial returns such as value-at- risk and expected shortfall as well as volatility forecasts and those forecasts are evaluated by various tests and performance measures. Empirical results with the US and Japanese stock indices, Dow Jones Industrial Average and Nikkei 225, show that the extended model improves the volatility and quantile forecasts especially in some volatile periods.

CO1232: **Dynamic Bayesian forecasting of multivariate volatilities and portfolio decisions**
Presenter: **Lutz Gruber**, Duke University, United States
Co-authors: Mike West
The recently introduced simultaneous graphical dynamic linear model (SGDLM) facilitates on-line Bayesian analysis of high-dimensional time series. We discuss a case study in financial forecasting and portfolio optimization using the SGDLM applied to a series of 400 daily stock prices from the S&P500 over 2007-2014. Our primary interest lies in sequential step-ahead forecasting of multivariate across-series co-volatilities; a secondary interest lies in application to portfolio decisions. We show the SGDLM to forecast volatilities and across-series co-volatilities substantially more accurately than the standard multivariate DLM, and we are able to substantially improve portfolio outcomes across a range of portfolio utility functions.

CO1234: **Dynamic block-equicorrelation, realized stochastic volatility and cross leverage**
Presenter: **Yuta Kurose**, Kwansei Gakuin University, Japan
Co-authors: Yasuhiro Omori
Multivariate daily returns and realized measures are simultaneously modeled in multivariate realized stochastic volatility model with dynamic block-equicorrelation and cross leverage effect. Using a state space representation, we propose a Bayesian estimation algorithm implemented by Markov chain Monte Carlo (MCMC) method. With additional information to estimate unobserved variables by using the realized measures, we can obtain an estimation result efficiently by the simple algorithm. Numerical examples are provided and the proposed model is applied to the multivariate daily stock price data.

CO1287: **Realized stochastic volatility model with multiple realized measures**
Presenter: **Tsunehiro Ishihara**, Osaka University, Japan
We extend the stochastic volatility model with realized measures of volatility (realized stochastic volatility model-RSV model). We incorporate the multiple realized measures. Our particular interest is how to introduce the different type realized measures simultaneously, e.g. realized volatility with and without jump parts, and semivariance. Moreover, we consider the link of the volatility latent variables and the mean of the return equations. Proposed models are applied for the daily returns and several realized measures. Several models are compared with the predictive performance.

CO1849: **Combining multivariate volatility forecasts: An economic-based approach**
Presenter: **Andre Portela Santos**, Universidade Federal de Santa Catarina, Brazil
Co-authors: Guilherme Moura, Joao Caldeira, Francisco J Nogales
We devise a novel approach to combine predictions of high dimensional conditional covariance matrices using economic criteria based on portfolio selection. The combination scheme takes into account not only the portfolio objective function but also the portfolio characteristics in order to define the mixing weights. Three important advantages are that i) it does not require a proxy for the latent conditional covariance matrix, ii) it does not require optimization of the combination weights, and iii) it holds the equally-weighted model combination as a particular case. Empirical applications involving three large data sets from different markets show that the proposed economic-based combinations of multivariate GARCH forecasts leads to mean-variance portfolios with higher risk-adjusted performance in terms of Sharpe ratio as well as to minimum variance portfolios with lower risk on an out-of-sample basis with respect to a number of benchmark specifications.

CO558 Room Bloomsbury BEHAVIOURAL AND EMOTIONAL FINANCE: THEORY AND EVIDENCE I Chair: Richard Fairchild

CO1122: **Entrepreneurial under diversication, over optimism and overcondence: Theory and evidence**
Presenter: **Enrico Maria Cervellati**, University of Bologna, Italy
Co-authors: Pierpaolo Pattitoni, Marco Savioli
We want to explain how overconfidence and over optimism lead entrepreneurs to overinvest in their companies, underestimating risks and overestimating expected returns. We propose a theoretical model where entrepreneurs choose which part of their wealth to invest in their private company and which one in the stock market. Overconfidence and over optimism are parameters in our model, and they bias entrepreneurs portfolio allocation. With a simulation analysis, we calculate overconfidence and over optimism levels implicit in the entrepreneurs observable portfolio, instead of using proxies or indirect measures. We use these measure of entrepreneurial under-diversification in an empirical analyses based on data on 2295 SMEs in Italy. Our data come from a detailed survey performed between 2008 and 2009. This unique database combines both entrepreneurs personal information, and on their companies. We are able to investigate the degree to which entrepreneurs overconfidence underestimate the riskiness associated to their business. Because entrepreneurs in our sample invest about 40% of their household wealth in their company, overinvesting in their firms leads to bear too much risk, that may eventually be transferred to their household, and affect their wealth.

CO1027: **Visceral emotions, within-community communication and (ill-judged) endorsement of financial propositions**
Presenter: **Kim Kaivanto**, Lancaster University, United Kingdom
The 2007-08 financial crisis exposed poignant examples of ill-judged risk accretion in both tails of the Lorenz curve: concentrations of inappropriate mortgages within low-income neighborhoods, and concentrations of Bernard Madoff's victims within wealthy, predominantly Jewish country-club communities. These examples share three key elements. First, individual behavioral decision makers take decisions privately but contribute to the build-up of risk within the community. Second, sales agents employ psychological persuasion techniques (bypassing logical processes), and trigger visceral emotions (overriding rational deliberation). Third, community membership immerses individuals within information flows that trigger invidious visceral emotions, and leads to biased inferences due to sample-size illusion and persuasion bias. We develop a closed-form model based on Signal-Detection Theory (SDT) that incorporates all three above-mentioned elements: it is behavioral in employing a Prospect Theory (PT) objective function; peripheral-route persuasion and visceral emotions are incorporated through their impacts on discriminability d'; and sample-size illusion and persuasion bias are incorporated through their effects on the score θ. This PT-SDT model predicts that visceral-emotion-charged hot states can short-circuit the capacity to practice caveat emptor, carrying implications for regulation and for our understanding of US household-borrowing growth 2001–2006.

CO0906: **Decision-making: Conceptual levels of examination**
Presenter: **Graham Mallard**, Cheltenham College, United Kingdom
The fundamental criticism levelled at behavioural economics is that, having not yet generated an overarching conceptual framework, it inevitably leads to works grounded on rather arbitrary and ad hoc behavioural assumptions. An overarching framework, then, is of paramount importance if the field is to present a feasible alternative to the standard rational choice approach. Adopting this stance as the starting point, a discussion of the conceptual levels at which decision making can be examined is presented. The standard approach in economics is to focus on the level of choice: the level at which an agent makes a selection from amongst a specified range of options, be they goods, bundles, lotteries or actions. The higher order level, in contrast, is that at which an agent decides the level of precision with which he subsequently makes such a particular choice. Based on the growing evidence of decision fatigue within the literature, it is proposed that (1) to assume global rationality at the level of choice is internally inconsistent, and (2) examining decision making behaviour at the higher order level rather than at the level of choice is a potentially fruitful avenue of future research, potentially overcoming the fundamental criticism of behavioural economics identified above. The possible nature of such study is discussed, drawing on a range of econometric analysis from the field of marketing as an example.

CO0804: **Joint audit, competence, independence, judgement, scepticism, bias mitigation, audit quality**
Presenter: **Oliver Marnet**, University of Southampton, United Kingdom
Joint audits have been identified as a potential measure to address deficiencies in the audit function identified as an outcome of investigations into the 2008-2009 global financial crisis. However, possible adoption of joint audits is a highly controversial issue. Arguments against their implementation are primarily based on concerns about cost and a perceived lack of evidence on their benefits. Arguments in their favour focus on their ability to enhance audit quality. The debate is widen by reflecting on joint audits from a behavioural perspective. Auditors operate at the centre of a complex interaction between heuristics and biases which can affect the quality of judgement and decision-making that underpin the audit opinion. It is suggested that the impact of social and psychological factors on auditor independence and competence is less pronounced under joint

audit arrangements than for a single engagement team. Theoretical frameworks derived from behavioural research are evaluated to guide future applied studies on the potential of joint audits in bias mitigation and enhancement of scepticism.

CC025 Room Athlone CONTRIBUTIONS IN PANEL DATA Chair: Josep Lluis Carrion-i-Silvestre

CC1440: **Empirical likelihood-based inference for categorical varying coefficient panel data models with fixed effects**
Presenter: **Luis Antonio Arteaga**, Universidad de Cantabria, Spain
Co-authors: Juan Manuel Rodriguez-Poo

We propose an empirical likelihood-based method to compute confidence bands based on non-parametric estimators of fixed effects varying coefficient panel data models that might include categorical covariates. In order to do so, a non-parametric version of the Wilk's theorem is derived. Furthermore, we obtain the mximum empirical likelihood estimator of the varying parameters. We also provide its asymptotic properties. Compared with normal approximation techniques, the confidence bands based on the empirical likelihood achieve more accuracy and have two main advantages: (1) confidence bands automatically reflect emphasis in the observed dataset, which means they adjust to the true shape of the underlying distribution; (2) no explicit covariance matrix estimator is needed, and as a result they are robust against heterocedasticity. To illustrate the proposed method we also present some simulation results.

CC1509: **Hurdle models of repayment behaviour in personal loan contracts**
Presenter: **Jose Murteira**, FEUC and CEMAPRE, Portugal
Co-authors: Mario Augusto

A hurdle model of repayment behavior in personal loans with fixed installments is proposed. Under plausible assumptions, a debtor in each period faces a number of missed payments, depending on his previous repayment decisions; meanwhile, as most debtors are expected to meet their financial obligations, the number of missed payments is bound to display excess zeros, with reference to a single-part law. Each sequence of missed payments is modeled by using the binomial thinning, a conceptual tool that allows for dependence between integers by defining the support of consecutive counts. Nonlinear least squares, quasi-maximum likelihood and maximum likelihood are viable estimators of the model, also viewed as a random effects model, under suitable assumptions on heterogeneity, which yield a two-part panel data model, estimable by quasi-maximum likelihood. The proposed approach is illustrated with a panel data set on personal loans granted by a Portuguese bank.

CC1821: **Corporate social responsibility performance and firm performance in the global shipping industry**
Presenter: **Dimitris Tsouknidis**, Cyprus University of Technology, Cyprus
Co-authors: Manolis Kavussanos

The aim is to investigate whether corporate social responsibility (CSR) performance causes higher firm performance in the global shipping industry. By utilizing a global sample of 33 shipping firms over the period 2002 to 2013, we provide strong evidence of a positive relationship between firm performance and CSR. In addition the study corrects for the fact that previous studies in the general finance literature and in the shipping literature, suffer from endogeneity problems such as simultaneity (reverse causality) and unobserved heterogeneity. This is achieved through the estimation of a dynamic panel (system) generalized method of moments (GMM) estimator, which takes into account that endogeneity often arises with a time lag (dynamic endogeneity), leading to biased results and doubtful inferences. Results reveal the financial benefits of engaging in CSR activities for the shipping sector and contribute to a better understanding of the governing mechanisms between financial performance and CSR. Thus, devoting resources on increasing CSR performance is expected to yield significant future financial benefits for shipping companies worldwide.

CC1741: **Shrinkage PCA for efficient estimation of large approximate factor models**
Presenter: **Rachida Ouysse**, University of New South Wales, Australia

A new approach is developed for the efficient estimation of large dimensional approximate factor models. Efficient estimation of factor models is attracting considerable attention, because recent empirical evidence suggests the estimates are adversely affected by the inability to account for the cross sectional dynamics. A factor structure is approximate when the idiosyncratic errors are weakly correlated across the variables. Principal components analysis (PCA) provides consistent estimation of the factor structure and efficiency can be achieved using robust econometric tools such as generalized PCA and quasi maximum likelihood. However when $N > T$, the sample covariance matrix is singular and accounting for cross-sectional dynamics is challenging without imposing a structure on these dynamics. We propose to use the approximate structure assumption of bounded cross-section correlation as a constraint in the PCA framework. The proposed penalized PCA can be interpreted as a shrinkage regression where the off diagonal elements of the covariance matrix are shrunk towards zero as N grows large. The new approach performs well in a series of Monte carlo simulations against PCA and other competing alternatives.

EO132 Room CLO 204 STATISTICS IN IMAGING Chair: Michele Guindani

EO0174: **A scalable multi-resolution model for activation and spatial connectivity in fMRI data**
Presenter: **Stefano Castruccio**, Newcastle University, United Kingdom
Co-authors: Marc Genton, Hernando Ombao

The aim is to introduce a scalable, multi-resolution spatio-temporal model and a computationally efficient methodology to estimate cognitive control related activation and whole-brain connectivity. The proposed model allows to test for voxel-specific activation of Blood-Oxygen-Level-Dependent response while accounting for a non-stationary local spatial dependence within pre-specified Regions Of Interest (ROIs), as well as global (among ROIs) dependence. Further, the model allows us to detect interpretable activation patterns among ROIs via graphical LASSO estimation of the inverse covariance matrix. The model is applied on an fMRI data set with 210,000 voxels per time frame, for a total of 30 million data points, using high performance clusters to parallelize the inference.

EO1143: **Estimating sparse graphical models for neuroimaging data**
Presenter: **Ivor Cribben**, Alberta School of Business, Canada

Graphical models are frequently used to explore networks among a set of variables. Several methods for estimating sparse graphical models have been proposed and their theoretical properties have been explored. However, their practical performance has not been studied in detail. First, we compare several estimation procedures (graphical lasso, SCAD, DP-glasso, ...) and several selection criteria (AIC, BIC, CV, ...) using both simulated multivariate normal data and autocorrelated data. We use evaluation criteria to compare the models and thoroughly discuss the superiority and deficiency of each of them. Second, we consider methods that jointly estimate functional connectivity (or graphs) for all subjects. These graphs borrow strength across subjects that share certain network characteristics. Finally, we estimate the functional connectivity networks between regions of interest (ROIs) from a resting state functional magnetic resonance imaging (fMRI) experiment and from a language processing experiment using the various procedures. The estimation of brain networks allows for a deeper understanding of the functional architecture and large scale characterizations of the brain.

EO1262: **Statistical methods for joint structural-functional connectomes**
Presenter: **Russell Shinohara**, University of Pennsylvania Perelman School of Medicine, United States

We propose statistical methods for quantifying variability in a population of connectomes using general weighted graphical representations. While neuroimaging analyses of graphical models center almost entirely on scalar summaries, little work has been done comprehensively studying populations of graphs. To study this, we propose generalized variances for complex objects based on distance statistics and develop methods for two-sample testing at the whole connectome and the subnetworks. We demonstrate the utility of these methods in a connectomic study of autism spectrum disorders using diffusion tensor imaging and magnetoencephalography.

EO1448: **A Bayesian approach to the study of dynamic functional connectivity networks in fMRI data**
Presenter: **Michele Guindani**, MD Anderson Cancer Center, United States

fMRI studies have traditionally assumed stationarity of the connectivity patterns observed in a subject during a fMRI experiment. While the assumption has successfully allowed to study large-scale properties of brain functioning, it is generally recognized that functional connectivity varies with time and tasks performed. We describe a novel Bayesian methodological framework for the analysis of temporal dynamics of functional networks in task-based fRMi data collected on a single subject. Our proposed formulation allows joint modeling of the task-related activations in addition to the dynamics of individual functional connectivity. Furthermore, we allow simultaneous learning of the common and differential edges (interactions) in the inferred time-varying functional networks. We illustrate the proposed approach by means of simulation and an analysis on a real fMRI dataset.

EO220 Room MAL 539 TEXT MINING, NETWORK DATA AND MIXTURE MODELING Chair: Annie Qu

EO0229: **Joint estimation of multiple Granger causal network models**
Presenter: **George Michailidis**, U of Florida, United States

Network Granger causality focuses on estimating the corresponding Granger causal effects from p-time series and it can be operationalized through Vector Autoregressive Models (VAR). The latter represent a popular class of time series models that has been widely used in applied econometrics and finance and more recently in biomedical applications. We discuss joint estimation and model selection issues of multiple Granger causal networks. We present a modeling framework for the setting where the same variables are measured on different entities (e.g. same set of economic activity variables for related countries). The framework involves the introduction of appropriate structural penalties on the transition matrices of the respective VAR models that link the underlying network Granger models and sparse covariance matrices for capturing latent idiosyncratic factors. A fast estimation strategy is presented and the model is evaluated on both synthetic and real macroeconomic data.

EO0382: **Classification with unstructured predictors and its application to sentiment analysis**
Presenter: **Junhui Wang**, City University of Hong Kong, China

Unstructured data refers to information that lacks certain structures and cannot be organized in a predefined fashion. Unstructured data involve heavily on words, texts, graphs, objects or multimedia types of files that are difficult to process and analyze by traditional computational tools and statistical methods. We discuss ordinal classification with unstructured predictors and ordered class categories, where imprecise information concerning strengths between predictors is available for predicting the class labels. We integrate the imprecise predictor relations into linear relational constraints over classification function coefficients, where large margin ordinal classifiers are introduced, subject to quadratically many linear constraints. The proposed methods are implemented via a scalable quadratic programming algorithm based on sparse word representations. The advantage is demonstrated in a variety of simulated experiments as well as one large-scale sentiment analysis example on TripAdvisor.com customer reviews. If time permits, the asymptotic properties will also be discussed, which confirm that utilizing relationships among unstructured predictors can significantly improve prediction accuracy.

EO1834: **Inference for logistic-normal mixtures with heterogeneous components**
Presenter: **Juan Shen**, Fudan University, China

Subgroup analysis with unspecified subgroup memberships has received much attention in recent years. We firstly propose a structured logistic-normal mixture model to model the subgroup distributions and the subgroup membership simultaneously, but under the assumption that the subgroups differ only in the means. Next, we consider a penalized likelihood approach for more general cases with heterogeneous subgroup variances. Empirical results with a simulation study and two real data examples demonstrate the usefulness of the proposed method.

EO647 Room MAL B34 DEPENDENCE MODELS AND COPULAS III Chair: Wolfgang Trutschnig

EO0311: **Failure probabilities and Hazard scenarios in environmental sciences: A multivariate perspective via copulas**
Presenter: **Gianfausto Salvadori**, Universita Del Salento, Italy

Several guidelines, released by international agencies, have recently recommended to individuate suitable Hazard scenarios in order to deal with extreme natural phenomena. In particular, the scenarios should be multivariate, taking into account the fact that the variables at play may not be independent. A Hazard scenario is characterized by (i) a specific geometry, and (ii) a suitable probability distribution. Several scenarios are investigated, and due comparisons are carried out. In addition, it is shown how the notion of Failure Probability, a tool traditionally used for design and risk assessment in engineering practice, is well suited to cope with the illustrated approach. All the results are based on the mathematics of Copulas, which turns out to be a fundamental theoretical apparatus for doing multivariate risk assessment. A possibly non-stationary hydrological case study is used to illustrate the procedures. Preliminary results concerning the construction of a multivariate conditional framework are also shown, to cope with multiple sources of risk (as it often happens, e.g., in environmental engineering).

EO0435: **Directional approach to identify extremes**
Presenter: **Raul Andres Torres Diaz**, Universidad Carlos III de Madrid, Spain
Co-authors: Rosa Lillo, Henry Laniado Rodas, Carlo De Michele

Extremes identification is a very important task in fields as insurance, finance, economy and environmental science. It has been widely studied in the univariate setting but events such as economic crisis and environmental disasters suggest that it is necessary to analyze extremes considering the joint information provided by all the variables implied in the problem. This fact motivates the extension of some typical concepts of the extreme value theory to the multivariate framework. Specifically, the quantile notion is of special interest for extremes recognition, although, the lack of a total order in R^n is a serious drawback. However, some definitions of multivariate quantiles have been introduced through partial orders or generalizing a specific property of the univariate quantile. One example is the case of the definition of quantile based on copulas that allows to obtain closed and analytics expressions of the quantiles easy to manage in practice. The aim is to highlight that the classic analysis of extremes using copula could be improved using a directional approach based on a non parametric estimation. We provide analytical properties and relationships between our proposal and the methods based on copulas. Finally, we illustrate the methodology with a dataset used in hydrology.

EO0341: **Calculating the Linfoot correlation of Archimedean copulas**
Presenter: **Soeren Moeller**, University of Southern Denmark, Denmark
Co-authors: Thomas Scheike, Jacob Hjelmborg
Linfoot's information theory based measure of correlation is one of the few dependence measure which fulfils all seven desired properties for a dependency measure. Still the Linfoot correlation has not come to widespread use in practice, as it is relatively hard to calculate. We prove that the Linfoot correlation of a multivariate distribution only depends on its copula, and not on its marginals. In the special case of Archimedean copulas, this enables us to express the correlation by the generator of the Archimedean copula. We find closed expressions for the Linfoot correlations of the Clayton copula, and compare it to Kendall's tau and Spearman's rho for this copula. We also compare it to the Pearson correlation, assuming differently distributed marginals, as the Pearson correlation depends on the marginals as well as the copula. Furthermore, modern computer power enables us, to estimate the Linfoot correlation of bivariate data non-parametrically via the empirical copula, as well as using kernel estimators and a k-nearest neighbor approach and compare these estimates with the estimates obtained from fitting an Archimedean copula to the data. We apply these methods to Danish twin data on lifespan and discuss future applications of these results to the development of a more robust measure for biometric heritability.

EO0841: **Extensions of subcopulas**
Presenter: **Fabrizio Durante**, Free University of Bozen-Bolzano, Italy
Copulas represent one of the building blocks of modern multivariate analysis since it was shown that the probability law of any random vector can be expressed as a composition of the distribution functions of all one-dimensional margins and a suitable copula. However, while the copula associated with a random vector is unique when the margins are continuous, in the non-continuous case, various copulas can be associated with the same vector, all being coincident in a subset of the copulas domain. This fact poses the natural question of how it is possible to construct a copula given some partial information about the values that it assumes. One of the most common extension procedures is given by the multilinear interpolation (or checkerboard construction), which also plays a central role in characterizing dependence concepts for discrete random vectors. Extension procedures are presented in a high-dimensional framework to transform a specific subcopula to a copula. Moreover, convergence results are given in order to check how these extensions approximate (in different metrics) a given target copula.

EO258 Room MAL B36 ROBUST METHODS Chair: Stefan Van Aelst

EO0329: **The rtclust procedure for robust clustering**
Presenter: **Alessio Farcomeni**, Sapienza - University of Rome, Italy
Co-authors: Francesco Dotto, Luis Angel Garcia-Escudero, Agustin Mayo-Iscar
A new robust model based clustering method is proposed, which is based on trimming and reweighting. In parallel with the case of location and scatter estimation, reweighting allows to achieve high breakdown together with high efficiency. Additionally, the user does not need to set the trimming level in advance. The method proceeds by fitting an initial robust finite mixture model based on a high trimming level. Observations with high robust Mahalanobis distance from the closest centroid are then flagged as outlying, and centroid estimation is repeated based on the other ones. The procedure is iterated until convergence. It is shown formally, with examples, and with extensive simulation studies that the resulting rtclust procedure can resist to different outlier generating schemes, and is highly efficient in the presence of little or no contamination. Additionally, the new procedure compares well and does not need much tuning.

EO0677: **Multichannel sparse approximations of complex-valued signals using Huber's criterion**
Presenter: **Esa Ollila**, Aalto University, Finland
Huber's criterion is generalized to multivariate regression and multichannel sparse recovery problem of complex-valued measurements where the objective is to find good recovery of jointly sparse unknown signal vectors from the given multiple measurement vectors which are different linear combinations of the same known elementary vectors. This requires careful characterization of robust complex-valued loss functions as well as Huber's criterion function for the multivariate sparse regression problem. We devise a greedy algorithm based on simultaneous normalized iterative hard thresholding (SNIHT) algorithm. In the developed algorithm, referred to as HUB-SNIHT, novel adaptive stepsize selection is performed for both the scale and signal matrix part at each iteration which is shown to speed up convergence considerably. Unlike the conventional SNIHT method, our method is robust under heavy-tailed non-Gaussian noise conditions, yet has a negligible performance loss compared to SNIHT under Gaussian noise. Theoretical recovery guarantees are discussed and usefulness of the method is illustrated in source localization application with sensor arrays and in image denoising problems.

EO1296: **Robust testing and confidence regions in high-dimensional quantile regression**
Presenter: **Tianqi Zhao**, Princeton University, United States
Co-authors: Mladen Kolar, Han Liu
We propose a robust procedure for assessing uncertainties of parameter estimation in high-dimensional linear models, where the dimension can grow exponentially fast with the sample size. Our first contribution is to develop a new de-biasing framework tailored for non-smooth loss functions, under rather weak conditions on the design and noise term. Moreover, the framework enables us to exploit the composite quantile function to construct a de-biased CQR estimator. This estimator is robust, and preserves efficiency such that the worst-case efficiency loss is less than 30% compared to square-loss-based procedures. In many cases our estimator is close to or better than the latter. Furthermore, we show that the computational complexity of the de-biased CQR estimator is similar to that of the de-biased single quantile estimator, so that the improvement in statistical performance is almost free. Our second contribution is to prove the finite sample properties of the L_1-penalized CQR estimator, and show that it achieves the minimax rate of convergence. The proof of the results relies on applying modern empirical process theories. Our framework can be extended to broader applications such as the high-dimensional simultaneous test and divide-and-conquer inference. We provide empirical results to support the theories.

EO1324: **Fast and robust bootstrap of robust estimators in SUR models**
Presenter: **Kris Peremans**, KU Leuven, Belgium
Robust estimators of the seemingly unrelated regression model are considered. First, S-estimators are studied which can attain a high breakdown value, but their normal efficiency can be quite low. Therefore, MM-estimators are introduced to obtain estimators that have both a high breakdown value and a high normal efficiency. Furthermore, the problem of statistical inference is studied. Classical inference relies on assumptions which are hard to verify. Moreover, classical inference may be non-robust. Therefore, a fast and robust bootstrap procedure is developed. Bias corrected and accelerated confidence intervals of the estimated parameters are constructed and their performance is analyzed in simulation studies. The robust estimators and the fast and robust bootstrap procedure are illustrated on some examples.

EO272 Room MAL 402 MULTIVARIATE ANALYSIS AND CHANGE-POINT DETECTION Chair: Daniel Vogel

EO0355: **Sequential bootstrap procedures for non-degenerate U-statistics under mixing with applications to change-point detection**
Presenter: **Ivan Kojadinovic**, University of Pau, France
Co-authors: Axel Buecher
Resampling schemes for two sequential processes constructed from non-degenerate U-statistics will be presented along with theoretical results establishing their asymptotic validity under mixing conditions. The proofs exploit recent results on degenerate U-statistics. A data-driven approach for estimating a key bandwidth parameter involved in the bootstrapping procedures is also suggested, making the use of the resampling schemes fully automatic. The derived results can for instance be used to derive tests for detecting changes in Gini's mean difference or Kendall's tau. For such applications, Monte Carlo experiments suggest that the use of the proposed bootstrap procedures can have advantages over that of estimated asymptotic distributions.

EO0953: **Asymptotic distribution for a class of general change-point tests and estimators**
Presenter: **Stefanie Schwaar**, University of Kaiserslautern, Germany
Many data sets exhibit changes at some unknown point in time. Therefore, we are interested in obtaining estimators as well as their asymptotic distribution in such a setting. To this end, we give examples of models with at most one change and describe a general class of change-point tests. Based on these generalised change-point tests we derive generalised estimators for the change-point, whose asymptotic behaviour we are interested in. We are able to prove that these estimators are asymptotically consistent for many alternatives with rate $O_P\left(\frac{1}{n}\right)$. We can show that this rate cannot be improved.

EO0940: **Multivariate procedures for detection of changes based on Fourier method**
Presenter: **Marie Huskova**, Charles University, Czech Republic
We propose and study procedures for detection of changes in two different setups. Procedures are based on empirical characteristic functions. At first we deal with off-line procedures for a detection of changes in distribution in a series of independent multivariate observations. Then we focus on procedures for monitoring changes in the error distribution of autoregressive time series when the overall size of the sequential test. Both under the null hypothesis (no change) as well as alternatives (there is a change) limit properties of the proposed test procedures are studied. They are accompanied by a simulation study.

EO1356: **Testing for structural breaks via ordinal pattern dependence**
Presenter: **Herold Dehling**, Ruhr-University Bochum, Germany
Co-authors: Alexander Schnurr
We propose new concepts in order to analyse and model the dependence structure between two time series. Our methods rely exclusively on the order structure of the data points. Hence, the methods are stable under monotone transformations of the time series and robust against small perturbations or measurement errors. Ordinal pattern dependence can be characterised by four parameters. We propose estimators for these parameters, and we calculate their asymptotic distributions. Furthermore, we derive a test for structural breaks within the dependence structure. All results are supplemented by simulation studies and empirical examples.

EO084 Room MAL B35 MULTIPLE TESTS II Chair: Jacobo de Una-Alvarez

EO0391: **New procedures controlling the false discovery proportion under dependence**
Presenter: **Etienne Roquain**, University Pierre et Marie Curie, France
Co-authors: Sylvain Delattre
The false discovery proportion (FDP) is a convenient way to account for false positives when a large number of tests are performed simultaneously. A general principle has been proposed that builds FDP controlling procedures from k-family-wise error rate controlling procedures while incorporating dependencies in an appropriate manner. We introduce suitable modifications of this heuristic that provide new methods overcoming the existing procedures with a proven FDP control.

EO0997: **Union-intersection based goodness-of-fit tests in terms of local levels**
Presenter: **Veronika Gontscharuk**, German Diabetes Center - Leibniz Institute for Diabetes Research at Heinrich-Heine-University Duesseldorf, Germany
Co-authors: Helmut Finner
Union-intersection based goodness-of-fit (GOF) tests are related to multiple test procedures in the sense the underlying GOF null hypothesis can often be represented as an intersection of suitable elementary hypotheses. An interesting characteristic of such GOF tests are so-called local levels. A local level is defined as the probability to reject an elementary hypothesis under the global null. Many GOF tests, e.g., the Kolmogorov-Smirnov test and the supremum version of the Anderson-Darling test as well as GOF tests based on phi-divergences, can be represented in terms of local levels. This allows a simple comparison of tests with respect to regions of sensitivity. Moreover, prior information about possible alternative distributions may be incorporated into the choice of local levels in order to tailor new GOF tests. We mainly focus on GOF tests with equal local levels. We provide various representations of these tests and discuss their relationship to several well-known GOF procedures. Thereby, the asymptotics of local levels is of special interest.

EO1339: **FDR control in Hierarchical hypothesis testing under dependence**
Presenter: **Anat Reiner-Benaim**, University of Haifa, Israel
If one main hypothesis is tested for numerous cases, it may be of interest to perform an initial screen, such that only potential cases will be tested. Such testing scheme can be expressed in a form of a hierarchical tree of hypotheses. A suitable FDR controlling procedure has been proposed previously for independent tests. Yet dependence between the hypotheses in the tree may be expected for various reasons. For instance, strong dependence may exist between the hypotheses if each parent hypotheses is the intersection of its child hypotheses. We first demonstrate by simulation that the bound parameter is variant to dependence, and thus a formulation adaptive to dependence can help in obtaining more accurate bound. We show that the bound under dependence can be too conservative, and thus dependence-adaptive form can increase power. Next we offer an extended formulation for the full-tree FDR that allows accounting for dependence under hierarchical testing. The previous upper bound is modified to include dependence parametrization, which is represented by the covariance matrix of the test statistics. We implement the procedure within an expression QTL study which is aimed to detect epistasis. We also address practical considerations such as characterization of the dependence form and estimation of bound parameters.

EC1289: **A discrete modification of the Benjamini-Hochberg procedure using a weighting approach**
Presenter: **Sebastian Doehler**, Darmstadt University of Applied Science, Germany
In many multiple testing problems control of the false discovery rate is desired. This can be achieved by the Benjamini-Hochberg (BH) procedure when the $p-$values satisfy certain dependence properties. It is also known that the BH procedure controls the FDR if the $p-$values are discrete and

e.g. independent. We take discreteness into account by using a weighted version of the BH procedure. We investigate the power of this method for empirical and simulated data.

EO324 Room MAL B30 ANALYSIS OF COMPLEX TIME SERIES DATA Chair: Marina Knight

EO0394: **Detecting multiple change-points in panel data via Double CUSUM Binary Segmentation**
Presenter: **Haeran Cho**, University of Bristol, United Kingdom
The aim is to propose a method for detecting multiple change-points in the mean of (possibly) high-dimensional panel data. CUSUM statistics have been widely adopted for change-point detection in both univariate and multivariate data. For the latter, it is of particular interest to exploit the cross-sectional structure and achieve simultaneous change-point detection across the panel, by searching for change-points from the aggregation of multiple series of CUSUM statistics, each of which is computed on a single series of the panel data. For panel data of high dimensions, the detectability of a change-point is influenced by several factors, such as its sparsity across the panel, the magnitude of jumps at the change-point and the unbalancedness of its location, and having a method that handles a wide range of change-point configurations without any prior knowledge is vital in panel data analysis. The double CUSUM statistic is a determined effort in this direction, where the key ingredient is data-driven, point-wise partitioning of the multiple series into those contributing to change-point detection and those which do not. We show that the double CUSUM statistic, combined with a binary segmentation algorithm, attains consistent change-point detection in terms of both the total number and the locations of detected change-points, and conduct a comparative simulation study in which its good performance is demonstrated.

EO1045: **On the impact of autocorrelation and heterogeneity on functional connectivity**
Presenter: **Mark Fiecas**, University of Warwick, United Kingdom
Co-authors: Ivor Cribben
We discuss functional connectivity analyses of the human brain. The conventional estimate of functional connectivity does not account for temporal autocorrelation or heterogeneity across the subjects of the experiment; the former leads to inflated Type I errors in a single-subject analysis and the latter leads to low power in a multi-subject analysis. To address this, we propose a flexible general linear model framework for estimating functional connectivity that accounts for i) temporal autocorrelation in a nonparametric manner and ii) heterogeneity across subjects by allowing for subject-specific estimates of the variance. We use simulated data to assess the performance of our proposed method with respect to Type I and II errors, and we illustrate the utility of our proposed method on a depression study.

EO1085: **Current challenges in the analysis and forecasting of time series**
Presenter: **Marina Knight**, University of York, United Kingdom
Co-authors: Matthew Nunes, Guy Nason, Rebecca Killick, Idris Eckley
Time series analysis is common place in a variety of fields, from the energy sector to biology. Traditional statistical techniques assume observations are collected at regular intervals, coupled with assumed stationarity of the past in order to facilitate data modelling and/or to produce forecasts. However, for data arising in many fields, either one or both assumptions may fail: the stationarity assumption is often either violated, hence certain conditions need to be imposed on the time-varying structure of the process in order to achieve meaningful estimation, or the data may have been irregularly sampled and in fact exhibit stationary behaviour hidden by long-range dependence. We propose adaptive solutions to each case, and illustrate them with real data examples.

EO0915: **Analysis of time series observed on networks**
Presenter: **Matthew Nunes**, Lancaster University, United Kingdom
Co-authors: Marina Knight, Guy Nason
We consider analysis problems for time series that are observed at nodes of a large network structure. Such problems commonly appear in a vast array of fields, such as environmental time series observed at different spatial locations or measurements from computer system monitoring. The time series observed on the network might exhibit different characteristics such as nonstationary behaviour or strong correlation, and the nodal series evolve according to the inherent spatial structure. The new methodology we develop hinges on reducing dimensionality of the original data through a change of basis. The basis we propose is a second generation wavelet basis which operates on spatial structures. As such, the (large) observed data is replaced by data over a reduced network topology. We give examples of the potential of this dimension reduction method for time series analysis tasks.

EO198 Room CLO 203 IN-SAMPLE FORECASTING IN INSURANCE, LONGEVITY AND THE LABOUR MARKET Chair: Jens Perch Nielsen

EO0412: **Forecasting macroeconomic labour market flows: What we can learn from micro level analysis**
Presenter: **Ralf Wilke**, Copenhagen Business School, Denmark
Forecasting labour market dynamics and flows is important for budgeting and decision making in government departments and public administration. Macroeconomic forecasts are normally obtained from aggregate or time-series data. The purpose is to suggest and compare alternative approaches that employ individual level statistical analysis to predict labour market flows, taking the number of exits out of unemployment benefits as an example. This is done by means of putting an interdisciplinary method mix to various data structures. In particular, we consider how forecast models based on micro level estimates compare to a classical time-series forecast. The results with data from Germany suggest that prediction based on individual level analysis using large administrative individual data constitutes an alternative to aggregate data based forecasting. Despite that results are encouraging, further development work is required to fully exhaust the potential of the alternative approaches.

EO0513: **Prediction outstanding insurance claims via in-sample forecasting**
Presenter: **Munir Hiabu**, Cass Business School, United Kingdom
Co-authors: Enno Mammen, Lola Martinez-Miranda, Jens Perch Nielsen
Non-life insurance companies traditionally use the so called chain ladder method to reserve for outstanding liabilities. We will show how to translate the chain ladder method into a continuous framework while keeping the basic structure. The problem will hereby be translated into a survival analysis setting. As it turns out, chain ladder, and thus also our continuous analogue, is a in-sample technique where no extrapolation is needed to forecast the reserve. The in-sample area is defined as one triangle and the forecasting area as the second triangle that added to the first triangle produces a square. We call our approach in-sample forecasting. It is defined as forecasting a structured density to sets where the density is not observed. The in-sample forecasting will be done with nonparametric kernel methods. Firstly we will focus on the multiplicative density structure which also is the underlying assumption of chain ladder. Secondly, we will show how to go beyond this multiplicativity assumption in order to get more accurate forecasts.

EO0575: **Calendar effect and continuous chain ladder**
Presenter: **Michael Vogt**, University of Bonn, Germany
Co-authors: Lola Martinez-Miranda, Jens Perch Nielsen
One of the most important problems in non-life insurance is the estimation of reserves for outstanding liabilities. This problem is traditionally

approached by the chain ladder methodology. Recently, the so-called continuous chain ladder approach has been introduced to improve on the classical technique. The statistical problem underlying this approach is as follows: Let (X,Y) be a two dimensional random variable with a multiplicative density $f(x,y) = f_1(x)f_2(y)$ which is supported on a proper subset S of the rectangle $[0,1]^2$, specifically on the triangle $S = \{(x,y) \in [0,1]^2 : x+y \leq 1\}$. The aim is to estimate the multiplicative components f_1 and f_2 from a sample of data $\{(X_i,Y_i) : 1 \leq i \leq n\}$. In non-life insurance, there are many situations where the reserve is affected by so-called diagonal effects such as economic or claim inflation. The two-dimensional density model $f(x,y) = f_1(x)f_2(y)$ is unable to capture such effects. We present an extension of the model that is able to do so. The extended model assumes that the density can be written as $f(x,y) = f_1(x)f_2(y)f_3(x+y)$, where the function f_3 captures the diagonal or calendar time effect. We discuss the problems of identification, estimation and forecasting in this framework. Moreover, we illustrate our proposal by applications to several real datasets.

EO0644: **Operational time and in-sample density forecasting**
Presenter: **Young Kyung Lee**, Kangwon National University, Korea, South
Co-authors: Enno Mammen, Jens Perch Nielsen, Byeong Park
A new structural model for in-sample density forecasting is considered. In-sample density forecasting is to estimate a density function on a region where the density is observed and then re-use the estimated density while estimating the density on a region it is not observed. Our structural assumption is that the density is a product of one-dimensional functions with one function sitting on the scale of a transformed space of observations. The transformation involves another unknown one-dimensional function, so that our model is formulated via a known smooth function of three underlying unknown one-dimensional functions. We present an innovative way of estimating the one-dimensional functions and show that all the estimators of the three components achieve the optimal one-dimensional rate of convergence. We illustrate how one can use our approach by analysing a real data from the insurance business. We also investigate the finite sample performance of the method via a simulation study.

EO318 Room CLO 306 STATISTICAL INFERENCE FOR BIOMEDICAL DATA Chair: Ming-Yen Cheng

EO0450: **Two-sample density estimation with likelihood ratio ordering**
Presenter: **Tao Yu**, National Unviersity of Singapore, Singapore
The aim is to propose a method for estimating the probability density functions in a two sample problem where the ratio of the densities is monotonic. Such a problem is well motivated from medical data and has been widely identified in the literature. However, effective methods for solving this problem are not yet available in the community. Clearly, an effectively method should at least satisfy that the resulting estimates are probability densities and that the corresponding density ratio inherits the monotonic property, otherwise it is difficult to explain the analysis results. We propose estimates for these densities whose ratio inherits the monotonic property, and explore their theoretical properties. One interesting application of our density estimates is that the corresponding receiver operating characteristic (ROC) curve estimate is concave. Through numerical studies, we observe that both the density estimates and the ROC curve estimate from our method outperform their competitors, particularly when the sample size is relatively small.

EO0661: **Penalized feature selection with an application to detection of herbal medicine**
Presenter: **Ying Zhu**, Nanyang Technological University, Singapore
Co-authors: Augustine Tuck Lee Tan
Fourier transform infrared (FTIR) spectra of herbal medicine consist of many overlapping absorption bands sensitive to the physical and chemical states of compounds. Direct implementation of linear discriminant methods in high-dimensional data setting provides poor classification results and the interpretation of the results is challenging. A model based on a penalized discriminant analysis (PDA) using FTIR spectroscopy was developed for the purpose of discrimination and interpretation. The discriminant vector of PDA involves a linear combination of spectral features by introducing an additional penalty. PDA model with L1 norm produced sparse solutions with small number of selected wavelength regions. However the model behavior may not always be desirable due to its loss of spectral information. The model involving both L1 and L2 norms overcame the over-selective problem when seeking the optimal combination of spectral features. This model has been applied to distinguish between different species of Ganoderma lucidum with better classification performance than the full-spectrum model. The informative spectral absorption bands for discrimination have provided quantitative interpretations of the bioactive chemical constituents of Ganoderma lucidum regarding its anticancer effects.

EO0862: **Noise cancellation and classification of mouse ultrasonic vocalization data**
Presenter: **Xiaoling Dou**, Waseda University, Japan
Mouse ultrasonic vocalizations (USVs) are studied in various fields of science. However, background noise and varied USV patterns in observed signals make complete automatic analysis difficult. To analyze mouse USV data automatically, we propose noise reduction methods and functional cluster methods. These methods are programed with free software R and shown work well for mouse USVs data taken from laboratory mice.

EO1245: **Forward selection in varying coefficient models with application to genetic data**
Presenter: **Ming-Yen Cheng**, National Taiwan University, Taiwan
Varying coefficient models have numerous applications. While enjoying nice interpretability, they also allow for flexibility in modeling dynamic impacts of the covariates. In the new era of big data, it is challenging to select the relevant variables when the dimensionality is very large. Recently several works are focused on this important problem based on sparsity assumptions; they are subject to some limitations, however. We introduce an appealing forward selection procedure. It selects important variables sequentially according to a reduction in sum of squares criterion and it employs a BIC-based stopping rule. Clearly it is simple to implement and fast to compute, and possesses many other desirable properties from theoretical and numerical viewpoints. The BIC is a special case of the EBIC when an extra tuning parameter in the latter vanishes. We establish rigorous screening consistency results when either BIC or EBIC is used as the stopping criterion. The theoretical results depend on some conditions on the eigenvalues related to the design matrices, which can be relaxed in some situations. Results of an extensive simulation study and a genetic data example are also presented to show the efficacy and usefulness of our procedure.

EO088 Room CLO B01 FUNCTIONAL DATA AND RELATED TOPICS Chair: Alicia Nieto-Reyes

EO0481: **Independent component models for replicated point processes**
Presenter: **Daniel Gervini**, University of Wisconsin-Milwaukee, United States
A semiparametric independent-component model for the intensity function of a doubly-stochastic point process is introduced. As an example of application we analyze the spatial distribution of street robberies in the city of Chicago.

EO1196: **Clustering fragmentary functional data**
Presenter: **Yuguang Fan**, University of Melbourne, Australia
Co-authors: Aurore Delaigle
We study curves that are observed only on short and irregular intervals. Such fragmentary functional data is often collected in practice but it poses several difficulties when we try to apply the usual methods designed for curves defined on the entire interval. Under these circumstances, we propose a method to cluster such fragments addressing the difficulties in lateral connection and design an algorithm to optimize the cluster performance.

EO0960: **Symmetry for distributions on functional spaces**
Presenter: **Alicia Nieto-Reyes**, Universidad de Cantabria, Spain
Co-authors: Heather Battey
Distributions on function spaces are intangible objects for which no formally defined notions of symmetry currently exist. This is a serious hindrance to the methodological development of modern statistics, as the absence of a notion of centre of symmetry precludes a well conceived generalisation of the sample median and other sample quantiles to functional data sets, which are ever more prevalent in a society overwhelmingly rich in data. After exposing, through examples, difficulties associated with crude extensions of multivariate notions of symmetry to function spaces, we propose a notion of symmetry for distributions on a general functional metric space (F,d). Our notion is based on halfspaces induced by the metric d, and engenders symmetries in other domains, as we expose through simulations. In demonstrating that the corresponding sample centre of symmetry satisfies a robustness property for which the median was originally conceived, we ratify our sample centre of symmetry as a generalization of the sample median on function space. Finally, an application to handwriting data underlines the practical importance of our work.

EO0443: **A topologically valid definition of depth for functional data**
Presenter: **Heather Battey**, Imperial College London and Princeton University, United Kingdom
Co-authors: Alicia Nieto-Reyes
Several properties which constitute a definition of depth for functional data will be discussed. These properties recognise topological features such as continuity, smoothness and contiguity, and intrinsically address the delicate challenge of partial observability of functional data. Fulfillment thus gives rise to a minimal guarantee on the performance of the empirical depth beyond the idealised and practically infeasible case of full observability. As an incidental product, functional depths satisfying our definition achieve a robustness that is commonly ascribed to depth, despite the absence of a formal guarantee in the multivariate definition of depth. Some widely used functional depth proposals are discussed with reference to the aforementioned properties.

EO156 Room MAL 421 SPATIAL STATISTICS AND ECONOMETRICS Chair: Soumendra Lahiri

EO0558: **Circulant embedding of approximate covariances for inference from Gaussian data on large lattices**
Presenter: **Joseph Guinness**, NC State University, United States
Recently proposed computationally efficient Markov chain Monte Carlo and Monte Carlo Expectation-Maximization (EM) methods for estimating covariance parameters from lattice data rely on successive imputations of values on an embedding lattice that is at least two times larger in each dimension. These methods can be considered exact in some sense, but we demonstrate that using such a large number of imputed values leads to slowly converging Markov chains and EM algorithms. We propose instead the use of a discrete spectral approximation to allow for the implementation of these methods on smaller embedding lattices. While our methods are approximate, our examples indicate that the error introduced by this approximation is small compared to the Monte Carlo errors present in long Markov chains or many iterations of Monte Carlo EM algorithms. Our results are demonstrated in simulation studies, as well as in numerical studies that explore both increasing domain and fixed domain asymptotics. We compare the exact methods to our approximate methods on a large satellite dataset, and show that the approximate methods are also faster to compute, especially when the aliased spectral density is modeled directly.

EO0642: **Autoregressive spatial spectral estimates**
Presenter: **Abhimanyu Gupta**, University of Essex, United Kingdom
Autoregressive spectral density estimation for stationary random fields on a regular spatial lattice has many advantages relative to kernel based methods. It provides a guaranteed positive-definite estimate even when suitable edge-effect correction is employed, is simple to compute using least squares and necessitates no choice of kernel. We truncate a true half-plane infinite autoregressive representation to estimate the spectral density. The truncation length is allowed to diverge in all dimensions in order to avoid the potential bias which would accrue due to truncation at a fixed lag-length. Consistency and strong consistency of the proposed estimator, both uniform in frequencies, are established. Under suitable conditions the asymptotic distribution of the estimate is shown to be zero-mean normal and independent at fixed distinct frequencies, mirroring the behaviour for time series. A small Monte Carlo experiment examines finite sample performance. We illustrate the technique by applying it to Los Angeles house price data and a novel analysis of voter turnout data in a US presidential election. Technically the key to the results is the covariance structure of stationary random fields defined on regularly spaced lattices. We study this in detail and show the covariance matrix to satisfy a generalization of the Toeplitz property familiar from time series analysis.

EO0721: **Confidence regions for excursion sets in asymptotically Gaussian random fields with an application to climate**
Presenter: **Armin Schwartzman**, North Carolina State University, United States
Co-authors: Max Sommerfeld
The goal is to give confidence regions for the excursion set of a spatial function above a given threshold from repeated noisy observations on a fine grid of fixed locations. Given an asymptotically Gaussian estimator of the target function, a pair of data-dependent nested excursion sets are constructed that are sub- and super-sets of the true excursion set, respectively, with a desired confidence. Asymptotic coverage probabilities are determined via a multiplier bootstrap method, not requiring Gaussianity of the original data nor stationarity or smoothness of the limiting Gaussian field. The method is used to determine regions in North America where the mean summer and winter temperatures are expected to increase by mid 21st century by more than 2 degrees Celsius.

EO1140: **An empirical likelihood method for spatial data with stochastic locations**
Presenter: **Daniel Nordman**, Iowa State University, United States

An empirical likelihood (EL) method for irregularly located spatial data is described. It is difficult to determine distributions and standard errors for many common estimators based on such spatial data, because asymptotic variances often depend on a intricate combination of several quantities, including the spectral density of the spatial process, the concentration of spatial locations, and the spatial asymptotic structure. The proposed spatial EL method, formulated in the frequency domain, allows inference without knowledge or estimation of these factors. The main results show that, regardless of these factors, the logarithm of the proposed EL ratio statistic has a chi-squared limit. As a result, the proposed spatial method can be used to build asymptotically correct confidence regions and tests for covariance parameters specified through spectral estimating equations. However, unlike frequency domain EL for time series (on a regular grid), the spatial EL formulation needs special care due to lack of the usual orthogonality properties of the discrete Fourier transform for irregularly spaced data and due to presence of nontrivial bias in the periodogram under different spatial asymptotic structures. Some simulation results illustrate the finite sample properties of the method.

EO128 Room CLO 101 STATISTICS IN FUNCTIONAL AND HILBERT SPACES Chair: Gil Gonzalez-Rodriguez

EO0645: **Classification of functional data using bi-level selection**
Presenter: **Hidetoshi Matsui**, Kyushu University, Japan

Penalties with an L_1 norm provide solutions in which some coefficients are exactly zero and can be used for selecting variables in regression settings. When applied to the logistic regression model, they also can be used to select variables which affect classification. We focus on the form of L_1 penalties in the logistic regression models for functional data, in particular, their use in classifying functions into three or more groups while simultaneously selecting variables or classification boundaries. We propose a new class of penalties in order to appropriately estimate and select variables or boundaries for the functional multiclass logistic regression model. In order to do this we extend the existing bi-level variable selection methods. The parameters involved in the model is estimated by the framework of the blockwise descent algorithm, and then a value of tuning parameters included in the regularization method is decided by a model selection criterion. We then apply the proposed method to the analysis of real data.

EO0733: **Canonical correlation analysis for multivariate functional data**
Presenter: **Michio Yamamoto**, Kyoto University, Japan
Co-authors: Yoshikazu Terada

The generalized functional canonical correlation analysis (GFCCA) is the extension of functional canonical correlation analysis from pairs of random functions to the case where a data sample consists of multiple square-integrable stochastic processes more than two. Since the dimension of a function is typically infinite, GFCCA may not be meaningful. To address this issue, sufficient conditions under which GFCCA has a meaningful solution are derived. Furthermore, as with the classical generalized canonical correlation analysis for multivariate data, another formulation of GFCCA based on homogeneous analysis is discovered. The equivalence between the two different formulations is provided, and it enables researchers to use GFCCA for various purposes.

EO0833: **Using functional singular components to obtain correlations with an application to a dietary study**
Presenter: **Christian Ritz**, University of Copenhagen, Denmark

Dietary studies commonly evaluate effects by means of outcomes such blood parameters, visual analogue scores, and physical activity. These outcomes are typically recorded repeatedly over time on the same subjects, resulting in (sometimes sparse) functional data. The standard approach for evaluating the intervention effect is to use separate univariate analyses for each outcome. Such analyses are usually based on linear mixed model methodology, which is very flexible in accommodating various study designs, but it does not fully allow for a multivariate appreciation of the results in terms of understanding to which degree the different outcomes reflect the same trend in the data. Therefore, it is often also of interest to understand how the different outcome profiles over time are mutually correlated. It is shown how to exploit a functional singular value decomposition to derive correlation coefficients between pairs of outcome profiles. Singular components and in particular singular functions are introduced and used for defining functional correlation coefficients. Results obtained from the analysis of a dietary intervention study are presented and compared to other methods used.

EO0951: **Functional graphical models**
Presenter: **Xinghao Qiao**, London School of Economics, United Kingdom
Co-authors: Gareth James, Jinchi Lv

Graphical models have attracted increasing attention in recent years, especially in settings involving high dimensional data. In particular Gaussian graphical models are used to model the conditional dependence structure among p Gaussian random variables. As a result of its computational efficiency the graphical lasso (glasso) has become one of the most popular approaches for fitting high dimensional graphical models. We extend the graphical models concept to model the conditional dependence structure among p random functions. In this setting, not only is p large, but each function is itself a high dimensional object, posing an additional level of statistical and computational complexity. We develop an extension of the glasso criterion (fglasso), which estimates the functional graphical model by imposing a block sparsity constraint on the precision matrix, via a group lasso penalty. The fglasso criterion can be optimized using an efficient block coordinate descent algorithm and our theoretical results demonstrate that, with high probability, the fglasso will correctly identify the true conditional dependence structure. Finally we show that the fglasso significantly outperforms possible competing methods through both simulations and an analysis of a real world data set.

EO064 Room MAL B20 CALIBRATION OF MISSPECIFIED BAYESIAN MODELS Chair: Natalia Bochkina

EO0687: **Risk of Bayesian inference in misspecified models, and the sandwich covariance matrix**
Presenter: **Ulrich Mueller**, Princeton University, United States

It is well known that in misspecified parametric models, the maximum likelihood estimator (MLE) is consistent for the pseudo-true value and has an asymptotically normal sampling distribution with "sandwich" covariance matrix. Also, posteriors are asymptotically centered at the MLE, normal and of asymptotic variance that is in general different than the sandwich matrix. It is shown that due to this discrepancy, Bayesian inference about the pseudo-true parameter value is in general of lower asymptotic frequentist risk when the original posterior is substituted by an artificial normal posterior centered at the MLE with sandwich covariance matrix. An algorithm is suggested that allows the implementation of this artificial posterior also in models with high dimensional nuisance parameters which cannot reasonably be estimated by maximizing the likelihood.

EO0692: **Learning the learning rate: How to repair Bayes when the model is wrong**
Presenter: **Peter Grunwald**, CWI and Leiden University, Netherlands
Co-authors: Thijs van Ommen

If the model is misspecified, then the Bayes posterior may fail to concentrate even for large samples, leading to extreme overfitting (and bad calibration) in practice. We demonstrate this on a very simple regression problem. The problem goes away if we make the so-called learning rate small enough, which essentially amounts to making the prior more and the data less important. Standard Bayes sets the learning rate to 1, which

can be too high under model misspecification; close cousins of Bayes popular in the learning theory community set the learning rate to 1/sqrtsample size, which is too low if the setting is not adversarial. We introduce the safe Bayesian estimator, which learns the learning rate from the data. It behaves essentially as well as standard Bayes if the model is correct but continues to be calibrated in a relevant sense with wrong models.

EO0788: **The Bernstein-von Mises theorem and misspecified nonregular models**
Presenter: **Peter Green**, UTS Sydney, United Kingdom
Co-authors: Natalia Bochkina
The asymptotic behaviour of the posterior distribution is studied in a broad class of statistical models that can be misspecified, in that the true distribution of the data does not belong to the assumed parametric family. We focus on the case where the parametric model best approximating the true distribution is non-regular, that is, where the parameter giving the best parametric model is on the boundary of the parameter space. We show that in this case, asymptotically, the posterior distribution has not only Gaussian components, as in the case of regular misspecified models, but also Gamma distributed components. The form of these depends on the behaviour of the prior near the boundary, and the rate of convergence is faster than the parametric rate. We illustrate the application of these results as a diagnostic tool for possible model misspecification, for instance, missing covariates or incorrect link functions in generalised linear models.

EO1039: **Calibration of conditional composite likelihood for Bayesian inference on Gibbs random fields**
Presenter: **Julien Stoehr**, University College Dublin, France
Co-authors: Nial Friel
Gibbs random fields play an important role in statistics, however the resulting likelihood is typically unavailable due to an intractable normalizing constant. Composite likelihoods offer a principled means to construct useful approximations. An approach will be presented to calibrate the posterior distribution resulting from using a composite likelihood and illustrate its performance in several examples.

EO118 Room MAL 414 STATISTICAL REGULARIZATION Chair: Subir Ghosh

EO0785: **Functional Fisher rules and reproducing kernel Hilbert spaces**
Presenter: **Jose Berrendero**, Universidad Autonoma de Madrid, Spain
Co-authors: Antonio Cuevas, Jose Luis Torrecilla
The Hajek-Feldman dichotomy establishes that two Gaussian measures are either equivalent (and hence there is a Radon-Nikodym (RN) density for each measure with respect to the other one) or mutually singular. Unlike the case of finite dimensional Gaussian measures, there are non-trivial examples of both situations when dealing with Gaussian stochastic processes. It is often possible to derive the optimal (Bayes) rule and the minimal classification error probability in several relevant problems of supervised binary functional classification defined by two equivalent measures. On the other hand, near perfect classification phenomena arise when the measures are mutually singular. We establish some results to formalize these ideas, which rely on the theory of Reproducing Kernel Hilbert Spaces (RKHS). We also propose a new method for variable selection in binary classification problems, which arises in a very natural way from the explicit knowledge of the RN-derivatives and the underlying RKHS structure. As a consequence, the optimal classifier in a wide class of functional classification problems can be expressed in terms of a classical, linear finite-dimensional Fisher's rule.

EO1235: **Variable selection in high dimensional models using generalized cross validation**
Presenter: **Maarten Jansen**, ULB Brussels, Belgium
We develop the use of generalized cross validation (GCV) as a tool for tuning the selection of variables from a high-dimensional parameter vector under the assumption of sparsity. As a selection procedure we consider the Lasso (L_1-regularized least squares) where GCV is used in the optimization of the regularization parameter or penalty. The use of GCV in this context rests on the combination of two results. One result links GCV to Mallows's C_p and Stein's Unbiased Risk Estimator for a near-optimal value of the tuning parameter. The second result ensures that GCV behaves well for large (near-full) models, i.e., for penalties tending to zero. The second result is necessary to prevent a zero penalty from being a local or global optimum in the optimization procedure. The two results provide the necessary framework for extending GCV for use in L_0-regularized least squares, leading to a variable selection scheme without shrinkage. Simulations illustrate the robustness of GCV as an implicit estimator of the variance of the errors.

EO1237: **Multiscale variational methods with applications in statistical imaging and signal recovery**
Presenter: **Axel Munk**, Georg-August-University Goettingen, Germany
We review a general class of statistical multiresolution estimators (MIND) which arise from variational analysis. These estimators that are defined as solutions of convex optimization problems with sup norm type constraints. We employ a combination of the alternating direction method of multipliers with Dykstras algorithm for computing orthogonal projections onto intersections of convex sets. MIND comprises many known estimators including the Dantzig selector, wavelet thresholding and the LASSO, and some new ones. The proposed approach is illustrated by various examples from imaging and signal detection.

EC0230: **An adaptive ridge procedure for L_0 regularization**
Presenter: **Florian Frommlet**, Medical University Vienna, Austria
Co-authors: Gregory Nuel
Penalized selection criteria like AIC or BIC are among the most popular methods for variable selection. Their theoretical properties have been studied intensively and are well understood, but making use of them in case of high-dimensional data is difficult due to the non-convex optimization problem induced by L_0 penalties. We introduce an adaptive ridge procedure (AR), where iteratively weighted ridge problems are solved whose weights are updated in such a way that the procedure converges towards selection with L_0 penalties. After introducing AR its theoretical properties are studied in the particular case of orthogonal linear regression. For the non-orthogonal case extensive simulations are performed to assess the performance of AR. In case of Poisson regression and logistic regression it is illustrated how the iterative procedure of AR can be combined with iterative maximization procedures. We conclude with an efficient implementation of AR in the context of least-squares segmentation.

EO212 Room MAL G15 CAUSAL DISCOVERY IN PRACTICE: ADVANCES AND CHALLENGES Chair: Jonas Peters

EO0948: **Faithfulness of probability distributions to graphs**
Presenter: **Kayvan Sadeghi**, Carnegie Mellon University, United States
We provide sufficient conditions for a given probability distribution to be Markov to a graph with the minimum possible number of edges, and more importantly, necessary and sufficient conditions for a given probability distribution to be faithful to a graph. We present our results briefly for the general case of mixed graphs, but in more details, specialize the definitions and results to the subclass directed acyclic graphs, which are essential elements of causal inference. Based on the results, we discuss new methods and algorithms for selecting graphical models that capture a given set of conditional independence statements.

EO1255: **Challenges in experimenting across advertisers**
Presenter: **Elon Portugaly**, Microsoft, United Kingdom

It is by now standard practice for online service providers to experiment across users: in simplified terms, users are randomly assigned to either a treatment or a control group, and each group experiences a somewhat different flavour of the service. Then some measurement of interest (e.g. revenue) is taken for each user. Assuming each user reacts only to the service variant it experiences, and assuming a large number of users, this allows calculating the effect of the treatment as the difference between what we would have measured had all users experienced treatment and what we would have measured had all users experienced control. Online advertising platforms interact with another type of agent beside the end user, namely the advertiser. To correctly estimate the effect of a treatment, it is therefore necessary to consider the reaction of advertisers, as well as users, to the treatment. However, advertisers are fewer, more variant, slower to react, and are more interdependent than end users. We will describe the challenges of designing experiments aiming to measure advertiser reaction, and some of the methods we use to tackle these challenges.

EO1800: **Modeling local causal gene networks to address the covariate dilemma in association genetics**
Presenter: **Oliver Stegle**, EMBL European Bioinformatics Institute, United Kingdom

The genetic analysis of high-dimensional phenotypes is often compromised by hidden structure between samples. In genetics, this structure can be linked to differences between individuals, either reflecting their genetic makeup (such as population structure) or environmental and technical factors. We discuss challenges and successful applications of principles from causal reasoning to estimate sample covariance matrices to account for such hidden structure. We show that simplistic methods based on correlative evidence alone are prone to overfitting, resulting in poor covariance estimation. To address this, we propose a simple yet efficient approach that first reconstructs local directed gene networks from the observed expression data to then condition on covariance matrices that are estimated from genes with a causal effects on the phenotype of interest. We validate this approach extensively in simulations and find substantial gains in statistical power and accuracy in several applications to problems in genetics.

EO1842: **backShift: Learning causal cyclic graphs from unknown shift interventions**
Presenter: **Dominik Rothenhaeusler**, ETH Zurich, Switzerland
Co-authors: Christina Heinze, Jonas Peters, Nicolai Meinshausen

We propose a simple method to learn linear causal cyclic models in the presence of latent variables.The method relies on equilibrium data of the model recorded under a specific kind of interventions ("shift interventions"). The location and strength of these interventions do not have to be known and can be estimated from the data. Our method, called "backShift", only uses second moments of the data and performs simple joint matrix diagonalization, applied to differences between covariance matrices. We give a sufficient and necessary condition for identifiability of the system. We demonstrate the performance on some simulated data and applications in flow cytometry and financial time series.

EO260 Room MAL B33 MODEL SELECTION AND COMPARISON IN HIGHLY-STRUCTURED SETTINGS Chair: David van Dyk

EO1161: **Comparing non-nested models in the search for new physics**
Presenter: **Sara Algeri**, Imperial College London, United Kingdom
Co-authors: David van Dyk, Jan Conrad

Astrophysicists and particle physicists are often interested in non-standard model comparisons. The search for the Higgs boson, for example, involves quantifying evidence for a narrow component added to a diffuse background distribution. The added component corresponds to the Higgs mass distribution, accounting for instrumental effects, and cannot be negative. Thus, not only is the null distribution on the boundary of the parameter space, but the location of the added component is unidentifiable under the null. In other settings, for example in the search for dark matter, physicists may aim to compare non-nested models. This can occur when known sources mimic a new signal or when two differently parametrized models must be compared. Because many researchers have a strong preference for frequency-based statistical methods, they use Likelihood Ratio Tests (LRT) and when appropriate use Wilks and Chernoff-type asymptotics (Monte Carlo methods are often infeasible owing to extreme significance criteria). We consider the case of non-nested models. By formulating an additive composite model, we are able to view the LRT statistics as a random process indexed by the parameter that is unidentified under the null. This allows us to leverage methods developed to handle so called trial factors and obtain asymptotically valid frequentist tests. We illustrate our proposed method in a series of numerical studies that validate its power and nominal false positive rate.

EO0367: **Robust sparse variable selection via flexible errors**
Presenter: **David Rossell**, University of Warwick, United Kingdom

A main challenge in high-dimensional variable selection is to enforce sparsity adequately. Because of theoretical and computational difficulties in non-standard settings, most research efforts have been directed towards linear regression with Normal errors. Naturally, in actual applications errors may not be Normal, hence it is important to relax these assumptions and consider their implications. We extend the usual Bayesian variable selection framework for Normal linear models to more flexible errors that can capture asymmetries and heavier-than-normal tails. Importantly the error structure is learnt from the data, so that the model automatically reduces to Normal errors when the further flexibility is not needed. We show that the corresponding log-likelihoods are concave, leading to computationally efficient optimization and integration and hence rendering the approach practical in high dimensions. Further, although the models are slightly non-regular we show that one can obtain asymptotic characterizations in an M-open situation, i.e. where data are truly generated from a model outside the considered families. Our examples show that while the extra flexibility has no significant cost when errors are truly Normal, when they are non-normal it helps increase statistical power. Our current theory addresses sparsity in finite dimensions and provides a promising basis upon which to build high-dimensional robust sparse inference approaches.

EO0585: **Power weighted densities for time series data**
Presenter: **Shane Jensen**, The Wharton School of the University of Pennsylvania, United States

A crucial issue in modeling time series data is the possibility of non-stationarity in the underlying data generating process. We examine two time series applications in the presence of non-stationarity: predicting future hitting performance in major league baseball and predicting future returns of stock portfolios. We develop a simple and effective approach to allow non-stationarity in time series modeling when the primary goal is forecasting future time points, while also allowing the practitioner to choose the data model. In our power-weighted density (PWD) approach, observations in the distant past are down-weighted in the likelihood function relative to more recent observations under the practitioners chosen data model. Our PWD approach is a simpler alternative for allowing non-stationarity compared to popular state-space methods that explicitly model the evolution of an underlying state vector. We present specific PWD approaches for simple exponential families, hierarchical models and linear regression models, which are needed for our two applications. We demonstrate the benefits of our PWD approach in terms of predictive performance compared to both stationary models and alternative non-stationary methods such as state-space and integrated moving average models.

EO1145: **Valid statistical comparisons without valid MCMC output**
Presenter: **Nathan Stein**, University of Pennsylvania, United States
Co-authors: Xiao-Li Meng, David van Dyk

We consider the problem of model comparison when test statistics are derived from the output of a Markov chain Monte Carlo algorithm. This problem is motivated by an astrophysical application, the analysis of X-ray images of quasars, in which the goal is to compare a null model to a model that allows flexible, nonparametric departures from this null model. We investigate this comparison from a classical hypothesis testing perspective and find that the power of such tests can exhibit surprising behavior. In particular, there is no guarantee that power monotonically increases with the number of MCMC iterations; in other words, stopping the MCMC runs early can lead to better performance. We discuss strategies for obtaining high power when little is known in advance about the Markov chains' stationary distributions.

EO264 Room CLO 102 DIRECTIONAL STATISTICS Chair: Thomas Verdebout

EO1847: **Testing uniformity on high-dimensional spheres against rotationally symmetric alternatives**
Presenter: **Christine Cutting**, Universite Libre de Bruxelles, Belgium
Co-authors: Davy Paindaveine, Thomas Verdebout
We will consider non-null issues in the problem of testing uniformity on high-dimensional unit spheres. Throughout, the dimension p is allowed to go to infinity in an arbitrary way as a function of the sample size n.We will see that rotationally symmetric alternatives lead to two Local Asymptotic Normality (LAN) structures. The first one is for fixed modal location theta and allows to derive locally asymptotically most powerful tests under specified theta. The second one, that addresses the Fisher–von Mises–Langevin (FvML) case, relates to the unspecified-theta problem and shows that the high-dimensional Rayleigh test is locally asymptotically most powerful invariant. Under mild assumptions, the asymptotic non-null distribution of this test allows to extend away from the FvML case the asymptotic powers obtained there from Le Cam's third lemma.

EO0294: **Testing for spherical location in the vicinity of the uniform distribution**
Presenter: **Davy Paindaveine**, Universite libre de Bruxelles, Belgium
Co-authors: Thomas Verdebout
The problem of testing the null hypothesis that the location parameter of a rotationally symmetric distribution is equal to a given value is considered. Motivated by a real-data example that shows little deviation from uniformity on the sphere, we investigate the robustness of the Watson test and of the Wald test based on the spherical mean under a sequence of null hypotheses that are contiguous to the uniform. We show that, while these classical tests are asymptotically equivalent away from uniformity, they exhibit very different behaviours in the vicinity of uniformity. Our asymptotic investigation also extends to non-null sequences of hypotheses. The finite-sample relevance of our theoretical results is illustrated through Monte Carlo simulations.

EO0368: **Goodness-of-fit test for noisy directional data**
Presenter: **Thanh Mai Pham Ngoc**, University Paris Sud Orsay, France
Co-authors: Peter T Kim, Ja Yong Koo, Claire Lacour
The aim is to consider the nonparametric goodness-of-fit test of the uniform density on the sphere when we have observations whose density is the convolution of an error density and the true underlying density. We will deal specifically with the smooth and supersmooth error case, this latter includes the Gaussian distribution. Similar to deconvolution density estimation, the smoother the error density the harder is the rate recovery of the test problem. When considering nonparametric alternatives expressed over Sobolev and analytic classes, we show that it is possible to obtain original separation rates. Furthermore, we show that our adaptive statistical procedure attains these optimal rates. Simulations and some applications in astrophysics are tackled.

EO0870: **Multimodality tests for circular data**
Presenter: **Jose Ameijeiras-Alonso**, University of Santiago de Compostela, Spain
Co-authors: Rosa Crujeiras, Alberto Rodriguez-Casal
In applied statistics, the identification of the number of preferred directions in animal movements or the assessment of peak times in data observed across time, can be reduced to the same statistical problem: determining the number of modes (local maxima of the probability density function) in the underlying distribution of a sample of angles. Testing for multimodality provides a formal way to determine if the (circular) distribution of the sample has a specific number of modes. Within this context, there are various proposals, both for scalar and for circular data. However, the poor calibration of the methods provides unsatisfactory results in practice. A method for assessing multimodality in circular data will be presented, jointly with a calibration algorithm whose performance will be supported by simulation experiments and a real data application.

EO0805: **Probability integral transforms in directional statistics**
Presenter: **Peter Jupp**, University of St Andrews, United Kingdom
Co-authors: Alfred Kume
The standard method of transforming a continuous distribution on the line to the uniform distribution on the unit interval is the probability integral transform. Such a transform can be defined also for distributions on the circle. A version of probability integral transform is introduced for distributions with continuous positive density on compact Riemannian manifolds (such as spheres or rotation groups). It is a continuous mapping of the manifold to itself that transforms the distribution into the uniform distribution. It is based on the usual probability integral transform along each geodesic through a given point. Although the mapping is not unique, there are 'almost canonical' choices. Applications include (i) decomposition of distributions on product spaces into marginal distributions and copulae, (ii) derivation of tests of goodness of fit from tests of uniformity. A non-parametric analogue produces uniform scores, which give rise to multi-sample tests. The construction can be extended to some other sample spaces, including simplices and simply-connected spaces of non-positive curvature.

EG135 Room MAL 415 CONTRIBUTIONS IN EXTREME VALUES THEORY AND APPLICATIONS Chair: Vicky Fasen

EC1497: **Application of the PORT methodology to recent extreme value index estimators**
Presenter: **Ivette Gomes**, University of Lisbon, Portugal
Co-authors: Ligia Henriques-Rodrigues
Given a sample of size n of either independent, identically distributed or possibly stationary weakly dependent random variables from a cumulative distribution function (CDF) F, let us assume that F is in the domain of attraction for maxima of an extreme value (EV) CDF with a positive extreme value index (EVI). For this type of Pareto right-tailed parents, the mean-of-order-p (MOP) EVI-estimator has revealed to be highly flexible, but just like the Hill EVI-estimator, is not location-invariant, a property valid for the EVI itself. The application of the PORT methodology based on the excesses over a q-quantile, and adequate algorithms for the choice of the tuning parameters under play, enable a reliable estimation of parameters of rare events and a location-invariant EVI-estimation.

EC1726: **Searching for robust estimates of the extremal index**
Presenter: **Cristina Miranda**, University of Aveiro, Portugal
Co-authors: Manuela Souto de Miranda, Ivette Gomes
Many practical problems deal with extreme values above fixed levels and occurring in clusters of exceedances. The dependence among clusters

is characterized by the extremal index of the process, which can be interpreted as the reciprocal of the clusters size mean. Several estimators of the extremal index have been proposed. They essentially differ in the identification of the groups. One of the most used proposals is known as the blocks estimator. Under specific local dependence and stationarity conditions the limit distribution of the number of exceedances follows a compound Poisson process where clusters occurrences are modeled by a Poisson distribution and clusters dimensions are determined by the process multiplicities. Since clusters dimension is generally reduced and can have a discrete asymmetric distribution, a robust version of the estimator should be based on a robust estimator of the mean clusters dimension taking into account those distributional properties. Assuming the Poisson distribution of the number of exceedances per cluster, the extremal index is estimated considering a robust Poisson regression estimate of the mean dimension of the clusters. The performance of the method is evaluated through the comparison of the results with those obtained when the observations are generated by a contaminated distribution.

EC1613: **A latent trawl process model for extreme values**
Presenter: **Ragnhild Cassel Noven**, Imperial College London, United Kingdom
Co-authors: Axel Gandy, Almut Veraart

Trawl processes are a class of stationary, continuous-time stochastic processes driven by an independently scattered random measure. They belong to the wider class of so-called Ambit fields, which were originally developed as spatio-temporal models for studying turbulence, and give rise to a flexible class of models that can accommodate non-Gaussian distributions and a wide range of covariance structures. We explore a purely temporal, hierarchical model for exceedances over a threshold, based on a mixture decomposition of the generalised Pareto distribution. This model uses a trawl process as a latent component, which allows for a flexible temporal correlation structure in the exceedances. We present properties of this model, focusing on the dependence structure of the exceedances, and discuss frequentist and Bayesian approaches to parameter estimation. Finally, we explore a reparameterisation that allows the model to capture a wider range of distributions and improves the parameter estimation.

EC1494: **Nonparametric estimators for conditional extreme quantiles**
Presenter: **Fatima Palacios Rodriguez**, University of Seville, Spain
Co-authors: Elena Di Bernardino

Let $\mathbf{X} = (X_1, \ldots, X_d)$ be a risk vector and $\partial L(\alpha)$ be the associated multivariate critical layer at level $\alpha \in (0,1)$. We provide a nonparametric extreme estimation for the $(1-p_n)$-quantile of $T_i := [X_i \,|\, \mathbf{X} \in \partial L(\alpha)]$ for a fixed α and when $p_n \to 0$, as the sample size $n \to +\infty$. To this end, by considering an Archimedean copula as dependence structure for $\mathbf{X}$ and the Von Mises condition for marginal X_i, an extrapolation procedure is developed. The asymptotic convergence of our estimator is studied for $p = p_n \to 0$, as n tends towards infinity. On the other hand, an adaptive version of the proposed estimator is presented. Finally, the performance of our estimator is evaluated in simulated data.

EC034 Room MAL B29 CONTRIBUTIONS IN METHODOLOGICAL STATISTICS Chair: John Hinde

EC1487: **Scoring rules for prediction**
Presenter: **Valentina Mameli**, Ca Foscari University of Venice, Italy
Co-authors: Federica Giummole

Most of the methods nowadays employed for comparing probability forecasts are based on proper scoring rules. To our knowledge, the use of scoring rules is exclusively restricted to assess the quality of a given probabilistic forecast for a future random variable. However, in the prediction framework the use of these objects can be extended to provide a whole predictive distribution for the unknown of interest. We discuss the asymptotic properties of predictive distributions obtained by minimizing the divergence associated to different scoring rules. Some examples dealing with the Tsallis score, which includes as special cases some of the most used scoring rules, are taken into consideration. For some Gaussian models, the predictive distribution obtained by minimizing the Tsallis divergence is asymptotically equivalent to the one obtained from the Kullback-Liebler divergence. We also consider the class of weighted scoring rules that evaluate probability forecasts on the basis of a non-uniform baseline distribution representing the available information at the time of prediction. The divergences associated to certain weighted scoring rules are shown to be asymptotically equivalent to α-divergences, for which optimal predictive distributions exist.

EC1665: **Estimation of delta-contaminated density of the random intensity of Poisson data**
Presenter: **Daniela De Canditiis**, CNR, Italy
Co-authors: Marianna Pensky

An estimator of a delta contaminated mixing density function $g(\lambda)$ of the intensity λ of the Poisson distribution is constructed. The estimator is based on an expansion of the continuous portion $g_0(\lambda)$ of the unknown pdf over an overcomplete dictionary with the recovery of the coefficients obtained as solution of an optimization problem with Lasso penalty. In order to apply Lasso technique in the so-called prediction setting where it requires virtually no assumptions on dictionary and, moreover, to ensure fast convergence of Lasso estimator, we use a novel formulation of the optimization problem based on inversion of the dictionary elements. The total estimator of the delta contaminated mixing pdf is obtained using a two-stage iterative procedure. Conditions on the dictionary and the unknown mixing density that yield a sharp oracle inequality for the norm of the difference between $g_0(\lambda)$ and its estimator are formulated. Numerical simulations and comparisons with other recently constructed procedures are shown.

EC1216: **Confidence interval-driven selective inference**
Presenter: **Asaf Weinstein**, University of Pennsylvania, United States
Co-authors: Daniel Yekutieli

Given m unknown parameters with corresponding independent estimators, the BH procedure can be used to classify the sign of parameters such that the proportion of incorrect directional decisions is controlled. More ambitiously, our goal is to construct sign-determining confidence intervals (CIs) such that the proportion of non-covering constructed intervals (FCR) is controlled. We suggest a valid procedure, which is governed by a choice of a marginal $1-\alpha$ CI. In particular, the shape of the marginal CI controls the tradeoff between the tendency to construct more confidence intervals (power) and the length of the constructed intervals. A one-sided marginal CI, for example, results in selecting according to a level-$2q$ BH procedure, but construct intervals of unbounded length; while using the two-sided marginal CI results in selecting according to a level-q BH procedure, and in turn construct shortest CIs. We propose a new marginal CI that, when used in our procedure nicely balances this tradeoff. Specifically, we show in simulation that typically, it is possible to select according to BH at a level very close to $2q$ and at the same time construct short intervals.

EC1683: **Testing covariance structures: Non-normal approach**
Presenter: **Tonu Kollo**, University of Tartu, Estonia

Test-statistics for sphericity and uncorrelatedness hypothesis on covariance structure are examined for non-normal populations. We use test-statistics based on trace functions. In a special case all the fourth order moments are assumed to be equal. Taylor expansions of the test statistics have been derived, asymptotic normality and chi-square distributions have been established and their behaviour examined in the situation when both, sample size n and number of variables p tend to infinity, $p/n < 1$. Parallel to theoretical study a simulation experiment is carried out to

investigate empirically speed of convergence to asymptotic distributions depending on sample size, number of variables and parameters of the population distribution.

EC1721: **A modified class of estimators for the population mean with auxiliary information in presence of non-response**
Presenter: **Saba Riaz**, Riphah International University Islamabad, Pakistan

The estimation of unknown population mean of the variable of interest with the auxiliary variable is addressed when the problem of non-response is present. A general class of biased estimators is suggested using the known auxiliary information. The asymptotic bias and the asymptotic variance of the suggested class are obtained up to the first degree of approximation. The theoretical results are compared with the linear regression estimator. A numerical study is provided to show the efficiency and superiority of the proposed class w.r.t. the linear regression estimator and some already existing estimators.

EG055 Room MAL 540 CONTRIBUTIONS IN MULTIVARIATE ANALYSIS Chair: Robert Serfling

EC1363: **Global minimum variance portfolio weights for small sample and singular covariance matrix: estimation and test theory**
Presenter: **Stepan Mazur**, Lund University, Sweden
Co-authors: Krzysztof Podgorski, Taras Bodnar

The main contribution is the derivation of the distribution of the estimated GMV portfolio weights in the elliptically contoured model for the case of the portfolio size exceeding the sample size and the covariance matrix with rank smaller than the sample size. Moreover, we derive a test for the hypothesis of linear dependence between GMV portfolio weights, which is an important issue in the case of large portfolio size where one can expect some linear dependencies. The results are illustrated using actual stock returns and a discussion of practical relevance of the model is presented.

EC1464: **Bias-corrected inference for multivariate nonparametric regression**
Presenter: **Maria Lucia Parrella**, University of Salerno, Italy
Co-authors: Francesco Giordano

The local polynomial estimator is particularly affected by the curse of dimensionality, which reduces the potential of this tool for large-dimensional applications. We propose an estimation procedure based on the local linear estimator and a sparseness condition that focuses on nonlinearities in the model. Our procedure, called BID (bias inflation-deflation), is automatic and easily applicable to models with many covariates without requiring any additivity assumption. It is an extension of the RODEO method, and introduces important new contributions: consistent estimation of the multivariate optimal bandwidth (the *tuning parameter* of the estimator); consistent estimation of the multivariate bias-corrected regression function and confidence bands; and automatic identification and separation of nonlinear and linear effects. Some theoretical properties of the method are discussed. In particular, we show the nonparametric oracle property. For linear models, BID automatically reaches the optimal rate $O_p(n^{-1/2})$, equivalent to the parametric case. A simulation study shows the performance of the procedure for finite samples.

EC1638: **Multivariate coefficients of variation: A full inference toolbox**
Presenter: **Stephanie Aerts**, University of Liege, Belgium
Co-authors: Gentiane Haesbroeck

The univariate coefficient of variation (CV) is a widely used measure to compare the relative dispersion of a variable in several populations. When the comparison is based on p characteristics however, side-by-side comparison of marginal CV's may lead to contradictions. Several multivariate coefficients of variation (MCV) have been introduced and used in the literature but, so far, their properties have not been much studied. Based on one of them, i.e. the inverse of the Mahalanobis distance between the mean and the origin, the aim is to demonstrate the usefulness of MCV's in several domains (finance and analytical chemistry) as well as provide a complete inference toolbox for practitioners. Some exact and approximate confidence intervals are constructed, whose performance is analyzed through simulations. Several bias-correction methods, either parametric or not, are suggested and compared. Finally, since MCV's are used for comparison purposes, some test statistics are proposed for the homogeneity of MCV's in K populations. The robustness of the techniques will be discussed. As a by-product, a test statistic allowing to reliably compare K univariate CV's even in presence of outliers will be outlined.

EC1761: **Multinomial multiple correspondence analysis**
Presenter: **Patrick Groenen**, Erasmus University Rotterdam, Netherlands
Co-authors: Julie Josse

Multiple correspondence analysis (MCA) is a well-known visualization technique for studying the relations between the categories of two or more nominal variables and the similarites between individuals. Often, the results are represented in a biplot showing individuals and the categories simultaneously. We present multinomial correspondence analysis, a maximum likelihood version of MCA based on the multinomial logit model. Multinomial correspondence analysis fits main effects for the categories along with the coordinates needed for the biplot representation for individuals and categories as is usual in MCA. The likelihood of such a model cannot be straightforwardly maximized and we tackle this issue by suggesting a majorization algorithm. The trick is to bound the Hessian with a very simple matrix. It is possible to include in the procedure a ridge and/or nuclear norm penalty to avoid problems due to the large dimensionality of the parameters set. Another advantage is that missing values can be easily handle and we can provide probabilities for each category of the missing variable for an individual.

EC1528: **Causal transmission in reduced-form models**
Presenter: **Vassili Bazinas**, University of Oxford, United Kingdom

We propose a method to explore the causal transmission of a catalyst variable through two endogenous variables of interest. The method is based on the reduced-form system formed from the conditional distribution of the two endogenous variables given the catalyst. The method combines elements from instrumental variable analysis and Cholesky decomposition of structural vector autoregressions. We give conditions for uniqueness of the causal transmission.

EC035 Room MAL 541 CONTRIBUTIONS IN COMPUTATIONAL STATISTICS Chair: Axel Gandy

EC1458: **D-trace precision matrix estimation using adaptive lasso penalties**
Presenter: **Vahe Avagyan**, Universidad Carlos III de Madrid, Spain
Co-authors: Andres M Alonso, Francisco J Nogales

An accurate estimation of a precision matrix has a crucial role in the current age of high-dimensional data explosion. To deal with this problem, one of the prominent and commonly used techniques is the ℓ_1 norm (Lasso) penalization for a given loss function. This approach guarantees the sparsity of the precision matrix estimator for properly selected penalty parameters. However, the ℓ_1 norm penalization often fails to control the bias of the obtained estimator because of its overestimation behavior. We introduce two adaptive extensions of the recently proposed ℓ_1 norm penalized D-trace loss minimization method. The proposed approaches intend to diminish the produced bias in the estimator. Extensive numerical results, using both simulated and real datasets, show the advantage of our proposed estimators.

EC1739: **The chopthin algorithm for resampling**

Presenter: **Din-Houn Lau**, Imperial College London, United Kingdom

Co-authors: Axel Gandy

Resampling is a standard step in particle filters and more generally sequential Monte Carlo methods. We present an algorithm, called chopthin, for resampling weighted particles. In contrast to standard resampling methods the algorithm does not produce a set of equally weighted particles; instead it merely enforces an upper bound on the ratio between the weights. A simulation study shows that the chopthin algorithm consistently outperforms standard resampling methods. The algorithms chops up particles with large weight and thins out particles with low weight, hence its name. It implicitly guarantees a lower bound on the effective sample size. The algorithm can be implemented very efficiently, making it practically useful. We show that the expected computational effort is linear in the number of particles. Implementations for C++, R (on CRAN), Python and for Matlab are available.

EC1314: **Autoregressive conditional duration model with an extended Weibull error distribution**

Presenter: **Boris Choy**, University of Sydney, Australia

Trade duration and daily range data often exhibit asymmetric shape with long right tail. In analysing the dynamics of these positively valued time series under autoregressive conditional duration (ACD) models, the choice of the conditional distribution for innovations has posed challenges. A suitably chosen distribution, which is capable of capturing unique characteristics inherent in these data, particularly the heavy tailedness, is proved to be very useful. A new Weibull distribution is introduced which is shown to perform better than the existing Weibull distribution in ACD and CARR modelling. By incorporating an additional shape parameter, the Weibull distribution is extended to the extended Weibull (EW) distribution to enhance its flexibility in the tails. An MCMC based sampling scheme under a Bayesian framework is employed for statistical inference and its performance is demonstrated in a simulation experiment. Empirical application is based on trade duration and daily range data from the Australian Securities Exchange (ASX). The performance of EW distribution, in terms of model fit, is assessed in comparison to two other frequently used error distributions, the exponential and Weibull distributions.

EC1472: **A new method for circadian gene identification using order restricted inference**

Presenter: **Yolanda Larriba**, Universidty of Valladolid, Spain

Co-authors: Cristina Rueda, Miguel Fernandez

Identification of periodic patterns in gene expression data is important for studying the regulation mechanism of the circadian system. However, the information available is often given by one or two cycles. Consequently, the number of observations is not enough to fit certain models, such as Fouriers models. Some authors have yet developed procedures or algorithms among which is the JTK Cycle Algorithm. We propose a new method to address this question based on order restricted inference, which allows to determine, in terms of an Euclidean or circular order, if the gene expression given is or not cyclic. Validation of the method is made by evaluating of real data sets and simulations. Moreover, we compare the results obtained by the method with others detecting methods developed in the literature, mainly with the JTK Cycle Algorithm.

Sunday 13.12.2015 16:40 - 18:20 Parallel Session J – CFE-CMStatistics

CO458 Room MAL G15 MODELLING AND FORECASTING CYCLICAL FLUCTUATIONS I Chair: Gian Luigi Mazzi

CO0243: **The Great Moderation in historical perspective: Is it that great?**
Presenter: **Lola Gadea**, University of Zaragoza, Spain
The Great Moderation (GM) is widely documented in the literature as one of the most important changes in the US business cycle. All the papers that analyze it use post WWII data. For the first time we place the GM in a long historical perspective, stretching back a century and a half, which includes secular changes in the economic structure and a substantial reduction of output volatility. We find two robust structural breaks in volatility at the end of WWII and in the mid-eighties, showing that the GM still holds in the longer perspective. Furthermore, we show that GM volatility reduction is only linked to expansion features. We also date the US business cycle in the long run, finding that volatility plays a primary role in the definition of the business cycle, which has important consequences for econometricians and forecasters.

CO0265: **The low-variance, high-risk economy: Lessons from the higher moments of MSI-VARs**
Presenter: **Alexander Karalis Isaac**, Warwick, United Kingdom
The aim is to determine whether Markov-switching models capture the non-Gaussian features of economic data evident since the Financial Crisis. We derive exact solutions for the for third and fourth moments of MSI-VARs under mean square stability. This allows us to model the Financial Crisis and the Great Moderation in a single framework. For U.S. data, the post 1983 business cycle describes a low-variance, high-risk economy, with skewness -1.1 and kurtosis 6.6. A Markov-switching model with four states splits the sample irreversibly in 1983 and captures the new moment structure. This enables economists to model both the asymmetry and probability of rare disasters in GDP growth, consistent with data generated in the era of global financial liberalisation.

CO0705: **Nested dynamic factor modeling: A coherent approach to measure national and state coincident indexes**
Presenter: **Juan-Carlos Martinez-Ovando**, ITAM, Mexico
Dynamic factor models have been used as a workhorse to measure business cycles from several economic information. However, when the economic information is available at aggregated and disaggregated levels (state or sectorial), the computations derived from this methodology exhibit some sorts of inconsistencies. A solution to that problem was proposed previously by deriving an ad-hoc procedure to consistently measure coincident indexes for the 50 states of the US economy. We develop an alternative procedure based on the notion of nested dynamic factor model, i.e. a dimensional reduction technique which takes into consideration the information contained in the coincident economic information for the states' economies and the aggregate, simultaneously. Our procedure, in this way, generalizes the approach previously adopted, and allows us to provide a coherent reading of local and aggregated business cycles. We illustrate our proposal by means of computing coherent national and states coincident indexes for the US and Mexico.

CO0854: **Combining composite indicators and advanced graphical tools for monitoring Euro area and member states cycles**
Presenter: **Gian Luigi Mazzi**, Eurostat, Luxembourg
Co-authors: Jacques Anas, Monica Billio, Ludovic Cales
Since several years, Eurostat is monitoring the cyclical situation of the Euro area and its largest economies by means of cyclical composite indicators. Such indicators based on MS-VAR models aim to simultaneously detect peaks and troughs of the growth and business cycles within the so-called ABCD sequence. Furthermore, at the Euro area level, also the acceleration cycle is monitored by means of a univariate MS model. Firstly we present the preliminary results of a project targeting a full coverage monitoring of the Euro area cycles, obtained by developing composite indicators, similar to those already in use, to all Euro area member countries plus the UK. Problems encountered in constructing such indicators, especially due to data availability, are analysed and related solutions are presented. Secondly we show how the results of the cyclical composite indicators can be presented in an intuitive, easy to read and friendly graphical representation. The core of such a graphical tool is constituted by a clockwise representation of the cyclical fluctuations. The characteristics of the tool are presented and some examples are proposed to show the potentials of the tool from the analysts' point of view.

CO562 Room MAL B20 RECENT ADVANCES IN BAYESIAN COMPUTATIONAL METHODS Chair: Gael Martin

CO0295: **Fast and efficient MCMC for large data problems using data subsampling and the difference estimator**
Presenter: **Matias Quiroz**, Stockholm University and Sveriges Riksbank, Sweden
Co-authors: Mattias Villani, Robert Kohn
The aim is to propose a generic Markov Chain Monte Carlo (MCMC) algorithm to speed up computations for datasets with many observations. A key feature of our approach is the use of the highly efficient difference estimator from the survey literature to estimate the log-likelihood accurately using only a small fraction of the data. Our algorithm improves on the $O(n)$ complexity of regular MCMC by operating over local data clusters instead of the full sample when computing the likelihood. The likelihood estimate is used in a Pseudo-marginal framework to sample from a perturbed posterior which is within $O(m^{-1/2})$ of the true posterior, where m is the subsample size. The method is applied to a logistic regression model to predict firm bankruptcy for a large data set. We document a significant speed up in comparison to the standard MCMC on the full dataset.

CO1282: **Accelerating Metropolis-Hastings algorithms by delayed acceptance**
Presenter: **Christian Robert**, Universite Paris-Dauphine, France
Co-authors: Marco Banterle, Clara Grazian, Anthony Lee
MCMC algorithms such as Metropolis-Hastings algorithms are slowed down by the computation of complex target distributions as exemplified by huge datasets. We offer a useful generalisation of the Delayed Acceptance approach, devised to reduce the computational costs of such algorithms by a simple and universal divide-and-conquer strategy. The idea behind the generic acceleration is to divide the acceptance step into several parts, aiming at a major reduction in computing time that out-ranks the corresponding reduction in acceptance probability. Each of the components can be sequentially compared with a uniform variate, the first rejection signalling that the proposed value is considered no further. We develop moreover theoretical bounds for the variance of associated estimators with respect to the variance of the standard Metropolis-Hastings and detail some results on optimal scaling and general optimisation of the procedure.

CO0157: **On consistency of approximate Bayesian computation**
Presenter: **David Frazier**, Monash University, Australia
Co-authors: Gael Martin, Christian Robert
Approximate Bayesian computation (ABC) methods have become increasingly prevalent of late, facilitating as they do the analysis of intractable, or challenging, statistical problems. With the initial focus being primarily on the practical import of ABC, exploration of its formal statistical properties has begun to attract more attention. The aim is to establish general conditions under which ABC methods are Bayesian consistent, in the sense of producing draws that yield a degenerate posterior distribution at the true parameter (vector) asymptotically (in the sample size). We

derive conditions under which arbitrary summary statistics yield consistent inference, with these conditions linked to the identification of the true parameters. Using simple illustrative examples that have featured in the literature, we demonstrate that identification, and hence consistency, is unlikely to be achieved in many cases, and propose a simple diagnostic procedure that can indicate the presence of this problem. We also touch upon the link between consistency and the use of auxiliary models within ABC, and illustrate the subsequent results in a simple Lotka-Volterra predator-prey model. Lastly, we explore the relationship between consistency and the use of marginalization to obviate the curse of dimensionality.

CO1187: **On the properties of variational approximations of Gibbs posteriors**
Presenter: **James Ridgway**, University Paris Dauohine, France
Co-authors: Nicolas Chopin, Pierre Alquier
The PAC-Bayesian approach is a powerful set of techniques to derive non-asymptotic risk bounds for random estimators. The corresponding optimal distribution of estimators, usually called the Gibbs posterior, is unfortunately intractable. One may sample from it using Markov chain Monte Carlo, but this is often too slow for big datasets. We consider instead variational approximations of the Gibbs posterior, which are fast to compute. We undertake a general study of the properties of such approximations. Our main finding is that such a variational approximation has often the same rate of convergence as the original PAC-Bayesian procedure it approximates. We specialise our results to several learning tasks (classification, ranking, matrix completion), discuss how to implement a variational approximation in each case, and illustrate the good properties of said approximation on real datasets.

CO554 Room MAL 414 MIXED-FREQUENCY TIME SERIES Chair: J Isaac Miller

CO0325: **Simple robust tests for the specification of high-frequency predictors of a low-frequency series**
Presenter: **J Isaac Miller**, University of Missouri, United States
Two simple variable addition test statistics are proposed for three tests of the specification of high-frequency predictors in a model to forecast a series observed at a lower frequency. The first one is similar to existing test statistics and we show that it is robust to biased forecasts, integrated and cointegrated predictors, and deterministic trends, while it is feasible and consistent even if estimation is not feasible under the alternative. It is not robust to biased forecasts with integrated predictors under the null of a fully aggregated predictor, and size distortion may be severe in this case. The second test statistic proposed is an easily implemented modification of the first one that sacrifices some power in small samples but is also robust to this case.

CO0461: **The Beveridge-Nelson decomposition of mixed-frequency series**
Presenter: **Yasutomo Murasawa**, Konan University, Japan
Gibbs sampling for Bayesian VAR with mixed-frequency series draws latent high-frequency series and model parameters sequentially. Applying the multivariate Beveridge-Nelson (B-N) decomposition in each Gibbs step, one can simulate the joint posterior distribution of the B-N permanent and transitory components in latent and observable high-frequency series. This method is applied to mixed-frequency series of macroeconomic variables including quarterly real GDP to estimate the monthly natural rates and gaps of output, inflation, interest, and unemployment jointly. The resulting monthly real GDP and GDP gap are complementary coincident indices, measuring classical and deviation cycles respectively.

CO0618: **Time-varying mixed-frequency vector autoregressive models**
Presenter: **Thomas Goetz**, Deutsche Bundesbank, Germany
Co-authors: Klemens Hauzenberger
Many of the existing macroeconomic forecasting models ignore the mismatch in the series sampling frequencies, the possibility of (smooth) structural changes, or the joint dynamics between the variables involved (or all of the above). To simultaneously address the aforementioned data features, we introduce a time-varying parameters mixed-frequency vector autoregressive (TVP-MF-VAR) model. To keep our approach feasible beyond small VARs we limit time variation to the constants and error variances. We estimate the time-varying parameters using two approximation techniques: forgetting factors in the prediction step of the Kalman filter; and exponentially weighted moving averages (EWMA) for the error variances. This approach reduces the computational burden, thus allowing us to evaluate many relatively large VARs (up to 20 variables) in a recursive forecasting exercise in a reasonable amount of time. For a small VAR, we examine the validity of our approximate approach by comparing it to a model that is based on exact MCMC methods. Furthermore, we assess our models forecasting ability by comparing it to a pure TVP-, a pure MF- and a classical VAR using German data.

CO0781: **The estimation of continuous time models with mixed frequency data**
Presenter: **Marcus Chambers**, University of Essex, United Kingdom
We consider exact representations for discrete time mixed frequency data generated by an underlying multivariate continuous time model. Allowance is made for different combinations of stock and flow variables as well as deterministic trends, and the variables themselves may be stationary or nonstationary (and possibly cointegrated). The resulting discrete time representations allow for the information contained in high frequency data to be utilised alongside the low frequency data in the estimation of the parameters of the continuous time model. Monte Carlo simulations explore the finite sample performance of the maximum likelihood estimator of the continuous time system parameters based on mixed frequency data, and a comparison with extant methods of using data only at the lowest frequency is provided. An empirical application demonstrates the methods and some ways in which the present analysis can be extended and refined are discussed.

CO552 Room MAL B33 APPLIED ECONOMETRICS Chair: Michael Owyang

CO0356: **Taylor type monetary policy rules with financial market expectations**
Presenter: **Michael Owyang**, Federal Reserve Bank of St Louis, United States
Co-authors: Eric Ghysels
Taylor rules are often used to characterize the systematic component of monetary policy. In the U.S., changes in the policy rate are typically made at scheduled meetings. The data that are available at these times are of different vintages. We develop a model that accounts for the variation in hard data vintage and uses soft data–high frequency financial data–to update hard data releases in computing expectations. The expectations are then used to estimate Taylor-type rules.

CO0546: **A comprehensive evaluation of macroeconomic forecasting methods**
Presenter: **Ana Galvao**, University of Warwick, United Kingdom
Co-authors: George Kapetanios, Andrea Carriero
The proposed forecasting evaluation compares the performance of four state-of-art multivariate forecasting models: Factor-Augmented Distributed Lag (FADL) Models, Mixed Data Sampling (MIDAS) Models, Bayesian Vector Autoregressive (BVAR) Models and a medium-sized Dynamic Stochastic General Equilibrium Model (DSGE). We look at these models to predict output growth and inflation with datasets from US, UK, Euro area, Germany, France, Italy and Japan. Our evaluation considers both the accuracy of point and density forecasts, and forecast horizons from nowcasting up to two-years ahead. We find predictability of inflation at all horizons, but no predictability of output growth at the two-year-ahead

horizon. MIDAS models are the adequate choice for nowcasting output growth and quarterly inflation, but at longer horizons, BVAR and Factor specifications are a better choice. The medium-sized DSGE model is able to deliver superior long horizon forecast of US and UK inflation. There is no clear evidence that a large set of predictors (one-hundred) may improve the accuracy of forecasts in comparison with a medium set (a dozen predictors). If there are gains from the use of large datasets, they are likely to be during the more recent period (2008-2011) and using BVARs for output growth and combination MIDAS models for inflation. We also observe that UK and US output growth density forecasts of models with large datasets may be better calibrated than with smaller datasets.

CO0564: **Adaptive state space models**
Presenter: **Ivan Petrella**, Bank of England, United Kingdom
Co-authors: Davide Delle Monache, Fabrizio Venditti
The estimation of state-space models with time-varying parameters typically implies the use of computationally intensive methods. Moreover, when volatility evolves stochastically the model ceases to be conditionally Gaussian and requires nonlinear filtering techniques. We model parameters' variation in a Gaussian state-space model by letting their dynamics to be driven by the score of the predictive likelihood. In this setup, conditionally on past data, the model remains Gaussian and the likelihood function can be evaluated using the Kalman filter. We derive the analytical expressions for the score and the information matrix which are needed to update the time varying system matrices. We show that this leads to a new set of recursions running in parallel with the standard Kalman filter recursions. The resulting algorithm allows us estimate simultaneously the unobserved state vector and the time-varying parameters by maximum likelihood. The model is further extended to handle data at mixed frequencies.

CO1058: **Real-time forecasting with a large, mixed frequency, Bayesian VAR**
Presenter: **Tatevik Sekhposyan**, Texas A and M University, United States
Co-authors: Michael McCracken, Michael Owyang
Point forecasts from a large, mixed-frequency, structural vector autoregression (VAR) are assessed. The VAR we consider uses data at monthly and quarterly frequencies to obtain forecasts of low frequency variables such as output growth on a more frequent basis. The structure imposed on the VAR allows us to account for the temporal ordering of the data explicitly, thus accounting for the effects of temporal surprises across the variables in a more interpretable manner. Our framework relies on a blocking model, i.e. econometric model specified at a low frequency, where high frequency observations of a particular variable are stacked, i.e. treated as individual economic series occurring at the low frequency. Since stacking results in a high- dimensional system of equations, we rely on Bayesian shrinkage techniques to mitigate parameter proliferation. We use our model for short-term forecasting of the U.S. economy, as well as for structural analysis. The relative performance of the model is compared to the factor model and private sector forecasts.

CO540 Room MAL B35 ECONOMETRICS OF DYNAMIC PORTFOLIOS AND RISK Chair: Jean-Michel Zakoian

CO0507: **Real uncertainty and the zero lower bound**
Presenter: **Guillaume Roussellet**, NYU Stern School of Business, United States
Both term structures of U.S. nominal and inflation-linked bonds are used to identify real uncertainty and the associated risk premia, in and out of the zero lower bound (ZLB). Regression analyses are first used to provide stylized facts on real term and inflation risk premia, and to derive new Fama conditions. We propose a ZLB-consistent affine pricing model for both term structures, encompassing simultaneously flexible inflation dynamics and providing non-negative nominal yields. We extract risk premia components, showing their consistency with the stylized facts. Decomposing the sources of real uncertainty, we document that although short-term inflation uncertainty is high at the ZLB, the predictability of nominal and real excess-returns is improved during this period.

CO0511: **Filtered historical simulations for estimating the conditional risk of a dynamic portfolio**
Presenter: **Christian Francq**, CREST and University Lille III, France
The estimation of the conditional Value-at-Risk (VaR) of a portfolio of assets is considered. The composition of the portfolio is time-varying and the vector of returns is assumed to follow a multivariate GARCH-type model. Under the assumption that the distribution of the innovations is spherical, the asymptotic distribution of an estimator of conditional VaR is established. We also derive the asymptotic properties of the so-called Filtered Historical Simulation (FHS) method, which does not need the sphericity assumption. We compare the FHS method with the method based on the sphericity assumption, via Monte Carlo experiments and empirical studies, and illustrate the superiority of the two multivariate approaches over a univariate approach based on the sole series of portfolio returns.

CO1369: **Deep conditional portfolio sorts**
Presenter: **Benjamin Moritz**, Ludwig Maximilian University of Munich, Germany
Co-authors: Tom Zimmermann
Which variables provide independent information about the cross-section of future returns? Standard techniques like portfolio sorts and Fama-MacBeth regressions cannot easily answer this question when the number of candidate variables is large and when cross-terms might be important as well. We introduce a new method, deep conditional portfolio sorts, that can be used in this context. To estimate the model, we import ideas from the machine learning literature and tailor them to our setting. We apply the method to past-return based predictions, and we recover short-term returns (i.e. the past six most recent one-month returns) as the most important predictors. A trading strategy based on these findings has Sharpe and information ratios that are about twice as high as in a Fama-MacBeth framework that accounts for two-way interactions. Transaction costs do not explain these results. Implications for the analysis of cross-sectional predictor variables going forward are discussed in the conclusion.

CO0782: **On the empirical saddlepoint approximation with application to asset pricing**
Presenter: **Benjamin Holcblat**, BI Norwegian Business School, Norway
We prove new theoretical results regarding the ESP (empirical saddlepoint) approximation, and apply the latter one to consumption-based asset pricing. Firstly, we prove the existence of the ESP estimand, which is the intensity of the solutions to estimating equations. The challenge of this proof comes from the possible multiplicity of the solutions. Secondly, we prove global consistency and asymptotic normality of the ESP approximation. The application suggests that the basic consumption-based asset-pricing model is more consistent with data than other inference approaches suggest.

CO414 Room MAL B34 DENSITY REGRESSION, TREE MODELS, AND VARIABLE SELECTION Chair: Carlos Carvalho

CO0601: **Shrinkage estimation of treatment effects: Dealing with many controls**
Presenter: **Carlos Carvalho**, The University of Texas at Austin, United States
Co-authors: Richard Hahn
A shrinkage strategy is presented to estimate linear treatment effects in the presence of potentially very many controls. The approach is based on a re-parametrization that allows us to identify treatment effects by leveraging all the positive aspects of variable selection and shrinkage priors. The method looks to find the sweet spot in the bias-variance trade-off where we reduce the variability of simple OLS while avoiding the extreme bias associate with naive applications of bayesian variable selection.

CO0629: **Multiscale spatial density smoothing**
Presenter: **James Scott**, University of Texas at Austin, United States
The estimation of a spatially varying density function is considered, motivated by problems that arise in large-scale radiological survey and anomaly detection. Four challenges make this a difficult problem. First, the density at any given spatial location may have both smooth and non- smooth features. Second, the spatial correlation is neither stationary nor isotropic. Third, the spatial correlation decays at different length scales for different parts of the density. Finally, at some spatial locations, there is very little data. We present a method called multiscale spatial density smoothing that successfully addresses these challenges. The method is motivated by the same construction that underlies a Polya-tree prior, in that it is based on a recursive dyadic partition of the underlying density function. We also describe an efficient algorithm for finding a maximum a posteriori (MAP) estimate that leverages recent advances in convex optimization for non-smooth functions.

CO1215: **Block hyper-g priors in Bayesian regression**
Presenter: **Christopher Hans**, The Ohio State University, United States
Thick-tailed mixtures of g priors have gained traction as a default choice of prior distribution in Bayesian regression. The motivation for these priors usually focuses on properties of model comparison and variable selection as well as computational considerations. Standard mixtures of g priors mix over a single, common scale parameter that shrinks all regression coefficients in the same manner. The particular form of the mixture distribution determines the model comparison properties. We focus on the effect of the mono-shrinkage induced by use of a single scale parameter and propose new mixtures of g priors that allow for differential shrinkage across collections of coefficients. We introduce a new "conditional information asymptotic" that is motivated by the common data analysis setting where at least one regression coefficient is much larger than others. We analyze existing mixtures of g priors under this limit and reveal two new behaviors, "Essentially Least Squares (ELS)" estimation and a "Conditional Lindleys Paradox (CLP)", and argue that these behaviors are undesirable. As the driver behind both of these behaviors is the use of a single, latent scale parameter that is common to all coefficients, we propose a block hyper$-g$ prior that allows for differential shrinkage across collections of covariates and provide conditions under which ELS and the CLP are avoided by the new class of priors.

CO1285: **Density regression with Bayesian additive regression trees**
Presenter: **Jared Murray**, Carnegie Mellon University, United States
Modeling how an entire density changes with covariates ("density regression") is an important but challenging generalization of mean and quantile regression models. We introduce a new continuous latent variable model for density regression. Treating this unobserved variable as the input to a nonparametric regression function induces a flexible model for the conditional density of the response (given the observed covariates) after marginalizing over the latent variable. This model has a natural interpretation in terms of omitted variables, and only requires prior distributions to be specified for one or two regression functions (in contrast to covariate-dependent mixture models). Bayesian additive regression trees (BART) are used as priors over location and scale (or bandwidth) regression functions, yielding attractive invariance properties and computationally efficient posterior inference.

CO657 Room MAL B36 REGIME CHANGE MODELING IN ECONOMICS AND FINANCE II Chair: Marco Gross

CO1123: **Simulated ML estimation of a financial agent-based herding model**
Presenter: **Jiri Kukacka**, Charles University in Prague - Faculty of Social Sciences, Czech Republic
Co-authors: Jozef Barunik
We apply the very recent simulated MLE methodology to a stylised financial agent-based herding model where noise traders switch between the optimistic and pessimistic states. We test small sample properties of the estimator via Monte Carlo simulations and confirm important theoretical features of the estimator such as consistency and asymptotic efficiency. Via exploring behaviour of the objective simulated log-likelihood function we also verify the identification of parameters and theoretical assumptions of the estimation method. Next, we estimate the model using three stock market indices (DAX, SP500, and Nikkei), price of gold in USD, and three exchange rates (USD/EUR, USD/YEN, and CHF/YEN). Results of the full sample as well as rolling simultaneous estimation of parameters a and b governing switches of opinion and sentiment dynamics together with standard deviation of the innovations of the fundamental value are presented. Finally, we compare and contrast the performance of the NPSMLE method to the simulated method of moments approach.

CO1345: **Convex Phillips curves: Literature review, a theoretical model and an empirical analysis for the Euro area**
Presenter: **Marco Gross**, European Central Bank, Germany
Co-authors: Willi Semmler
We develop a theoretical model that features a state-dependent relation between output, price inflation and inflation expectations, augmenting a previous model with a nonlinear Phillips curve that reflects the rationale underlying the capacity constraint theory. Our empirical assessment for the Euro area backs the theory - based on a regime-switching Phillips curve and a regime-switching monetary structural VAR - by confirming the presence of a significant *convex* relationship between inflation and the output gap. Convexity means that the beta of inflation on the output gap increases during times of economic expansion and abates during times of recession. The regime switching monetary SVAR reveals the business cycle dependence of macroeconomic responses to monetary policy shocks: Expansionary monetary policy (be it via conventional or unconventional measures) induces less pressure on price inflation at times of weak growth as opposed to strong growth; thereby rationalizing relatively stronger expansionary policy, including unconventional volume-based policy such as the Expanded Asset Purchase Programme (EAPP) of the ECB, at times of recession.

CO1386: **Macroeconomic regime switching and technological change**
Presenter: **Tommaso Ferraresi**, University of Pisa and IRPET, Italy
Co-authors: Willi Semmler, Andrea Roventini
We assess the effects of (adjusted) TFP on GDP and hours worked in bad and good times. More precisely, we estimate over several time spans (the longest being 1950:1-2011:4) different threshold vector autoregressions (TVAR) allowing for different threshold variables (e.g. GDP growth and financial stress indexes) and we assess the paths of generalized impulse response functions in order to judge whether TFP shocks differently affect hours and output according to the state of the economy. Moreover, consistently with the prevailing literature, we assess whether the relation is time

dependent and investigate some plausible explanations. As to our results, first, TFP shocks while spurring GDP growth, display a negative effect on hours worked at least on impact, independently of the state of the economy. However, the effects on aggregate demand depend on the sampled period, with a milder effect during the Great Moderation. Moreover, there is evidence of state dependent effects of TFP shocks irrespectively of the threshold variable at hand (either GDP growth or the financial stress index). However, the impact of TFP shocks in good or bad times depends on the chosen sample. In particular, we observe a rising effects of TFP shocks in bad times during the Great Moderation. According to our empirical exercises, this could be partly due to better stabilization policies, skill-biased technical change, labor market liberalization and financial development.

CO0969: **On the instability of the market beta: A three regime threshold GARCH model**
Presenter: **Fredj Jawadi**, University of Evry, France
Co-authors: Abdoulkarim Idi cheffou, Ben Ameur Hachmi, Louhichi Wael
The aim is to revisit an important issue in Finance associated with the instability of the market beta and therefore diversification strategies. In particular, we propose to capture further instability through the extension of the market model to a nonlinear framework. To this end, we develop an On/Off specification that captures further time variation in the market beta while identifying three main states: i) Low state market, ii) Medium state market, iii) High state market. Interestingly, our model identifies endogenously these states and captures the transition from one state to another according to the activation of an endogenous threshold. Thus, it provides evidence of different type of risk diversification according to the regime under consideration. The application of this three regime asymmetrical model to conventional and Islamic stock indexes provides two interesting results. First, the threshold model supplants the linear market model to capture asymmetry and instability in the market beta. Second, our nonlinear market model shows that the consideration of nonlinearity and asymmetry in estimating beta enables to improve investment and diversification strategies.

CO374 Room MAL 415 NON-LINEAR TIME SERIES MODELS Chair: Alessandra Canepa

CO1192: **Dynamic asymmetries in house price cycles: A generalized smooth transition model**
Presenter: **Alessandra Canepa**, Brunel University, United Kingdom
Dynamic asymmetries in house price cycles are investigated. To capture real estate cycles, an ad-hoc non-linear model is proposed. The suggested model involves using a particular parametrization of the logistic function used in the transition equation of a smooth transition model which improves the fit in the non-central probability region. Estimation of the empirical model suggests that dynamic symmetry in house price cycles is strongly rejected in a number countries taken into consideration. Further, our results show that the proposed model performs well in a out of sample forecasting exercise.

CC1829: **Mortage default, property price and banks lending behaviour in Hong Kong SAR**
Presenter: **Fawaz Khaled**, Brunel University, United Kingdom
The aim is to investigate the long-run equilibrium relationship and short run dynamic between the mortgage default, property prices, loan-to-value and bank lending behaviour in Hong Kong by employing the Autoregressive-Distributed Lag (ARDL) bounds test technique for cointegration on time series. Our findings indicate the existence of cointegrating relationships between banks' lending and property prices which governs the correction mechanism between banks' lending, property prices and mortgage default in the long-run. Particularly, we find positive impact of property prices and loan-to-value on the evolution of mortgage default. However, the role of banks lending behaviour on the level of mortgage default is found to be negative. Our finding shows that any disequilibrium in this relation is corrected and converges back, with a relatively good speed of adjustment, to its long-run equilibrium.

CO1832: **Bank lending and house price: The Hong Kong experience**
Presenter: **Yuefeng Wang**, Brunel University, United Kingdom
Cyclical patterns in real estate markets are investigated. Particular attention is given to the relation between housing market cycles and bank lending policy. Using nonlinear smooth transition model it is found that transition from booms to busts in housing market are endogenously driven by bank lending policy.

CC1667: **Stock market volatility and economic variables: A nonlinear approach**
Presenter: **Eduardo Rossi**, EC-JRC, Italy
The aim is to examine the role of macroeconomic and financial variables in predicting the monthly realized variance of the S&P500. Differently from existing studies, we focus on the functional form specification. We provide evidence that the relationship between macroeconomic and financial variables and the future monthly stock market volatility is nonlinear. To model the nonlinearity we adopt a smooth transition regression. We compare alternative nonlinear specifications characterized by different transition variables. The specification which prevails according to the standard specification tests is the one based on a threshold variable expressed as a linear function of a set of leading economic variables. Out-of-sample forecasting comparisons corroborates the in-sample results.

CO432 Room MAL B30 VOLATILITY MODELLING IN FINANCIAL MARKETS Chair: Menelaos Karanasos

CO1236: **Exchange rate volatility and emerging market portfolio flows**
Presenter: **Faek Menla Ali**, Brunel University London, United Kingdom
The aim is to examine empirically the impact of exchange rate uncertainty on the level and the variability of net equity flows of the US vis-vis 14 developing and emerging countries, with monthly data over the period January 1993-November 2012. Our model is based on a near bivariate near VAR GARCH-BEKK-in-mean and allow the exchange rate uncertainty to compete with pull and push factors. The results indicate that the impact of exchange rate uncertainty on net equity inflows towards the US is positive in seven countries, negative in three countries, whilst four countries showed insignificant responses. These findings suggest that an increase in the volatility of the US dollar against the currencies of developing and emerging countries induces equity inflows from these countries towards the US. Furthermore, the results provide evidence of volatility transmission between exchange rate changes and net equity flows, with the causality running mostly from the latter to the former. Therefore, as a policy tool, credit controls on equity flows may be deemed to stabilise the foreign exchange markets.

CO1240: **Inflation convergence in the EMU and the link between inflation differentials and their uncertainty**
Presenter: **Panagiotis Koutroumpis**, Brunel University, United Kingdom
Co-authors: Menelaos Karanasos, Yiannis Karavias
The aim is to provide evidence over the behavior of European inflation rates covering a period from 1980Q1 to 2013Q4. By applying univariate unit root and stationarity tests we show that convergence occured before the implementation of the common currency in 1999. Moreover, the influence of exchange rate mechanism can be attributed as positive since it stimulated the convergence process. For the high inflation countries deviations from the ERM policies led to higher inflation rates that were above the average.

CO1804: **Modelling time varying volatility spillovers and conditional correlations across commodity metal futures**
Presenter: **Menelaos Karanasos**, Brunel University, United Kingdom
Co-authors: Faek Menla Ali
The aim is to examine how the most prevalent stochastic properties of key metal futures returns have been affected by the recent financial crisis. Our results suggest that copper and gold futures returns exhibit time varying persistence in their corresponding volatilities during the crisis period. The estimation of a bivariate GARCH model further shows the existence of time varying shock and volatility spillovers between these returns during the different stages of such a crisis. Our results are broadly robust irrespective of whether mapped or unmapped data are employed.

CO1841: **Stock markets response to MPC unconventional monetary policy**
Presenter: **Emmanouil Noikokyris**, Kingston University London, United Kingdom
Co-authors: Menelaos Karanasos, Georgios Chortareas
The effects of Monetary Policy Committee's (MPC) asset purchase announcements and communications on the UK stock market are explored for the period 2009-2014. Using intraday aggregate stock market data in an event study framework, we assess the equities reaction, both their level and volatility, to survey-based measures for monetary policy stance over a variety of time frames both preceding and following the MPC announcements and Banks communications. Our results show significant effects of monetary policy shocks and communications on the level and volatility of UK equities which can be traced also on the days following the announcement.

CG493 Room MAL 540 CONTRIBUTIONS ON VAR AND EXTREME VALUE THEORY Chair: Stefan Straetmans

CC0989: **Specification testing in Hawkes models**
Presenter: **Francine Gresnigt**, Erasmus University Rotterdam, Netherlands
Co-authors: Philip Hans Franses
We propose various specification tests for Hawkes models based on the Lagrange Multiplier (LM) principle. Hawkes models can be used to model the occurrence of extreme events in financial markets. Our specific testing focus is on extending a univariate model to a multivariate model, that is, we examine whether there is a conditional dependence between extreme events in markets. Simulations show that the test has good size and power, in particular for sample sizes that are typically encountered in practice. Applying the specification test for dependence to US stocks and bonds data, we find evidence for the stock-bond contagion as well as for the flight-to-quality phenomenon.

CC1419: **Realized Peaks-Over-Threshold: A high-frequency extreme value approach for financial time series**
Presenter: **Luca Trapin**, IMT Institute for Advanced Studies Lucca, Italy
Co-authors: Marco Bee, Debbie Dupuis
Recent advances in financial econometrics have emphasized how high-frequency information can be incorporated into models for daily asset returns. We propose an extension of the classical Peaks-Over-Threshold method from Extreme Value Theory, with the aim of exploiting high-frequency data to model the time-varying behaviour of daily extreme returns. Specifically, we provide an approach that estimates and forecasts the tails of the conditional return distribution while accounting for stock returns dependence. Considering extreme returns the observations exceeding a high threshold, we model the exceedance rate of this threshold with Logit models and Poisson processes, while the size of the exceedances is assumed to follow a Generalized Pareto distribution with dynamic parameters. The in-sample fit of these models to the S&P 500 index shows that time variation in the extremes of daily returns can be modelled through high frequency measures such as the realized variance. Moreover, out-of-sample forecasts of the Value-at-Risk are superior to those from simpler models.

CC1672: **Forecasting value-at-risk for BRICS financial markets using high quantile estimation methods**
Presenter: **Pilar Grau**, Universidad Rey Juan Carlos, Spain
Co-authors: Luis Miguel Doncel, Jorge Sainz
The financial risk management depends critically on the assumptions about the distribution of asset returns. Although it is often assumed that the returns follow a normal distribution, the probability of large price movements is often greater than would be expected from the normal distribution. Thus, the results obtained when this distribution is used tend to underestimate the potential financial risks. Extreme Value Theory focuses on large price movements and their probabilities. Its application for the analysis of financial markets is not new. Although some authors state that it produces better results than other traditional methodologies, there is some controversy. We analyze extreme financial events in emerging markets using data from the major stock indices in emerging countries grouped within the acronym BRICS. We use the so-called high quantile (HQ) estimators, based on extreme value theory (EVT), in particularly through DPOT (duration-based Peak Over Threshold) and quasi-PORT (peaks over a threshold random) to estimate Value-at-Risk. The relative performance of these measures is benchmarked against alternative methods more commonly used as Riskmetrics and GARCH-type approaches. The empirical results show than this methodologies increase predictive performance according to number and independence of violations. They are also consistent with Basel agreements.

CC0178: **Modeling conditional skewness and kurtosis: A VaR application**
Presenter: **Luis Melo**, Universidad del Rosario, Colombia
Co-authors: Andres Jimenez
Standard methodologies for computing Value at Risk (VaR) and Conditional Value at Risk (CVaR) only consider the dynamics of the first two moments of the data generating process while assuming that all other moments are held constant. In contrast, we model the first four moments of the series. In particular, we use AR-GARCH specifications for both the mean and variance and employ various parametric specifications to estimate the conditional skewness and kurtosis. Additionally, we compare the performance of our models with traditional VaR and CVaR techniques for the period of 2008-2014, using Colombian financial data. Our findings indicate that in general, risk measures of our models outperform those of traditional methodologies.

CG377 Room MAL 421 CONTRIBUTIONS ON QUANTILE REGRESSION IN FINANCE AND ECONOMICS Chair: Matthew Harding

CC1184: **Improving accuracy of value at risk estimation using intra-day data**
Presenter: **Xiaochun Meng**, Oxford Unuversity, United Kingdom
Co-authors: James Taylor
Some novel VaR approaches are proposed that utilize intra-day data and overnight night return. These approaches incorporate lagged intra-day range, daily high, daily low and overnight return. A new parameter estimation procedure based on quantile regression is proposed and used. The proposed approaches are very flexible and easy to implement. An empirical analysis is conducted on five market indices. The performances of the proposed models, including a comparison with the established benchmark models, are examined using three standard VaR back-testing methods. The post-sample results show that the proposed models successfully capture the main characteristics of the financial returns and perform very competitively. One of the proposed models utilizing both the intra-day data and the overnight return is shown to forecast VaR more accurately than other models.

CC1563: **The quantile-heterogenous autoregressive model of realized volatility: Evidence from commodity markets**
Presenter: **Konstantin Kuck**, University of Hohenheim, Germany
Co-authors: Robert Maderitsch
A cross-asset perspective is provided on the state-dependence in the intertemporal dynamics of (realized) price volatility in the commodities market. Using high-frequency data for futures on Gold, Silver and Light Sweet Crude Oil, covering a period from 2007 to 2014, we estimate various Quantile-Heterogeneous Autoregressive models of daily realized volatility (Q-HAR-RV). Daily volatility is expressed as linear function of own lags measured over three different time horizons, particularly, the first lag as well as averages over the preceding 5 and 22 trading days. This allows to identify potential state-dependence and asymmetry in the short-, mid- and long-term autoregressive dynamics with respect to different volatility levels. Most importantly, we document considerable differences concerning both the structure and degree of volatility persistence regarding components realized over different time horizons which appear to be similar across the three assets. For instance, the daily and the monthly volatility components are found to decrease from lower to higher quantiles, indicating lower tail dependence. For the weekly volatility aggregate by contrast, we observe a distinct increase from lower to higher quantiles of the conditional volatility distribution.

CC1707: **Bayesian endogenous Tobit quantile regression**
Presenter: **Genya Kobayashi**, Chiba University, Japan
Parametric $p-$th Tobit quantile regression models with endogenous variables are proposed. The first stage regression of the endogenous variable on the exogenous variables where the $a-$th quantile of the error term is zero is introduced into the hierarchical model. Then, the residual of this regression model is included in the $p-$th quantile regression model in such a way that the regressors are uncorrelated with the error term. To meet the zero $a-$th quantile restriction, the error distribution of the first stage regression is modelled by using the asymmetric Laplace distribution, normal distribution, and their nonparametric scale mixtures. Since the value is a priori unknown, it is treated as an additional parameter and is estimated from the data. The proposed models are then demonstrated by using simulated data and real data on the labour supply of married women.

CC1371: **A copula quantile approach to conditional Value-at-Risk estimation**
Presenter: **Krenar Avdulaj**, Department of Econometrics-IITA-The Czech Academy of Sciences, Czech Republic
Co-authors: Jozef Barunik
We propose to use copula quantile regression models and realized volatility to estimate Value-at-Risk (VaR) of an institution conditional on some other institution being under financial distress. The proposed model uses copulas from elliptical family, Normal and t copula, and the realized volatility measure which is calculated using 5 minutes returns. Contrary to the literature which studies the systemic risk of financial institutions, we estimate the risk contribution of an institution to some other institution. We apply the model on 21 most liquid U.S. stocks from seven main market sectors. We find that stocks from financial sector have the highest risk transmission among each other, followed by the Information Technology sector. Consumer Staples industry stocks and Health Care have the lowest risk transmission.

CG349 Room MAL 539 CONTRIBUTIONS ON TIME-VAYING PARAMETERS AND KALMAN FILTER Chair: Rodney Strachan

CC1452: **Shrinking time varying VARs: A regularized score driven approach**
Presenter: **Davide Delle Monache**, Bank of Italy, Italy
Co-authors: Massimiliano Marcellino, Ivan Petrella, Fabrizio Venditti
Vector autoregressive models with time-varying parameters have become a popular tool both in structural macroeconomics as well as in forecasting. However, as the number of variables included in the model increases, their rich parameterization may rapidly result in a large number of drifting coefficients, with detrimental consequences for predictive accuracy as well as for impulse response analysis. One solution is to discipline the behavior of the time-varying parameters through an informative prior, thus reducing estimation uncertainty. We propose a method to shrink the time-varying parameters towards a given prior using a score-driven approach, with the updating mechanism governed by the regularized score of the log-likelihood function. Shrinkage is obtained by means of stochastic constraints, which work as additional artificial observations that tilt the model parameters towards their desired value. We show how to implement in this framework popular shrinkage methods, such as the Minnesota Prior, the sum of coefficients prior, long-run restrictions priors, as well as priors implied by structural models such as the term structure model and the DSGE model. In the empirical application we show the usefulness of the proposed approach in terms of forecasting performance and impulse response analysis.

CC0269: **Present value of houses: A state space approach**
Presenter: **Dooruj Rambaccussing**, University of Dundee, United Kingdom
The time series of expected returns and expected rent growth are derived from a state space representation of the present value of houses. Expected returns are found to be persistent and time varying. Expected rent growth is less persistent, but far from stochastic. Both series are good predictors for realized returns and rent growth, and are marginally better predictors than a present value constrained VAR representation. Variance decompositions show that most of the variation in house prices is caused by variation in discount rates. On the other hand unexpected returns may be caused by news from both discount rates and rent growth.

CC1388: **Time-dependent Black-Litterman**
Presenter: **Martin van der Schans**, Ortec Finance, Netherlands
Many (institutional) investors support their investment decisions with models that forecast future risk and return. In practice, it is often necessary to combine these models with views, e.g. for stress-testing purposes or because the views contain additional forward-looking information. With the Black-Litterman method, views can be combined with models in one-period setups: a density forecast for returns over the investor's investment horizon is combined with the investor's views expressed as expected returns over the investor's investment horizon. In practice, however, this setup is often too limited: the investor's views are formulated for only the near future while the investor's investment horizon is much longer. Using a time-series modelling approach, we extend the original Black-Litterman method to a time-dependent Black-Litterman method which is suitable for time- dependent setups: time-dependent views can be combined with time-dependent density forecasts varying over the investor's investment horizon. Through a stylized example, we show how the method can be used by institutional and other long-term investors in their risk management process.

CC1753: **Firm's dynamics and business cycle: New disaggregated data**
Presenter: **Emilio Zanetti Chini**, University of Pavia, Italy
Co-authors: Lorenza Rossi
We provide new stilized facts on firms dynamics. In particular, we improve the availability of the US data by disaggregating the only official long-span dataset on Business Dynamic Statistics from US Census Bureau, published at yearly frequency. We consider three different methods: i) a univariate dynamic regression in an unobserved component scenario; ii) a multivariate regression; iii) a multivariate dynamic regression. The methodologies based on augmented Kalman Filter conveys computationally more efficient and accurate results than previous techniques adopted.

CC030 Room MAL 541 CONSTRIBUTIONS IN FINANCIAL APPLICATIONS Chair: Michael Rockinger

CC1576: **Estimating nominal share price preferences**
Presenter: **Maria Chiara Iannino**, University of Vienna, Austria
Co-authors: Sergey Zhuk
A dynamic structural model of stock splits is proposed, which allows us to estimate the preferences of investors about nominal share price levels from stock split announcements data. The split announcement premium depends not only on such preferences, but also on the degree of asymmetric information between managers and investors. If information is symmetric there is no announcement premium. Our model allows us to decompose the effect of these factors. We estimate the model with a generalized method of moments, and our results are consistent with non-trivial nominal share price preferences. When prices decrease by 50% from the optimal level, investors require an additional premium of 23 basis points.

CC1786: **The foreign exchange market, the U.S. dollar and the oil price in an open medium-sized economy**
Presenter: **Katarzyna Leszkiewicz-Kedzior**, University of Lodz, Poland
Co-authors: Piotr Keblowski, Aleksander Welfe
The currencies of European countries that have not joined the Euro area are usually recognized to be related to the euro. This is intuitively a self-explanatory approach to the exchange rate modelling, due to the balance of payments perspective and high share of intra-EU trade. Using the Polish zloty as an example, we show that the currencies of European medium-sized economies are also affected by the euro-dollar exchange rate, therefore the tripolar model is an appropriate framework for exchange rate modelling. The euro-dollar exchange rate is found to be determined in the long-run by the crude oil prices, which prevail the terms of trade in fact, on the one hand. On the other hand, it is affected by the disparity of real (risk-free) interest rates, which in turn is stationary in itself. We show also that credit default swaps are important factors driving the exchange rates through the channel of the capital account (real risk-free interest rates) as well as through the channel of the current account (in case of medium-sized economies). The analysis is based on the conditional VEC model with four cointegration vectors, also the impulse-response functions from the SVMA model are reported.

CC1113: **Importance of network positioning in the interbank market**
Presenter: **Asena Temizsoy**, City University, United Kingdom
Co-authors: Giulia Iori, Gabriel Montes-Rojas
The effect of local and global network measures on interest rate spreads in the e-MID interbank market is empirically investigated. We hypothesize that not only the bank's position in the network is important, but also the counterparty's interconnectedness. The empirical results show that interbank spreads are significantly affected by the banks positioning in the network, measured by both local and global connectedness measures. Overall, lenders are willing to pay a premium (i.e. obtain lower rates) for better connections in the network, at both local and global measures. This effect is statistically significant for the pooled sample and for post-Lehman's Brothers collapse sub-period. Borrowers, on the other hand, pay a higher premium (i.e. higher rates) for better local connections, but significantly benefit for better global positioning.

CC1570: **Price drift before U.S. macroeconomic news: On private information about public announcements**
Presenter: **Georg Strasser**, European Central Bank, Germany
Co-authors: Alessio Sancetta, Alexander Kurov, Marketa Halova Wolfe
We examine stock index and Treasury futures markets around releases of U.S. macroeconomic announcements. Seven out of 18 market-moving announcements show evidence of substantial informed trading before the official release time. Prices begin to move in the "correct" direction about 30 minutes before the release time. The pre-announcement price drift accounts on average for about half of the total price adjustment. These results imply that some traders have private information about macroeconomic fundamentals. The evidence points to leakage and proprietary data collection as the most likely sources of that private information.

CG533 Room MAL B29 CONTRIBUTIONS ON SYSTEMIC RISK Chair: Christophe Hurlin

CC1679: **Systemic risk in the American financial system: A view from the top**
Presenter: **Hanane Dakhli**, Sorbonne, France
Previous research and the recent global financial crisis show that there is a persistent requirement of managing properly systemic risk in the financial system. The level of predictability of the recent financial crisis is empirically assessed based on five cross-sectional measures of systemic risk and contagion effects (Conditional Value-at-Risk, Distressed Insurance Premium, Co-Risk, Marginal Expected Shortfall and Systemic Expected Shortfall). Also, in order to capture the link between systemic risk and a range of balance sheet indicators related to a specific financial company, over 2007-2015, we use OLS panel estimation based on a sample of major US financial firms.

CC1557: **Sovereign adaptive risk modeling**
Presenter: **Morgan Escalera**, Rose Hulman Institute of Technology, United States
Co-authors: Wayne Tarrant
In the wake of the 2008 financial crisis, the FSB (Financial Stability Board) and the BCBS (Basel Committee on Banking Supervision) created a list of Globally Systematically Important Banks with the intention of determining which financial institutions were important enough to the global market that their failure would result in total systemic collapse. The purpose is to use econometrics and statistical analysis to create a mathematical model that generalizes the BCBSs five criteria that define the financial institutions and apply them to governmental bodies. These five criteria are size, interconnectedness, cross-jurisdictional activities, complexity, and substitutability. The model is created by utilizing a series of weighted directed graphs to analyze the interconnectedness of central banks in the system as well as creating a market valuation of each country based on their one-year bond yields and a complete regard of their holdings. The original application of the model was for the troubled economy of Greece in the Eurozone. Time permitting, we plan to discuss the cases of other countries such as Argentina or Saudi Arabia.

CC1467: **Large-dimensional factor modeling based on high-frequency observations**
Presenter: **Markus Pelger**, Stanford University, United States
A statistical theory is developed to estimate an unknown factor structure based on financial high-frequency data. We derive a new estimator for the number of factors and derive consistent and asymptotically mixed-normal estimators of the loadings and factors under the assumption of a large number of cross-sectional and high-frequency observations. The estimation approach can separate factors for normal continuous and rare jump risk. The estimators for the loadings and factors are based on the principal component analysis of the quadratic covariation matrix. The estimator for the number of factors uses a perturbed eigenvalue ratio statistic. The results are obtained under general conditions, that allow for a very rich class of stochastic processes and for serial and cross-sectional correlation in the idiosyncratic components.

CC1760: **Measuring nominal yield risk**
Presenter: **Galen Sher**, University of Oxford, United Kingdom
A framework is presented for modelling the effect of deformations in yield curves on the solvency position of financial intermediaries. The framework nests three strands of literature on measuring interest rate/credit risk as special cases. We show how to measure the sampling uncertainty in these risk measures due to yield curve estimation and measurement error in accounting data. This framework is applied to measure the sensitivity of the solvency position of US banks to revaluation in their loan books under yield curve scenarios recommended by the Bank for International Settlements.

CG353 Room MAL 402 CONTRIBUTIONS ON TIME SERIES Chair: Carlos Velasco

CC0281: **Approximating and reducing bias in 2SLS estimation of dynamic simultaneous equation models**
Presenter: **Gareth Liu-Evans**, University of Liverpool, United Kingdom
Co-authors: Garry Phillips
An order $O(T^{-1})$ approximation is made to the bias in 2SLS estimation of a dynamic simultaneous equation model, building on similar large-T moment approximations for non-dynamic models. The expression is long because it contains two distinct parts: a part due to the simultaneity which is directly related to the Nagar bias and a part due to the dynamics which has many component terms. However, the approximation is seen to work well when used for bias correction. We compare the bias of the resultant bias corrected estimator with the Quenouille half-sample jackknife and the residual bootstrap for 2SLS in dynamic models due to Freedman, and find that it is competitive, particularly in terms of mean squared error. The Monte Carlo and bias approximation also suggest that the bias in estimating endogenous variable coefficients in dynamic simultaneous equation models is non monotonic in the sample size, contrary to the well known theoretical result for static models.

CC1438: **Calculating joint bands for impulse response functions using highest density regions**
Presenter: **Anna Staszewska-Bystrova**, University of Lodz, Poland
Co-authors: Helmut Luetkepohl, Peter Winker
A new non-parametric method is proposed to construct joint confidence bands for impulse response functions of vector autoregressive models. The estimation uncertainty is captured by means of bootstrapping and the highest density region (HDR) approach is used to construct the bands. A Monte Carlo comparison of the HDR bands with existing alternatives shows that the former are less conservative than the bootstrap-based Bonferroni and Wald confidence regions. The relative tightness of the HDR bands matched with their good coverage properties makes them attractive for applications. An application to corporate bond spreads for Germany highlights the potential for empirical work.

CC1473: **Efficient deterministic filtering for state-space models**
Presenter: **Constantin Weiser**, Heinrich Heine University Duesseldorf, Germany
Co-authors: Florian Heiss
State-Space Models (SSM) have become a widely used framework for modeling time series. They are able to capture very flexible patterns inherent in econometric and financial data. The filtering procedure itself is difficult because in general they do not have an analytically tractable posterior-distribution. In the literature, approximations using stochastic methods known as particle filters have been suggested and successfully applied to a broad field of topics. They suffer from two potential problems: while the true solution can be approximated arbitrarily accurately, a close approximation can be computationally very demanding. And the approximated likelihood function is not smooth in the estimated parameters, impeding its numerical maximization. As an alternative, we suggest specific deterministic filters based on numerical integration methods to solve these problems and facilitate the use of low dimensional SSM in practice. We propose some new grid-based integration methods, the so-called sparse grids approach, in combination with an adaptive grid refinement. This algorithm allows to overcome to some extend the curse of dimensionality, which typically grid-based methods suffered. Monte Carlo simulations indeed show an impressive performance of the suggested algorithm which turns out to be more accurate than particle filters by orders of magnitude and gives a smooth approximation of the likelihood function which can be efficiently used within a maximum likelihood approach.

CC1611: **Quantile cross-spectral measures of dependence between economic variables**
Presenter: **Jozef Barunik**, UTIA AV CR vvi, Czech Republic
Co-authors: Tobias Kley
We introduce quantile cross-spectral analysis of multiple economic time series. Quantile coherence and quantile coherency measures are designed to detect general dependence structures between variables under study emerging from quantiles of the joint distribution in the frequency domain. Using a vector quantile auto-regression process as an example we illustrate how the dependence structures can arise between variables in different parts of the joint distribution across frequencies. We argue that this type of dependence is natural for many economic time series but remains unobserved when only the traditional measurements are employed. The quantile-based methodology is also very attractive as it does not require the existence of any moments. We provide a detailed analysis of the asymptotic properties of the estimators for a general class of (possibly nonlinear) processes, and discuss how inference can be conducted. In an empirical illustration we examine one of the most prominent time series in economics and document strong dependence in large negative bivariate stock market returns across various frequencies.

EI014 Room Beveridge Hall SPECIAL SESSION IN HONOR OF H. OJA'S 65TH BIRTHDAY Chair: Klaus Nordhausen

EI0171: **Monge-Kantorovich ranks and signs**
Presenter: **Marc Hallin**, Universite Libre de Bruxelles, Belgium
Unlike the real line, the real space, starting with dimension two, is not naturally ordered. As a consequence, such fundamental univariate concepts as quantile and distribution functions, ranks, and signs, do not straightforwardly extend to the multivariate context. Since no universal pre-existing order exists, each distribution, each data set, has to generate its own—the rankings behind sensible concepts of multivariate quantiles, ranks, or signs have to be distribution-specific and, in empirical situations, data-driven. The Monge-Kantorovich ranks and signs belong to that type. They are based on measure transportation ideas and an associated concept of statistical depth. They are fully distribution-free, and, under elliptical densities, asymptotically coincide with the Mahalanobis ranks and signs yielding semiparametrically efficient rank-based procedures for traditional multivariate analysis problems.

EI0556: **On invariant within equivalence coordinate system (IWECS) transformations**
Presenter: **Robert Serfling**, University of Texas at Dallas, United States
In exploratory data analysis and data mining in the very common setting of a data set $\mathbb{X}$ of vectors from $\mathbb{R}^d$, the search for important features and artifacts of a geometrical nature is a leading focus. One must insist that such discoveries be invariant under selected changes of coordinates, at least within some specified equivalence relation on geometric structures. Otherwise, interesting findings could be merely artifacts of the coordinate system. To avoid such pitfalls, it is desirable to transform the data $\mathbb{X}$ to an associated data cloud $\mathbb{X}^*$ whose geometric structure may be viewed as intrinsic to the given data $\mathbb{X}$ but also invariant in the desired sense. General treatments of such "invariant coordinate system" transformations have been developed from various perspectives. As a timely step, a more structured and unifying framework for the relevant concepts is formulated.

With this in hand, results that clarify the roles of so-called transformation-retransformation transformations are developed. These are illustrated by treating invariance properties of some outlyingness functions. Finally, some productive connections with maximal invariants are examined.

EI1214: **Nonlinear sufficient dimension reduction for functional data**
Presenter: **Bing Li**, The Pennsylvania State University, United States
Co-authors: Jun Song
We propose a general theory and the estimation procedures for nonlinear sufficient dimension reduction where the predictor or the response, or both, are random functions. The relation between the response and predictor can be arbitrary and the sets of observed time points can vary from subject to subject. The functional and nonlinear nature of the problem leads naturally to consideration of two layers of functional spaces: the first space consisting of functions of time; the second space consisting of functions defined on the first space. We take both spaces to be reproducing kernel Hilbert spaces. A particularly attractive feature of our construction is that the two functional spaces are nested, so that the kernel for the first space determines the kernel for the second. We propose two estimators, functional generalized sliced inverse regression (f-GSIR) and functional generalized sliced average variance estimator (f-GSAVE) for this general dimension reduction problem. We investigated the performances of our estimators by simulations, and applied them to data sets about phoneme recognition and handwritten symbols.

EO252 Room Woburn STATISTICAL METHODS FOR COMPLEX LONGITUDINAL DATA Chair: Yanqing Sun

EO0153: **Time-Varying coefficient models for joint modeling binary and continuous outcomes in longitudinal data**
Presenter: **Esra Kurum**, Yale University, United States
Co-authors: Runze Li, Saul Shiffman, Weixin Yao
Motivated by an empirical analysis of ecological momentary assessment data (EMA) collected in a smoking cessation study, a joint modeling technique for estimating the time-varying association between two intensively measured longitudinal responses, a continuous one and a binary one, is proposed. A major challenge in joint modeling these responses is the lack of a multivariate distribution. We suggest introducing a normal latent variable underlying the binary response and factorizing the model into two components: a marginal model for the continuous response, and a conditional model for the binary response given the continuous response. We develop a two-stage estimation procedure and establish the asymptotic normality of the resulting estimators. We also derive the standard error formulas for estimated coefficients. We conduct a Monte Carlo simulation study to assess the finite sample performance of our procedure. The proposed method is illustrated by an empirical analysis of smoking cessation data, in which the important question of interest is to investigate the association between urge to smoke, continuous response, and the status of alcohol use, the binary response, and how this association varies over time.

EO0164: **Generalized semiparametric varying-coefficient model for longitudinal data with applications to treatment switching**
Presenter: **Yanqing Sun**, University of North Carolina at Charlotte, United States
Co-authors: Li Qi, Peter Gilbert
The aim is to investigate a generalized semiparametric varying-coefficient model for longitudinal data that can flexibly model three types of covariate effects: constant effects, time-varying effects, and covariate-varying effects. Different link functions can be selected to provide a rich family of models for longitudinal data. The model assumes that the time-varying effects are unspecified functions of time and the covariate-varying effects are parametric functions of an exposure variable specified up to a finite number of unknown parameters. The estimation procedure is developed using the local linear smoothing and the profile weighted least squares estimation techniques. The asymptotic distributions of the proposed estimators are established. A working formula for bandwidth selection is discussed and examined through simulations. Our simulation study shows that the proposed methods have satisfactory finite sample performance. The proposed methods are applied to the ACTG 244 AIDS clinical trial to examine the effects of treatment switching before and after developing 215-mutation. Our analysis shows the benefit of treatment switching before developing 215-mutation while no benefit is observed for patients who switching treatments after developing 215-mutation.

EO0220: **Regression analysis of longitudinal data with informative and recurrent episode observation processes**
Presenter: **Jianguo Sun**, University of Missouri, United States
The analysis of longitudinal data with informative observation processes has recently attracted a great deal of attention and some methods have been developed. However, most of those methods treat the observation process as a recurrent event process, which assumes that one observation can immediately follow another. Sometimes, this is not the case, as there may be some delay or observation duration. Such a process is often referred to as a recurrent episode process. One example is the medical cost related to hospitalization, where each hospitalization serves as a single observation. For the problem, we present a joint analysis approach for regression analysis of both longitudinal and observation processes and a simulation study is conducted that assesses the finite sample performance of the approach. The asymptotic properties of the proposed estimates are also given and the method is applied to the medical cost data that motivated the study.

EO0797: **Survival analysis with longitudinal covariates measured with correlated errors**
Presenter: **Jianxin Pan**, The University of Manchester, United Kingdom
When covariates in Cox's proportional hazards model are time-dependent, statistical inferences may be similar to those with time-independent covariates provided that complete knowledge of the true covariates history is available. Time-dependent covariates, however, are usually measured intermittently and very likely to be measured with errors, so that joint modelling of survival and longitudinal data is much preferred. The existing methods such as sufficient statistical methods assume that longitudinal covariates are measured with mutually independent errors, which unfortunately is not always true in practices. It is evident through simulation studies that violation of the independent errors assumption can lead to very biased estimates of regression coefficients. Generalized least square estimates, rather than ordinary least square estimates, are adopted for time-dependent covariates to account for correlated measurement errors. Furthermore, covariance modelling strategy based on modified Cholesky decomposition is proposed to model the covariance structure of the measurement errors. Simulation studies show that the proposed method performs very well. Real data analysis is provided too.

EO592 Room CLO 101 ADVANCES IN FUNCTIONAL DATA ANALYSIS Chair: Jeng-Min Chiou

EO0182: **Varying-coefficient models for functional data**
Presenter: **Jane-Ling Wang**, University of California Davis, United States
Co-authors: Xiaoke Zhang
Both varying-coefficient and additive models have been studied extensively in the literature as extensions to linear models. They have also been extended to deal with functional response data. However, existing extensions are still not flexible enough to reflect the functional nature of the responses. We extend varying-coefficient and additive models to a much more flexible model and propose a simple algorithm to estimate its nonparametric additive and varying-coefficient components. We establish the asymptotic properties of each component function. We also demonstrate the applicability of the new model through analysis of traffic data.

EO0979: **Nonparametric scalar-on-function regression via regularized principal coordinates**
Presenter: **Philip Reiss**, New York University, United States
Co-authors: David Miller, Wen-Yu Hua

Three major approaches to nonparametric regression, in order of their historical development, are (i) kernel smoothing, (ii) reproducing kernel Hilbert space (RKHS) optimization solved by smoothing splines, and (iii) reduced-rank penalized splines. Functional-predictor analogues of the first two approaches, respectively, are the Nadaraya-Watson estimator and the RKHS estimator. A simple new nonparametric functional regression approach of type (iii) is introduced. The core idea is to choose a distance or kernel defined on pairs of functions, and to perform penalized regression on principal coordinates (in the former case) or kernel principal components (in the latter). Extensions to generalized linear models, multiple predictors and random effects are straightforward using existing software. An application to online signature verification illustrates the method's utility.

EO1194: **Intrinsic feature extraction via functional data analysis**
Presenter: **Yuko Araki**, Shizuoka University, Japan

Recent years have seen that functional data analysis are capable of extracting intrinsic features from recently arising complicated and high dimensional data, such as three dimensional brain sMRI, time course microarray data , or hundreds of records of human gait, for example. We introduce statistical methods for classifying individuals with such high dimensional covariates. We have also constructed a statistical model to characterize the relationship between the event time and a set of baseline covariates. The proposed method is based on composite basis function, which is an extended version of basis expansions with the help of sparse PCA. Further, L_1-type penalty constraints are imposed in the estimation of the parameters of logistic discrimination and Cox proportional hazard models. This two-step regularization method accomplishes both covariates selection and estimation of unknown model parameters simultaneously. The crucial issue is how to select the regularization parameters used in model estimation. We propose to use information criterion based model selection. The proposed models are applied to real data example and Monte Carlo simulations are conducted to examine the efficiency of our modeling strategies.

EO1818: **Functional principal component analysis over two-dimensional manifolds**
Presenter: **Eardi Lila**, University of Cambridge, United Kingdom
Co-authors: Laura Sangalli, John Aston

The human cerebral cortex is a folded layer of gray matter. This enables the brain to have a large surface area. It is now well established that it is beneficial to analyse cortical neuroimaging data with surface-constrained methods and it is natural to assume some spatial smoothness hypothesis on the signal measured over the cortical surface. Thus this data can be contextualized in the Functional Data framework. Moreover, in a multi-subject analysis, one such function will be available for each individual. Functional Principal Components is a powerful dimensionality reduction tool that allows the study of the main modes of variation of functional data. However, the classical literature can deal only with functions whose domain is an interval of the real line. We introduce a novel formulation that generalizes FPCA concepts to functions lying on two-dimensional manifolds. We deal with the problem of introducing an appropriate smoothing operator on the manifold and we approach the discretization task by means of finite element analysis. We compare the proposed algorithm with other approaches already proposed in literature. We finally apply it to neuroimaging data, proving that thanks to its efficiency, it is applicable to large-scale datasets.

EO124 Room Jessel TIME SERIES: NONPARAMETRICS AND EXTREMES Chair: Edit Gombay

EO0241: **Testing for independence between two functional time series**
Presenter: **Gregory Rice**, University of Waterloo, Canada
Co-authors: Lajos Horvath

Frequently econometricians are interested in verifying a relationship between two or more time series. Such analysis is typically carried out by causality and/or independence tests which have been well studied when the data is univariate or multivariate. Modern data though is increasingly of a high dimensional or functional nature for which finite dimensional methods are not suitable. We develop methodology to check the assumption that data obtained from two functional time series are independent. Our procedure is based on the norms of empirical cross covariance operators and is asymptotically validated when the underlying populations are assumed to be in a class of weakly dependent random functions which include the functional ARMA, ARCH and GARCH processes.

EO0510: **Robust and nonparametric detection of shifts using two-sample U-statistics and U-quantiles**
Presenter: **Roland Fried**, TU Dortmund University, Germany
Co-authors: Herold Dehling, Martin Wendler

Tests for detecting level shifts in near epoch dependent time series are studied. As examples, we will be able to treat most standard models of time series analysis, such as ARMA and GARCH processes. The popular CUSUM test is not robust to outliers and can be improved in case of non-normal data, particularly for heavy-tails. The CUSUM test can be modified using the Hodges-Lehmann 2-sample estimator, which is the median of all pairwise differences between the samples. It is highly robust and has a high efficiency under normality. Like for a related test based on the 2-sample Wilcoxon statistic, the asymptotics of the Hodges-Lehmann change-point test can be established under general conditions without any moment assumptions. Both tests offer similar power against shifts in the center of the data, but the test based on the Hodges-Lehmann estimator performs superior if a shift occurs far from the center. MOSUM-type tests restrict attention to data in two subsequent moving time windows. This may overcome possible masking effects due to several shifts into different directions. We investigate MOSUM-type tests based on the 2-sample Wilcoxon statistic or the Hodges-Lehmann estimator in case of independent data. While asymptotical versions of the tests turn out to be rather conservative if the sample size is not very large, versions deriving critical values using the permutation principle indeed provide improvements in the presence of multiple shifts.

EO0926: **Modeling time series of house insurance claims due to extreme weather and climate events**
Presenter: **Natalia Humphreys**, The University of Texas at Dallas, United States

There is a growing body of scientific evidence that extreme weather events are increasing in frequency and intensity. This phenomenon has already led to a significant upward trajectory in the dynamics of weather-related house insurance claims due to floods, storms, hurricanes, and other natural disasters. To mitigate the adverse effects of the changing climate, insurers need a comprehensive attribution analysis of weather-related claim dynamics at a local level and a reliable quantification of future risks. We propose a nonparametric approach to modelling a nonlinear relationship between severity of weather-related losses and atmospheric data, and detection of critical thresholds in weather variables leading to an increased magnitude of such claims. We develop a novel data-driven composite index that enables to assess the current weather-related claim dynamics and to forecast future short- and long-term trajectories of losses at a local level. We illustrate our methodology by application to modelling current and projected house insurance losses due to floods, heavy rain, hail and severe wind. The study will result in the better understanding of factors impacting claims dynamics and the better detection of local geographical areas of increased vulnerability due to projected climate change.

EO1054: **Confidence intervals for time series using the Wilcoxon signed rank statistic with drug policy intervention applications**
Presenter: **Ying Zhang**, Acadia University, Canada

We are interested in rank-based statistics to make inference for time series data. We propose a new approach for constructing robust confidence intervals for the mean/median in time series data using the Wilcoxon Signed Rank Statistic or Sign Statistic. The dependent nature of the data in time series analysis poses difficulties in constructing robust confidence intervals. The key to such construction lies in the appropriate estimation of the variance of the statistics considered. We discuss and compare three variance estimators: the bootstrap variances and the partial sum variance of dependent process. Finally we demonstrate our approach to an investigation of the effect of a provincial insurance policy change on the consumption of a heartburn medication.

EO310 Room Torrington MODERN STATISTICAL REGRESSION Chair: Roy Welsch

EO0291: **A new perspective on boosting in linear regression via subgradient optimization and relatives**
Presenter: **Robert Freund**, MIT, United States
Co-authors: Paul Grigas, Rahul Mazumder

The aim is to analyze boosting algorithms in linear regression from the perspective modern first-order methods in convex optimization. We show that classical boosting algorithms in linear regression, namely the incremental forward stagewise algorithm (FSe) and least squares boosting (LS-Boost-e), can be viewed as subgradient descent to minimize the loss function defined as the maximum absolute correlation between the features and residuals. We also propose a modification of FSe that yields an algorithm for the LASSO, and that may be easily extended to an algorithm that computes the LASSO path. These new algorithms for the LASSO may also be interpreted as the same master algorithm (subgradient descent), applied to a regularized version of the maximum absolute correlation loss function. We derive novel, comprehensive computational guarantees for several boosting algorithms in linear regression (including LS-Boost-e and FSe), which inform us about the statistical properties of boosting algorithms. In particular they provide, for the first time, a precise theoretical description of the amount of data-fidelity and regularization imparted by running a boosting algorithm with a pre-specified learning rate for a fixed but arbitrary number of iterations, for any dataset.

EO0371: **The role of robustness in modern statistical regression**
Presenter: **Martin Copenhaver**, MIT, United States
Co-authors: Dimitris Bertsimas

Sparsity is a key driver in modern statistical problems, from linear regression via the Lasso to matrix regression with nuclear norm penalties in matrix completion and beyond. In stark contrast to sparsity motivations for such problems, it is known in the field of robust optimization that a variety of vector regression problems, such as Lasso which appears as a loss function plus a regularization penalty, can arise by simply immunizing a nominal problem (with only a loss function) to uncertainty in the data. Such a robustification offers an explanation for why some linear regression methods perform well in the face of noise, even when these methods do not reliably produce sparse solutions. We deepen and extend the understanding of the connection between robustification and regularization in regression problems. Specifically, (a) in the context of linear regression, we characterize under which conditions on the model of uncertainty used and on the loss function penalties robustification and regularization are equivalent; (b) we show how to tractably robustify median regression problems; and (c) we extend the characterization of robustification and regularization to matrix regression problems (matrix completion and Principal Component Analysis).

EO0531: **Exact subset selection in regression**
Presenter: **Rahul Mazumder**, MIT, United States
Co-authors: Dimitris Bertsimas

The usual linear regression set up is considered where the task is to find the best least squares fit to data subject to a cardinality constraint in the regression coefficient vector. This problem, popularly dubbed as the best-subset selection problem, is known to be NP-hard and has been widely dismissed as computationally intractable. A novel framework is proposed via which the best-subset selection problem can indeed be solved to global (or near global) accuracy in problems of practical interest. At the core of our proposal is a computationally tractable framework that brings to bear the power of modern discrete optimization methods:(a) Mixed Integer Optimization (MIO) techniques and (b) Discrete first-order methods, which may be viewed as a "discrete" extension of first order continuous optimization methods. Superiority of best-subset selection procedures over popularly used computationally friendlier alternatives will be discussed. We describe how to address change point detection problems, and another classic but computationally difficult procedure in robust statistics, namely, the Least Median of Squares Regression.

EO0971: **Recent advances in high-dimensional robust regression**
Presenter: **Po-Ling Loh**, University of Pennsylvania, United States

We present results for high-dimensional linear regression using robust $M-$estimators with a regularization term. We show that when the derivative of the loss function is bounded, our estimators are robust with respect to heavy-tailed noise distributions and outliers in the response variables, with the usual order of $k \log p/n$ rates for high-dimensional statistical estimation. Our results continue a line of recent work concerning local optima of nonconvex $M-$estimators with possibly nonconvex penalties, where we adapt the theory to settings where the loss function only satisfies a form of restricted strong convexity within a local neighborhood. We also discuss second-order results concerning the asymptotic normality of our estimators, and provide a two-step $M-$estimation algorithm for obtaining statistically efficient solutions within the local region.

EO154 Room Bedford METHODS AND COMPUTATIONS FOR DEPENDENCE MODELLING Chair: Ivan Kojadinovic

EO0298: **Parallel and other simulations in R made easy: An end-to-end study**
Presenter: **Marius Hofert**, University of Waterloo, Canada
Co-authors: Martin Maechler

The world of copulas is highly non-linear. Many applications require simulation studies involving copulas. From a computational point of view, these studies can become quite demanding. In order to tackle such large-scale simulations, the Rpackage "simsalapar" has been developed. This package aims at simplifying statistical simulation studies and carefully deals with important task such as parallel computing, seeding, catching of warnings and errors, and measuring runtime. The approaches in "simsalapar" may be of interest to students, researchers and practitioners as a how-to for conducting realistic, large-scale simulation studies in R. A practical (copula) problem from the realm of Quantitative Risk Management serves as a motivating example.

EO0488: **Modeling bivariate extreme value copula with shape restricted splines**
Presenter: **Jun Yan**, University of Connecticut, United States
Co-authors: Danielle Prado, Devanil Souza

A bivariate extreme value copula is characterized by its Pickands dependence function, which is convex and restricted to a certain range. Two shape restricted spline regressions are proposed to estimate the Pickands dependence function. The first one uses the Bernstein polynomials, while the second one uses cubic splines, both with the appropriate restrictions imposed to ensure the shape of the estimated Pickands dependence function. The degrees of freedom is selected by the Shwartz information criterion. The methods are easy to implement and fast to compute.

The semiparametric estimation methods naturally lead to a hypothesis test for extreme value dependence. The performance of the inferences is competitive in a simulation study.

EO0533: **Detecting breaks in the dependence of multivariate extreme-value distributions**
Presenter: **Paul Kinsvater**, TU Dortmund University, Germany
Co-authors: Axel Buecher, Ivan Kojadinovic
In environmental sciences, it is often of interest to assess whether the dependence between extreme measurements has changed during the observation period. The aim is to propose a statistical test that is particularly sensitive to such changes. The resulting procedure is also extended to allow the detection of changes in the extreme-value dependence under the presence of known breaks in the marginal distributions. Simulations are carried out to study the finite-sample behavior of both versions of the proposed test. Illustrations on hydrological data sets conclude the work.

EO0794: **A local C-vine copula based composite likelihood approach to spatial dependency modeling**
Presenter: **Tobias Erhardt**, Technische Universitaet Muenchen, Germany
Co-authors: Claudia Czado, Ulf Schepsmeier
Classical approaches for spatial dependence modeling usually assume Gaussian dependence structures. This assumption may simplify the modeling process, however it is not always met, when we face real world problems. We introduce a composite likelihood based inference technique combining local C-Vine copulas, which is able to capture non-Gaussian spatial dependencies. Vine copulas are a class of flexible d-dimensional probability distributions, which are composed out of pair-copulas. For each of these bivariate building blocks we can choose among a variety of different dependence structures (copula families), which are well understood and easy to compute. The proposed composite likelihood approach for the modeling of spatial dependencies combines the flexibility of vine copulas with the geostatistical idea of modeling the extent of spatial dependencies using the distances between the variable locations. We outline the composite likelihood based parameter estimation, a method for spatial prediction and provide a data based illustration of the new methodology.

EO633 Room Court BAYESIAN SEMI- AND NONPARAMETRIC MODELLING II Chair: Matteo Ruggiero

EO0422: **Modeling scale-free networks**
Presenter: **Harry Crane**, Rutgers University, United States
In general applications, statistical models strike a balance between empirical observations and the logical properties needed for valid inferences. These considerations create issues when modeling networks, as empirical observations of sparsity and power law degree distribution are incompatible with the usual assumption of exchangeability. After reviewing previous attempts to circumvent this issue, we discuss a new framework within which certain scale-free networks can be modeled in an exchangeable way.

EO0505: **Nonparametric hierarchical models based on completely random measures**
Presenter: **Federico Camerlenghi**, University of Pavia and Collegio Carlo Alberto, Italy
Co-authors: Antonio Lijoi, Igor Pruenster
A very active line of research in Bayesian statistics has aimed at defining and investigating general classes of nonparametric priors. A notable example, which includes the Dirichlet process, is obtained through normalization or transformation of completely random measures. These have been extensively studied for the exchangeable setting. However in a large variety of applied problems data are heterogeneous, being generated by different, though related, experiments; in such situations partial exchangeability is a more appropriate assumption. In this spirit we propose a nonparametric hierarchical model based on transformations of completely random measures, which extends the hierarchical Dirichlet process. The model allows us to handle related groups of observations, creating a borrowing of strength between them. From the theoretical viewpoint, we analyze the induced partition structure, which plays a pivotal role in a very large number of inferential problems. The resulting partition probability function has a feasible expression, suitable to address predictionin its generality, as suggested by de Finetti. Finally we propose a set of applications which include inference on genomic and survival data.

EO0602: **Model-based clustering via normalized completely random measure mixtures**
Presenter: **Raffaele Argiento**, University of Kent, United Kingdom
Co-authors: Alessandra Guglielmi, Ilaria Bianchini
Mixtures of parametric densities arise as the natural statistical tool when dealing with model-based clustering problems, and a very flexible class of such mixtures is obtained when the mixing measure is a random probability measure, possibly a.s. discrete. We consider mixtures when the mixing probability measure is within a large class of random probability measures, that is normalized completely random measures (NCRM). However, the computational effort to compute the relevant posteriors in this case can be very burdensome, since MCMC schemes are complicated by the presence of an infinite number of parameters. We propose a truncation method to approximate the mixing measure and simplify the computations. Since a NCRM is a random discrete measure where the weights are obtained by normalization of the points of a Poisson process, we discard those larger than a threshold epsilon. Hence, the number of parameters becomes finite, so that an efficient blocked Gibbs sampler to simulate from the posterior is built. To show the performance of our algorithm and the flexibility of the model, we will illustrate two example via NCRM mixtures: the first consider a new NCRM called Bessel random probability measure as the mixing measure, and then apply the mixture to simulated and a real dataset, while the second deals with a linear dependent epsilon-NGG mixture to fit a well known dataset.

EO1007: **Bayesian adaptive quantile estimation in deconvolution**
Presenter: **Catia Scricciolo**, Bocconi University, Italy
Bayesian adaptive quantile estimation in deconvolution problems with unknown error distribution is considered: the objective is to estimate single quantiles of a distribution or the quantile function globally from indirect observations that are additively corrupted by unknown error measurements. Quantile estimation in deconvolution is an instance of nonlinear functional estimation in ill-posed inverse problems. We pursue the analysis for mildly ill-posed problems, namely, when the error distribution has characteristic function that decays polynomially. We propose a method to derive posterior contraction rates for the inverse problem from rates for the direct problem and establish adaptive pointwise and L_2-rates that coincide with the minimax rates. Also, credibles sets are constructed whose frequentist coverage is studied.

EO613 Room Senate INFERENCE IN LATENT VARIABLE MODELS Chair: Estelle Kuhn

EO0430: **Statistical clustering of temporal networks through a dynamic stochastic block model**
Presenter: **Catherine Matias**, CNRS - Universite Pierre et Marie Curie, France
Co-authors: Vincent Miele
Statistical node clustering in discrete time dynamic networks is an emerging field that raises many challenges. We explore statistical properties and deterministic inference in a model that combines a stochastic block model (SBM) for its static part with independent Markov chains for the evolution of the nodes groups through time. We model binary data as well as weighted dynamic random graphs (with discrete or continuous edges values). Our approach particularly focuses on the control for label switching issues across the different time steps. We study identifiability of the model parameters, propose an inference procedure based on a variational expectation maximization algorithm as well as a model selection criterion to select for the number of groups. We carefully discuss our initialization strategy which plays an important role in the method and compare our procedure with existing ones on synthetic datasets. We also illustrate our approach on a real data set of encounters among high school students and provide an implementation of the method into a R package called dynsbm.

EO0773: **Parametric estimation of complex mixed models based on meta-model approach**
Presenter: **Pierre Barbillon**, AgroParisTech, France
Co-authors: Celia Barthelemy, Adeline Leclercq-Samson
Complex biological processes are more and more experimented along time among a collection of individuals. Longitudinal data are then available and the statistical challenge is to analyse these data to better understand the underlying biological mechanisms. The parametric statistical approach commonly used with longitudinal data is mixed-effects model methodology. In these models, the regression functions are now highly-developed to describe precisely biological processes. They may be solutions of multi-dimensional differential equations. When there is no analytical solution, a classical approach for estimating the parameters relies on the coupling of a stochastic version of the EM algorithm (SAEM) with a MCMC algorithm. This procedure needs many evaluations of the regression function which is clearly prohibitive when a time-consuming solver is used for computing it. That is why we propose to replace this regression function with a meta-model based on a Gaussian process. The new source of uncertainty due to this approximation can be incorporated in the model which leads to what we call a mixed meta-model. We guarantee a control on the distance between the maximum likelihood estimates in this mixed meta-model and the maximum likelihood estimates obtained with the exact mixed model.

EO0425: **Inference for dynamic and latent variable models via iterated, perturbed Bayes maps**
Presenter: **Edward Ionides**, University of Michigan, United States
Iterated filtering algorithms are stochastic optimization procedures for latent variable models that recursively combine parameter perturbations with latent variable reconstruction. Previously, theoretical support for these algorithms has been based on the use of conditional moments of perturbed parameters to approximate derivatives of the log likelihood function. We introduce a new theoretical approach based on the convergence of an iterated Bayes map. A new algorithm supported by this theory displays substantial numerical improvement on the computational challenge of inferring parameters of a partially observed Markov process.

EO0729: **Approximate Bayesian model choice via random forests**
Presenter: **Pierre Pudlo**, CNRS Universite de Montpellier, France
When the likelihood depends on some latent process living in a space of high dimension, standard inference methods derived from the EM algorithm or MCMC on an augmented space are inefficient. The approximate Bayesian computation (ABC) method bypasses the computation of the likelihood by comparing the observed data with numerous simulated datasets. ABC has thus found an ever increasing range of applications covering complex models of population genetics, where the latent process explains the past genetic evolution of individuals sampled nowadays. We show how a random forest classifier can help ABC to compare the simulations to the observed dataset and to perform a reliable model choice. We also propose a solution to evaluate the posterior probability of the chosen model with a second resort to the random forest algorithm.

EO244 Room CLO 203 SEMI- AND NON-PARAMETRIC ESTIMATION OF FRONTIERS Chair: Leopold Simar

EO0442: **Cross-section dependence and latent heterogeneity to evaluate the impact of human capital: A robust nonparametric model**
Presenter: **Camilla Mastromarco**, University of Salento - Lecce, Italy
Co-authors: Leopold Simar
Human Capital has been recognized as the most important force behind economic growth of countries. However, the effect of this important growth factor on economic growth remains ambiguous due to endogeneity and latent heterogeneity. By using a dataset of 40 countries over 1970-2007, we estimate the global frontier and explore the channels under which human capital and time affect the production process and its components: impact on the attainable production set (input-output space), and the impact on the distribution of efficiencies. We extend existing methodological tools - robust frontier in non parametric location-scale models - to examine these interrelationships. We use a flexible nonparametric two-step approach on conditional efficiencies to eliminate the dependence of production inputs/outputs on common factors. We emphasize the usefulness of pre-whitened inputs/outputs to obtain more reliable measure of productivity and efficiency to better investigate the impact of human capital on the catching-up productivity process. Then, we take into account the problem of unobserved heterogeneity and endogeneity in the analysis of the influence of human capital on the production process by extending a previous instrumental nonparametric approach to account also for cross section and time dependence.

EO0614: **Confidence intervals for efficiency scores in non-convex technologies**
Presenter: **Luiza Badin**, Bucharest University of Economic Studies, Romania
Co-authors: Valentin Patilea, Leopold Simar
The most popular envelopment estimators in efficiency analysis rely on the assumption that all the observations fall on the same side of the frontier. The nonparametric Free Disposal Hull (FDH) estimator represents the smallest free disposal set covering all the observations and the technical efficiency of an arbitrary producer is measured with respect to the boundary of the free disposal hull of the whole sample. The asymptotic sampling distribution of the FDH estimator is the Weibull law depending on the dimension of the input $\times$ output space and a parameter $\mu_{NW,0}$, which has to be estimated. The literature on frontier estimation proposes a consistent estimator for $\mu_{NW,0}$, but Monte Carlo experiments available so far show rather poor performances of this estimator in terms of accuracy. These poor performances jeopardize the precision of the resulting confidence intervals for the efficiency scores. The objective is to provide alternative estimators of $\mu_{NW,0}$ and asymptotic confidence intervals for the efficiency scores, easy to implement and fast to compute in practical situations. The behavior of the resulting estimators and their effect on the achieved coverage of the confidence intervals for the efficiency scores are investigated and compared through a Monte Carlo simulation study.

EO0627: **Nonparametric robust stochastic frontier analysis: A Tikhonov regularization approach**
Presenter: **Jean-Pierre Florens**, Toulouse School of Economics, France
Co-authors: Abdelaati Daouia, Leopold Simar

In production theory and efficiency analysis, the interest is in estimating frontiers of the production set, i.e. the set of attainable combinations of inputs (the resources or production factors) and of outputs (the production: goods or services produced. We suggest an original and new approach to estimate non-parametrically and in a robust way stochastic frontier functions, i.e. frontier models where stochastic noise is allowed. We suppose that the noise has a given density (like the Gaussian) for identifiability. We suppose, as in most studies on deconvolution, that the variance of the noise is known. We consider the estimation of an minimum input (or cost) function reachable for given values of the outputs. The idea is to use deconvolution methods to estimate in a first step and for each value of the outputs, the conditional survivor function, as in previous deterministic models where no noise was allowed. This provides an estimate of the order-m frontier by integrating powers of the estimated survivor function. The asymptotic of the resulting estimators is derived and under some regularity condition, we reach the $\sqrt{n}$ rate of convergence and asymptotic normality. The procedure is based on Tikhonov regularization which is easy and fast to implement. It is illustrated with simulated and real data examples.

EO0732: **Testing the separability condition in two-stage nonparametric models of production**
Presenter: **Cinzia Daraio**, University of Rome La Sapienza, Italy
Co-authors: Leopold Simar, Paul Wilson

A statistical model was previously provided that can rationalize two-stage estimation of technical efficiency in non-parametric settings. Two-stage estimation has been widely used, but requires a strong separability' assumption: the second-stage environmental variables cannot affect the support of the input and output variables in the first stage. We provide a fully non-parametric test of this assumption, deriving test statistic having asymptotic normal distribution under the null. We obtain this for both the DEA and the FDH first stage estimators. Our simulation results indicate that our tests (including some bootstrap alternatives) perform well both in terms of size and power. We present a real-world empirical example by updating a previous analysis on U.S. commercial banks. Our tests easily reject the assumption required for two-stage estimation, calling into question results that appear in hundreds of papers that have been published in recent years.

EO086 Room G21A STATISTICAL ANALYSIS OF LONG-RANGE DEPENDENT DATA Chair: Herold Dehling

EC0470: **Some limit theorems for linear processes in a Hilbert space**
Presenter: **Alfredas Rackauskas**, Vilnius University, Lithuania

Consider for each $n \geq 1$ an array $\{X_{nk} : k = 1,\ldots,n\}$ of linear processes with values in a separable Hilbert space H and defined by $X_{nk} = \sum_{j=0}^{\infty} u_{nj}\varepsilon_{k-j}$, $k = 1,\ldots,n$, where $\{u_{nj} : j \geq 0\}$ are bounded linear operators and $\{\varepsilon_k : k \geq 1\}$ are independent and identically distributed H-valued random elements with zero mean and finite second moment. We discuss some limit theorems for sums $\sum_{j=1}^{n} X_{nj}, n \geq 1$ where the series of operator norms $\sum_{j=0}^{\infty} \|u_{nj}\|$ converge but the sum tends to infinity as $n \to \infty$. Applications in statistical inference are discussed too.

EO1387: **Weighted tests for distributional change in long-memory processes**
Presenter: **Johannes Tewes**, Ruhr-Universitaet Bochum, Germany

We study changes in the marginal distribution of a subordinated Gaussian process that exhibits long-range dependence. Tests for such change-points are occasionally based on Kolmogorov-Smirnov or Cramer-von Mises statistics. However, our simulation study indicates that the combination of strong dependence and change-points occurring near the beginning or near the end of the observation period drastically reduces the power of these tests. We therefore consider weighted versions of the statistics. In order to calculate critical values we show that the weighted sequential empirical process converges in distribution to a semi-degenerate process. The structure of the limiting process, e.g. the marginal distribution and the covariance structure, mainly depends on the Hemite rank. It is the index of the first nonzero coefficient in the Hermite expansion of one summand of the empirical process. A special feature of distributional changes is the fact that the Hermite rank may change, too. We consider local alternatives covering this scenario, and as a result, we may derive the asymptotic power of the change-point tests.

EO1389: **Quasi-MLE for quadratic ARCH model with long memory**
Presenter: **Andrius Skarnulis**, Vilnius University, Lithuania
Co-authors: Donatas Surgailis, Ieva Grublyte

The quasi maximum likelihood estimation for parametric quadratic ARCH model with long memory and the conditional variance σ_t^2 given by strictly positive quadratic form of an observable stationary sequence $r_s, s < t$, that was introduced recently and called the generalized quadratic ARCH model (GQARCH), is discussed. The consistency and asymptotic normality of the corresponding QMLE estimates, including the estimate of long memory parameter $0 < d < 1/2$, is proved. The presence of positive constant that separates the volatility from zero is very important for consistency of QMLE procedure. Finite sample performance of the QMLE for the GQARCH model is investigated in the simulation study, the empirical RMSE is presented. Some results from applications of GQARCH model on empirical stock market data are included.

EO1647: **The multivariate sequential empirical process under long-range dependence**
Presenter: **Jannis Buchsteiner**, Ruhr-University Bochum, Germany

An important object in nonparametric statistics is the sequential empirical process (SEP) $(R_N(x,t))$. It is well known that under a suitable long-range dependence setting, $(R_N(x,t))$ converges weakly to the product of a deterministic function and an Hermite process. We will discuss a number of new noncentral limit theorems for the SEP and their statistical application. Specifically, we study the SEP for multivariate subordinated Gaussian data. Furthermore, we investigate the SEP indexed by a class of functions, satisfying a bracketing condition.

EO120 Room Chancellor's Hall ON QUANTILES, EXPECTILES AND EXTREMILES Chair: Irene Gijbels

EO0593: **From quantiles and expectiles to extremiles**
Presenter: **Abdelaati Daouia**, Toulouse School of Economics, France
Co-authors: Irene Gijbels

Quantiles and expectiles of a distribution are found to be useful descriptors of its tail in the same way as the median and mean are related to its central behavior. A valuable alternative class to expectiles, called extremiles, is considered, which parallels the class of quantiles and includes the family of expected minima and expected maxima. The new class is motivated via several angles, which reveals its specific merits and strengths. Extremiles suggest better capability of fitting both location and spread in data points and provide an appropriate theory that better displays the interesting features of long-tailed distributions. We discuss their estimation in the range of the data and beyond the sample maximum. Implications for exploring trends and spread in regression analysis are also considered in some detail. A number of unrelated motivating examples are given to illustrate the utility of estimated extremiles in modeling noncentral behavior. There is in particular an interesting connection with coherent measures of risk protection.

EO0756: **Replacing moments by expectiles**
Presenter: **Paul Eilers**, Erasmus University Medical Centre, Netherlands

A number of statistical procedures compute moments from data and distributions from moments. For detailed results many moments might be needed, but it is well known that higher moments are very sensitive to small changes in the data, which is a disadvantage. As an alternative expectiles are proposed to be used. They are easy to compute, using asymmetric least squares, and one can construct a smooth density from a limited set of expectiles. In contrast to quantiles, expectiles are unique and also well-defined for discrete distributions. Applications are discussed, with an emphasis on extrapolation and forecasting of distributions.

EO0929: **A local moment type estimator for an extreme quantile in regression with random covariates**
Presenter: **Armelle Guillou**, Strasbourg, France
Co-authors: Yuri Goegebeur, Michael Osmann

A conditional extreme quantile estimator is proposed in the presence of random covariates. It is based on an adaptation of the moment estimator in the classical univariate setting and thus it is valid in the domain of attraction of the extreme value distribution, i.e. whatever the sign of the extreme value index is. Asymptotic normality of the estimator is established under suitable assumptions, and its finite sample behaviour is evaluated with a small simulation study, where a comparison with an alternative estimator already proposed in the literature is provided. An illustration to a real dataset concerning the world catalogue of earthquake magnitudes is also proposed.

EO1035: **On expectiles and quantiles**
Presenter: **Fabian Sobotka**, University Oldenburg, Germany
Co-authors: Goeran Kauermann, Thomas Kneib, Linda Schulze Waltrup

While quantile regression can be seen as a generalization of median regression, expectiles are a generalized form of the mean. Quantile regression estimates are acquired by minimizing the asymmetrically weighted sum of the absolute residuals while an expectile is computed from the least asymmetrically weighted squares (LAWS) of the residuals. This means, instead of a linear programming problem, we face an easier optimization problem with an objective function being convex and differentiable. Despite this ease in computation, expectiles lack the intuitive interpretation of quantiles. We contrast the two approaches and show how to get quantiles from a fine grid of expectiles. This has the additional advantage that we can directly compare quantiles and expectiles. We compare such quantiles from expectiles with native quantile estimates regarding efficiency on an asymptotic and an empirical basis. We also look at regression problems where both, quantile and expectile curves, have the undesirable property that individually estimated neighboring curves may cross each other. We show empirically that crossing curves are more frequent in quantile regression and we propose a method to estimate non-crossing expectile curves based on splines. In an application, we look at the expected shortfall, a risk measure used in finance, which can be calculated easily with the proposed methods.

EO188 Room Gordon COMPUTATIONAL ASPECTS OF ESTIMATING STOCHASTIC PROCESSES Chair: Kengo Kamatani

EO0704: **Quasi likelihood analysis for ultra high frequency data**
Presenter: **Nakahiro Yoshida**, University of Tokyo, Japan

The latest trend of financial statistics is toward analysis of ultra high frequency phenomena by modeling more precise mechanisms in a more precise time-scale. There is no Brownian motion as a driving process of the system since the central limit theorem is not effective at this level of description, differently from the standard framework. Point process modeling gives a promising approach to a description of microstructure. The quasi likelihood analysis (QLA) is a systematic analysis of the quasi likelihood random field and the associated estimators, with a large deviation method that derives precise tail probability estimates for the random field and estimators. QLA is constructed for point processes. The point process regression model can express asynchronicity of observations, lead-lag effects and microstructure. This model can incorporate nonstationarity under finite time horizon and self-exciting/self-correcting effects of the point processes as well as exogenous effects. Non-ergodic statistics is obtained when the intensities of the point processes become large. The point process regression model is applied to price models and limit order books. A related topic is a nonparametric method for estimating covariation between intensity processes. QLA can be developed also in long-time asysmptotics. Then establishing ergodicity of point processes becomes an issue.

EO0984: **On some ergodic properties of the MpCN algorithm**
Presenter: **Kengo Kamatani**, Osaka University, Japan

Ergodic properties of Mixed preconditioned Crank-Nicolson (MpCN) algorithm are studied. The MpCN algorithm is a Markov chain Monte Carlo which is similar to the Random walk Metropolis (RWM) algorithm. Unlike RWM, the proposal transition kernel is reversible with respect to a heavy-tailed distribution. The MpCN algorithm works well for complicated target distributions, and is proved to have good asymptotic properties in high-dimension. We study ergodic properties of the MpCN algorithm. It is proved that even for heavy-tailed distribution, the Markov chain generated by MpCN algorithm can be geometrically ergodic. We compare some MCMC algorithm by ergodic properties. Application to the parameter estimation of stochastic diffusion process is also studied.

EO0981: **Maximum-likelihood-type estimation for diffusion processes with noisy, nonsynchronous observations**
Presenter: **Teppei Ogihara**, The Institute of Statistical Mathematics, Japan

We study statistical inference for security prices modeled by diffusion processes with high-frequency observations. In particular, we focus on two specific problems on analysis of high-frequency data, that is, nonsynchronous observations and the presence of observation noise called market microstructure noise. We construct a maximum-likelihood-type estimator of parameters, and study their asymptotic mixed normality. We also study the local asymptotic mixed normality property and asymptotic efficiency of our estimator when diffusion coefficients are constants and observation noise is normal.

EO0839: **Hybrid type estimation for stochastic differential equations based on sampled data**
Presenter: **Masayuki Uchida**, Osaka University CSFI CREST JST, Japan

We consider the estimation procedure for unknown parameters of a multi-dimensional diffusion type process defined by a stochastic differential equation from discrete observations. We propose the hybrid type estimators, which are obtained by the following procedure. First, as initial estimators, we get the Bayes type estimators with the non-optimal rates of convergence under weak conditions. Next, using the initial Bayes type estimators, we obtain hybrid type estimators by means of the adaptive estimation or the method of scoring with the quasi-likelihood function based on locally Gaussian approximation of the transition density function of the diffusion type process. We notice that the number of iteration to obtain the hybrid type estimator strongly depends on the asymptotic properties of initial estimators. It is shown that the proposed estimators have asymptotic properties including the convergence of moments, and we also give an example and simulation results of the hybrid type estimator with the initial Bayes type estimator.

EO178 Room Montague ROBUST STATISTICAL MODELLING Chair: Alfio Marazzi

EO0740: **Higher-order adjustments of the signed scoring rule root statistic**
Presenter: **Laura Ventura**, University of Padova, Italy
Co-authors: Valentina Mameli, Monica Musio

Proper scoring rules can be used as an alternative to the full likelihood, when the aim is to increase the robustness. Proper scoring rule inference is usually based on the first-order approximations to the distribution of the scoring rule estimator or of the scoring rule ratio test statistic. However, several examples illustrate the inaccuracy of first-order methods, even in models with a scalar parameter, when the sample size is small or moderate. Analytical higher-order asymptotic expansions for proper scoring rules, generalizing results for likelihood quantities but allowing for the failure of the information identity, have been previously discussed. However, the calculation of the quantities involved in the analytical adjustments of the signed and signed profile scoring rule root statistic is cumbersome, even for simple models. The aim is to discuss the alternative approach to higher-order adjustments, based on a parametric bootstrap.

EO0807: **Robust state space models**
Presenter: **Eva Cantoni**, University of Geneva, Switzerland
Co-authors: William Aeberhard, Chris Field, Joanna Mills Flemming, Ximing Xu

In fishery science, state space models are often used because they explicitly include both process error in the dynamics and observation error in the data. The estimation of the parameters of the model is usually performed by maximum (marginal) likelihood. But the use of robust statistical methodologies is essential because the data being modelled in fisheries science and management may be subject to large measurement error. In addition, the state and observation processes are only approximations with the consequence that there may well be outlying observations or deviating substructure. We propose a way to robustify the estimation of the parameters of a state space model, which modifies the objective function to introduce robustness. The evaluation of the new marginal objective function being computationally challenging, we use software implementing automatic differentiation to reach our goal. We show simulation results supporting our proposal and applications to areal dataset.

EO1100: **Robust estimation for mixtures of Gaussian factor analyzers**
Presenter: **Alfonso Gordaliza**, Universidad de Valladolid, Spain
Co-authors: Luis Angel Garcia-Escudero, Francesca Greselin, Salvatore Ingrassia, Agustin Mayo-Iscar

Mixtures of Gaussian factors are powerful tools for modeling an unobserved heterogeneous population, offering at the same time dimension reduction and model-based clustering. Unfortunately, the high prevalence of spurious solutions and the disturbing effects of outlying observations, along maximum likelihood estimation, open serious issues. We consider restrictions for the component covariances, to avoid spurious solutions, and trimming, to provide robustness against violations of normality assumptions of the underlying latent factors. A detailed AECM algorithm for this new approach is presented. Simulation results and an application to the AIS dataset show the aim and effectiveness of the proposed methodology.

EO1323: **Robust Bayesian regression**
Presenter: **Marco Riani**, University of Parma, Italy
Co-authors: Anthony Atkinson, Aldo Corbellini

The forward search provides a flexible and informative form of robust regression. We describe two ways of introducing prior information into the regression model used in the search, either through fictitious observations or through prior distributions of the parameters. The relationship between the two methods is established. The extension to the forward search is not entirely straightforward, requiring weighted regression. Forward plots are used to exhibit the effect of correct and incorrect prior information on inferences. Analysis of the 546 observations of the Windsor housing data shows the presence of several outliers and illustrates the effect of incorrect prior specification.

EO268 Room SH349 ANALYSIS OF DATA FROM COMPLEX SURVEYS Chair: Paola Vicard

EO0822: **Empirical likelihood approaches under complex sampling**
Presenter: **Yves Berger**, University of Southampton, United Kingdom

Data are often collected with unequal probabilities from stratified population. Empirical Likelihood is widely used in mainstream statistics. We propose a new empirical likelihood approach for sample data selected with unequal probabilities. In this situation, the standard empirical likelihood approach cannot be applied. Under a set of regularity conditions, the empirical log-likelihood function has an asymptotic chi-squared distribution. The proposed approach does not rely on variance estimates, re-sampling or joint-inclusion probabilities, even when the parameter of interest is not linear and does not have a normal distribution. An alternative approach is the pseudoempirical log-likelihood function which is not entirely appealing from a theoretical point of view, because it relies on a parameter (the design effect) which need to be estimated. A previous approach does not rely on design effect, and can be more accurate than the adjusted pseudoempirical approach. Standard confidence intervals based on variance estimates may give poor coverages, when normality does not hold. This can be the case with skewed data and outlying values. The proposed empirical likelihood confidence interval has good coverages and balanced tail errors even when the sampling distribution of the point estimator is not normal.

EO0610: **Resampling from finite populations: An empirical process approach**
Presenter: **Pier Luigi Conti**, Sapienza University of Rome, Italy
Co-authors: Daniela Marella, Fulvia Fulvia Mecatti

In sampling finite populations, several resampling schemes have been proposed. The common starting point is that, despite its excellent asymptotic properties, Efron's original bootstrap only works for i.i.d. data. This condition is not met in sampling finite populations, because of the dependence among units due to the sampling design. Hence, adaptations are needed to account for the non i.i.d. nature of data. Different versions of the standard bootstrap algorithm have been proposed in the literature. A new class of resampling procedures for finite populations is defined. Such a class appears to provide a unified framework that allows for encompassing other resampling algorithms already proposed. Its main theoretical justification is based on asymptotic, large sample arguments: the probability distribution of the original statistic and its approximation based on resampling converge to the same limit. Technically speaking, it is shown that a "finite population version" of the empirical process and its "resampled form" weakly converge to the same limiting Gaussian process. In a sense, this justification is similar to those given for classical bootstrap.

EO0619: **PC algorithm for complex survey data via resampling**
Presenter: **Daniela Marella**, University Roma Tre, Italy
Co-authors: Paola Vicard

The PC algorithm is one of the main methods for learning the structure of a Bayesian network from sample data. The algorithm uses conditional independence tests for model selection in graphical modeling and it is based on assumption of independent and identically distributed observations (i.i.d). The i.i.d. assumption is almost never valid for sample surveys data since most of the commonly used survey designs employ stratification and/or cluster sampling and/or unequal selection probabilities. The impact of complex design on i.i.d. based procedures can be very severe leading

to erroneous results, then alternative procedures are needed which allow for complex designs. The aim is to modify the PC algorithm using resampling methods for finite population in order to take into account the complexity of sampling design in the learning process.

EO1221: **Principle component analysis under complex survey designs**
Presenter: **Natalie Shlomo**, University of Manchester, United Kingdom
Co-authors: Duncan Smith
Principal components analysis (PCA) is an important statistical technique that allows analysts to explore highly dimensional sample data. A covariance matrix and means are estimated from the data, and these are used to generate the PCA. Standard estimation methods are based on the sample data being representative of the population such as the case for simple random sampling. Complex survey designs do not generate simple random samples from a population. The population might be partitioned into subgroups (strata) and sampling might be in multiple stages where the first stage PSU is a cluster of sample elements. In addition, clusters might not be selected with equal probability. We explore how PCA can be performed with data arising from complex survey designs and demonstrate in a simulation study.

EO332 Room Athlone STATISTICS FOR FUZZY- OR SET-VALUED DATA II Chair: Ana Belen Ramos-Guajardo

EO0832: **Dispersion measures for interval data**
Presenter: **Przemyslaw Grzegorzewski**, Warsaw University of Technology, Poland
Co-authors: Adam Kolacz
In classical statistics we admit uncertainty caused by randomness but all other elements of the statistical model should be given precisely. Hence a typical outcome of the experiment is a set of real values. However, in many situations we deal with assessments that are more perceptions than actual measurements, as when the data represent linguistic label information or if we cannot measure the results of the experiment precisely. In both cases quite often we meet interval data. However, to handle such data we need new mathematical models, different than used in the classical statistics, both to describe observations and to perform statistical reasoning. It appears that a starting point to construct proper tools for handling interval data is to realize which of the two possible perspectives: ontic or epistemic, should be accepted since this distinction has a strong impact on the concepts and techniques of statistical reasoning. In particular, the ontic and epistemic points of view yield different approaches to model and measure the dispersion of a sample that consists of intervals. We illustrate these differences and propose the generalizations of well-known measures of dispersion, like sample variance or range, adequate to each of the discussed views on interval data.

EO0801: **Probabilistic and possibilistic representations of information entities about time series**
Presenter: **Katarzyna Kaczmarek**, Polish Academy of Sciences, Poland
Co-authors: Olgierd Hryniewicz
Randomness, uncertainty and imprecision related to time series data may be formally described with probability theory or possibility theory (depending on context). As to give an example, discrete time series may be represented as a sequence of crisp observations of given objects property measured at successive points in time and at uniform time intervals. Each observation is a realization of the random variable. The traditional approach to the time series analysis is an iterative process and consists of the identification of the probabilistic model, estimation of its parameters, verification techniques and finally, prediction. Alternatively, in the context of very uncertain environment, time series may be defined as a sequence of fuzzy numbers. Furthermore, thanks to the advances in granular computing, there are competitive definitions related to more complex information granules, like association rules or linguistic summaries, that describe or summarize time series data with the use of fuzzy sets. Nonetheless, the relevance of such information granules to the probabilistic models is not extensively explored in the literature. We introduce and discuss a catalogue of exemplary probabilistic and possibilistic representations of the information entities related to time series data. Then, some modifications of the Kullback-Leibler divergence are considered to measure the quality of information granules.

EO1071: **On the extension of the Fisher linear discriminant model for interval data classification**
Presenter: **Ana Belen Ramos-Guajardo**, University of Oviedo, Spain
Co-authors: Przemyslaw Grzegorzewski
Random intervals have been considered in the literature to model those situations in which the characteristic of study cannot be measured precisely (due to different reasons) but the results of the experiment are given by closed intervals. If two characteristics are observed over the examined objects which belong to only one of two possible classes, each observation can be perceived as a non-empty closed and bounded rectangle on the real plane. The aim is to propose a two-dimensional classification method for two classes by following the Fisher's method on linear discriminant analysis. Thus, the idea is to find a vector which maximizes the ratio of the between-classes dispersion and the within-classes dispersion of the observed rectangles. Some properties of such a vector and the goal proposed function will be discussed. In addition, a classification rule will be suggested and some simulations will be carried out to show the behaviour of the classification procedure.

EC1601: **Estimation of multiple linear regression models with random linear constraints**
Presenter: **Marta Garcia Barzana**, University of Oviedo, Spain
Co-authors: Ana Colubi, Gil Gonzalez-Rodriguez, Erricos John Kontoghiorghes
The least squares estimation of linear regression models involves an optimization problem that may be subjected to a certain group of constraints. The well-known constrained least-squares approach assumes that the number of inequality linear constraints is fixed. This framework will be extended by removing such an assumption. Thus, the number of constraints can vary depending on the sample size. This situation arises naturally in the context of the arithmetic-based linear regression for interval data, where the set of constraints involved in the estimation problem has the same dimension as the number of observations. An estimators will be presented and its consistency will be discussed.

EO230 Room Bloomsbury SIMPSON'S PARADOX Chair: Tamas Rudas

EO1195: **On the Yule-Simpson paradox as a logical consequence of particular types of generating processes**
Presenter: **Nanny Wermuth**, Chalmers University of Technology, Sweden
One speaks of the Yule-Simpson paradox when the type of dependence of a given response on an explanatory variable of interest is similar in direction and strength when a third variable is given, but it appears to be qualitatively different overall, that is when information on this third variable is not used. Such situations have been rediscovered frequently and often shocked those who described them. We use instead simple data generating processes to explain why these situations are logical consequences of particular data generating processes and when they cannot occur.

EO0720: **Simpsons paradox: Polysemy or phantom**
Presenter: **Anna Klimova**, IST Austria, Austria
Co-authors: Tamas Rudas
The concept of Simpsons paradox refers to a situation when the inference about treatment effect depends on whether it was evaluated from the joint distribution of the treatment and outcome or from their conditional distributions given a covariate. If the measure of association used is collapsible, the paradox does not arise, although, as demonstrated, if a different measure is employed, the data set may still exhibit the paradox. Multiple

discussions of the paradox and the collapsibility property in the literature convey that authors associate different meanings with these concepts. Many authors attribute collapsibility to the absence of confounding, but disagree about what confounding is. Two interpretations of confounding are briefly discussed. In causal inference, the paradox is explained by misreading the causal structure in the data, and is claimed to be resolved. An example often used to justify this claim is presented. Some papers use the concepts of collapsibility and confounding without specifying their precise meaning and underlying assumptions. The statements about Simpsons paradox given in these papers may make the reader confused, especially if his/her knowledge about the paradox contradicts with the claims given by the authors.

EO0717: **On a linear concept of association**
Presenter: **Tamas Rudas**, Eotvos Lorand University, Hungary
The deviation of a binary distribution from the uniform may be measured with the difference of the two probabilities, just as well as with their ratio. For 2x2 contingency tables, this suggests a linear contrast; its zero value indicates the lack of linear association. Higher order linear association terms may also be defined, and they constitute a parameterization of multivariate binary distributions. Higher order linear interactions may be present in a distribution, but variation independence from lower dimensional associations does not hold, in contrast with the multiplicative case. The linear association term is directionally collapsible, meaning that if in a 2x2x2 table, the two conditional associations of two variables have the same direction, then their marginal association has this direction, too. Therefore, the linear association avoids Simpsons paradox in treatment by response tables. Surprisingly, if the linear contrast is defined in terms of the row conditional distributions, the paradox appears again. It occurs for the same data as if the odds ratio was used. One interpretation of this is that the occurrence of the paradox is not a consequence of using ratios instead of differences, rather of neglecting the allocation in treatment categories.

EO0965: **On some proposals to block Simpson's paradox**
Presenter: **Prasanta Bandyopadhyay**, Montana State University, United States
Simpsons paradox (SP) poses problems for those who work with observational data. We will first explore and then reject some of the current proposals to block SP. One of such proposals is the theoretical one, the other the practical one. There are in turn two versions of the theoretical proposal: (i) Restrict the use of the collapsibility principle, thus preventing the formulation of SP, and (ii) since SP is measure-sensitive; use the cross sum ratio measure systematically so that SP wont be generated. The practical proposal, in contrast, stems from the need to apply causal inference-engine to real cases. Causal theorists who have pioneered the causal inference-engine approach think that if we consider only the faithfulness condition, along with the causal Markov condition then SP (i.e., the cessation of a specific relationship in the overall population which exists in its sub-populations) cannot be generated. We will point out that the theoretical proposals to block SP are similar to proposals for blocking paradoxes in the foundations of mathematics, and as in the case of foundations of mathematics, they also have to pay a heavy price. We will also critically review the current defence of the faithfulness condition.

EO334 Room CLO 306 STRUCTURED DATA ANAYSIS Chair: Arthur Tenenhaus

EO1130: **Applications of constrained generalized SVD to various multivariate techniques**
Presenter: **Vincent Guillemot**, ICM, France
Co-authors: Herve Abdi
Generalized Singular Value Decomposition (GSVD) is a rich technique that encompasses several well-known multivariate data analysis methods such as partial least squares, principal components analysis, (generalized) canonical correlation analysis, (Multivariate) Correspondance Analysis, to name but a few. A constrained version of GSVD is proposed. Famous and very useful examples of such constraints are the L1-penalty, the group-LASSO, the fused-LASSO,... A very efficient proximal based algorithm is derived. The efficiency and the flexibility of the proposed method is demonstrated on a real datasets.

EO0848: **Sparse common and distinctive components for multi-set data analysis**
Presenter: **Katrijn Van Deun**, Tilburg University, Netherlands
Co-authors: Zhengguo Gu
Multi-set data consist of several heterogeneous data sets that are obtained for the same set of objects. An example is the combination of questionnaire, health, and (epi)genomic data for the same set of persons. Often it is of interest to find the common sources of variation underlying such multi-set data. Yet, these common sources may be very subtle, being governed by a limited number of variables only and dominated by source specific variation. Hence, we propose a novel component-based approach that separates common sources of variation from distinctive sources of variation and with components that are sparse in the number of non-zero loadings. The method is based on structured penalties like the sparse group lasso and the elitist lasso. A highly efficient algorithm is proposed and a stability selection procedure is used to tune the variable selection.

EO0825: **Data fusion of heterogeneous data sets: The Tucker3-PCA model as an illustrative example**
Presenter: **Tom Frans Wilderjans**, Leiden University, Netherlands
Co-authors: Elisa Frutos Bernal, Eva Ceulemans
In different fields of science, challenging research questions often imply that different blocks of (heterogeneous) information pertaining to the same research units are to be analysed simultaneously (i.e., data fusion). Examples can be found in psychology (e.g., simultaneous analysis of behavioural and questionnaire data or fusion of brain data from different modalities) but also in other fields, like bio-informatics (e.g., fusion of Fluorescence Spectroscopy, Nuclear Magnetic Resonance and Liquid Chromatography-Mass Spectrometry data). One such challenging question pertains to the disclosure of the common processes underlying coupled data blocks. To tackle this question, we propose a data fusion strategy that consists of factorizing each data block with an appropriate decomposition method (e.g., PARAFAC, Tucker3, PCA) and imposing the parameters of the common mode, which are estimated by using the information present in all data blocks, to be equal for all data blocks this mode belongs to. As an illustration of the proposed data fusion approach, we present the new Tucker3-PCA model for fusing a three-way real-valued data array and a two-way real-valued data matrix that share a single mode. An application of the model to the fusion of behavioural (i.e., situation-specific behaviour profiles) and questionnaire (i.e., person dispositions) data is discussed.

EO0539: **Kernel non-parametric tests of relative dependency**
Presenter: **Wacha Bounliphone**, Paris Saclay - CentraleSupelec, France
Co-authors: Arthur Gretton, Arthur Tenenhaus, Matthew Blaschko
Tests of dependence are an important tool in statistical analysis, and are widely applied in many data analysis contexts. For many problems in data analysis, however, the question of whether dependence exists is secondary: there may be multiple dependencies, and the question becomes which dependence is the strongest. We present a novel non parametric statistical method which describes hypothesis test of relative dependence between a source variable and two candidate target variables. Such a test enables one to answer whether one source variable is significantly more dependent on the first target variable or the second. Dependence is measured via the Hilbert-Schmidt Independence Criterion (HSIC), resulting in a pair of empirical dependence measures (source-target 1, source-target 2). Modeling the covariance between these HSIC statistics leads to a provable more powerful test than the construction of independent HSIC statistics by sub-sampling. The resulting test is consistent and unbiased, and (being based on U-statistics) has favorable convergence properties. The test can be computed in quadratic time, matching the computational complexity

of standard empirical HSIC estimators. The effectiveness of the test is demonstrated on several real-world problems: we identify language groups from a multilingual corpus, and we prove that the tumor location is more dependent to gene expression than chromosomal imbalances.

EO200 Room CLO B01 NONPARAMETRIC FUNCTIONAL DATA ANALYSIS Chair: Alicia Nieto-Reyes

EO1179: **Modeling functional data with complex dependencies via partial differential regularizations**
Presenter: **Laura Sangalli**, Politecnico di Milano, Italy
Co-authors: Mara Bernardi, Gabriele Mazza, James Ramsay
A class of models will be presented for the analysis of functional data with complex dependencies, such as spatially dependent curves and time dependent surfaces, observed over complex domains. The models are based on the idea of regression with partial differential regularizations. The methodology is illustrated via an application to the study of the annual production of waste in the municipalities of Venice province.

EO1288: **Tests based on spatial signs and ranks for data in infinite dimensional spaces**
Presenter: **Anirvan Chakraborty**, Ecole Polytechnique Federale de Lausanne, Switzerland
Co-authors: Probal Chaudhuri
Tests based on signs and ranks for paired sample and two sample problems involving univariate data are robust competitors of mean based tests like the $t-$test. For finite dimensional multivariate data, several extensions of sign and rank based tests have been shown to have better performance than Hotelling's T^2 test for many non-Gaussian distributions of the data. Paired sample and two sample tests for data in infinite dimensional spaces are considered that are based on notions of spatial sign and spatial rank in such spaces. The asymptotic properties of the proposed tests are discussed under the null hypothesis as well as under sequences of appropriate shrinking alternatives. These tests are compared with a wide range of competing mean based tests for infinite dimensional data in terms of their powers and robustness properties using some real and simulated data.

EO1087: **Abstract definitions of statistical depth for functional data**
Presenter: **Pedro Teran**, Oviedo, Spain
A number of definitions of statistical depth have been proposed for functional data. An abstract approach to handling those concrete examples might proceed in one of two ways. Firstly, by proposing a list of desired properties a depth function might ideally meet. It is typical, however, that many depth functions accepted in practice do not meet all the ideal properties. Secondly, by trying to give an abstract method to construct functions which contains concrete examples as particular cases of the construction. That is, one may either focus at common properties of concrete examples or try to subsume concrete examples under a unifying definition. Two approaches in the second direction of research will be introduced.

EO1312: **Amplitude and phase variation of point processes**
Presenter: **Yoav Zemel**, Ecole polytechnique federale de Lausanne, Switzerland
Co-authors: Victor Panaretos
The amplitude variation of a real stochastic process $\{Y(x)\}$ consists in its random oscillations in the y-axis, typically encapsulated by its (co)variation around a mean level. In contrast, phase variation refers to fluctuations in the x-axis, often caused by random time changes or spatial deformations. We consider the problem of identifiably formalising similar notions for point processes, and of nonparametrically separating them based on realisations of iid copies of the phase-varying point process. First principles regarding the phase variation lead inevitably to the geometry of probability measures induced from the problem of optimal transportation of measure. It is precisely this geometry that allows the consistent separation of the two types of variation for point processes.

EG127 Room Holden CONTRIBUTIONS ON LIKELIHOOD Chair: Alastair Young

EC1396: **Probit transformation for nonparametric kernel estimation of the copula density**
Presenter: **Gery Geenens**, University of New South Wales, Australia
Co-authors: Arthur Charpentier, Davy Paindaveine
Copula modeling has become ubiquitous in modern statistics. The problem of nonparametrically estimating a copula density is addressed. Arguably the most popular nonparametric density estimator, the kernel estimator is not suitable for the unit-square-supported copula densities, mainly because it is heavily affected by boundary bias issues. In addition, most common copulas admit unbounded densities, and kernel methods are not consistent in that case. A kernel-type copula density estimator is proposed. It is based on the idea of transforming the uniform marginals of the copula density into normal distributions via the probit function, estimating the density in the transformed domain, which can be accomplished without boundary problems, and obtaining an estimate of the copula density through back-transformation. Although natural, a raw application of this procedure was, however, seen not to perform very well in the earlier literature. It is shown that, if combined with local likelihood density estimation methods, the idea yields very good and easy to implement estimators, fixing boundary issues in a natural way and able to cope with unbounded copula densities. The asymptotic properties of the suggested estimators are derived, and a practical way of selecting the crucially important smoothing parameters is devised. Finally, extensive simulation studies and a real data analysis evidence their excellent performance compared to their main competitors.

EC1399: **Higher-order corrected likelihood for nonlinear models with fixed effects**
Presenter: **Yutao Sun**, KULeuven, Belgium
Nonlinear models with fixed effects could be severely biased. This is known as the incidental parameter problem (IPP). While several first-order solutions to the IPP have been developed by many researchers, a higher-order solution has been absent from the literature. We present a general approach to derive a higher-order corrected likelihood that can produce less biased estimators, for the common parameter, under independence, and in the IPP setting where N is allowed to go to infinite with T almost fixed to be small. The approach is fully general and can be invoked to derive a corrected likelihood to an arbitrary order. We provide an explicit formula for the second-order corrected likelihood and several examples of the application of such corrected likelihood demonstrating the bias-reducing feature.

EC1486: **Approximate likelihood inference for the Bingham distribution**
Presenter: **Marco Bee**, University of Trento, Italy
Co-authors: Roberto Benedetti, Giuseppe Espa
Likelihood inference for the Bingham distribution is difficult because the density function contains a normalization constant that depends on unknown parameters and cannot be computed in closed form. We propose to estimate the parameters by means of Approximate Maximum Likelihood Estimation (AMLE), thus bypassing the problem of evaluating the likelihood function. The method can be seen as a frequentist reinterpretation of Approximate Bayesian Computation (ABC) techniques. Instead of approximating the likelihood and then maximizing it, AMLE directly approximates the MLE by simulating observations form the distribution of interest. The restriction to uniform prior distributions is the crucial difference between AMLE and a classical ABC approach. We study the impact of the input parameters of the AMLE algorithm and suggest some heuristic approaches for choosing their numerical values and for performing statistical inference. When the dimension of the problem is small, simulation experiments and real-data applications suggest that the performance of the method is in line with MLE based on the approximation of

the normalizing constant via the Holonomic Gradient Method. In large dimensional problems we find that, according to our simulation-based evidence, AMLE is more efficient in terms of RMSE.

EC1516: **Distribution of likelihood-based $p-$values under the alternative hypothesis**
Presenter: **Alastair Young**, Imperial College London, United Kingdom
Co-authors: Stephen Lee
We consider inference on a scalar interest parameter in the presence of a nuisance parameter, using a likelihood-based statistic which is asymptotically normally distributed under the null hypothesis. Higher-order expansions are used to compare the repeated sampling distribution, under a general contiguous alternative hypothesis, of $p-$values calculated from the asymptotic normal approximation to the null sampling distribution of the statistic with those calculated by bootstrap approximations. Comparisons of different testing procedures in terms of power under an alternative hypothesis are closely related to differences under the null hypothesis, specifically the extent to which testing procedures are conservative or anti-conservative under the null. Empirical examples are presented which demonstrate that higher-order asymptotic effects may be clearly seen in small sample contexts.

EG297 Room CLO 204 CONTRIBUTIONS ON MISSING DATA Chair: Christophe Biernacki

EC1512: **Jackknife empirical likelihood for order-restricted statistical inference with missing data**
Presenter: **Heng Wang**, Michigan State University, United States
Co-authors: Ping-Shou Zhong
Although missing values appear very often in many applications, the problem of testing means with an increasing order or a decreasing order for data with missing values has not received much attention. In the existing literature, no formal procedure is available. Under the missing at random (MAR) assumption, we impute the missing values nonparametrically using kernel regression. For data with imputation, the classical likelihood ratio test designed for testing the order restricted means is no longer applicable since the likelihood does not exist. We propose a novel method based on jackknife empirical likelihood (JEL) ratio statistics. It is shown that the JEL ratio statistic evaluated under the null hypothesis converges to a chi-bar distribution, which is the same as that of the classical likelihood ratio test statistic. A simulation study shows that our proposed test maintains the nominal level well under the null and has prominent power under the alternative. The test is also robust for normally and non-normally distributed data. The proposed method is applied to an ADNI study for helping to find out a biomarker for the diagnosis of the Alzheimer's disease.

EC1722: **Estimating exceedance probability in environmental data**
Presenter: **Giuseppina Albano**, University of Salerno, Italy
Co-authors: Michele La Rocca, Cira Perna
Environmental data are tipically characterized by some stylized facts, such as non-stationarity, non linearity, seasonality, conditional heteroscedasticity and missing data. In particular, missing values make it difficult to determine whether the limits set by the European Community on certain indicators of air quality are fulfilled or not. In this perspective, to estimate the probability of each indicator to have exceeded the admitted limit becomes fundamental. We propose a procedure for estimating the probability of exceeding the legal limit. It combines a local estimator for the trend-cycle and a feedforward neural network as an estimator of the non linear component of the detrended time series. The first choice is justified in order to have flexible local structures which are not influenced by missing values outside the estimation window. The latter choice appears to be necessary since the detrended time series might show a non linear structure due both to the intrinsic characteristic of the data and to the cycle-trend estimation step. Indeed, the difficult task of selecting tuning parameters inthat step might possibly induce neglected nonlinearities in the detrended series. Finally, a bootstrapping procedure is implemented in order to estimate the probability to observe an exceedance of an established limit. In order to validate the procedure a simulation study is performed. An application to real data is also presented and discussed.

EC1373: **An information criterion for a subset of MAR data**
Presenter: **Keiji Takai**, Kansai University, Japan
Co-authors: Kenichi Hayashi
We discuss how to select a best model among candidate models with an information criterion (IC) when the data are missing at random (MAR). The problem in using the MAR data is that a subset of the MAR data may not be MAR and thus the consistent maximum likelihood (ML) estimator of interesting parameter cannot be obtained. To those data, various ICs developed so far for missing data cannot be applied because they assume that subsets of data on the candidate models are also MAR. However, this assumption is not realistic. Some of the variables that cause missingness might be excluded from the model, which result in not MAR data. Thus, we need a method to obtain the consistent ML estimator and develop an IC that can be used for subset data that are not MAR. Our goals are (i) giving a framework to obtain a consistent ML estimator for the sub-model parameter and its asymptotic distribution, (ii) developing an information criterion for the data which may not be MAR, and (iii) showing through numerical simulations how well our method works under some practical conditions.

EC1125: **Messy data analysis for endogenously allocated categories**
Presenter: **Maria Odejar**, University of Aberdeen, United Kingdom
Bayesian Markov chain Monte Carlo (MCMC) algorithms are developed for estimation and inference about missing data via endogenous supervised classification of these nonresponse into regimes corresponding to individual latent decision or net-benefit score. The MCMC algorithms are applied to bivariate ordered threshold model to provide estimates for the unresolved cases like dont knows, protests and item non-response in contingent valuation surveys. The MCMC algorithms are also developed for imputing the item nonresponse for sequential discrete decision models and currency hedge switching trade level model. Results show that inclusion of the missing responses is not detrimental to the quality of parameter estimates as reflected in the sum of log conditional predictive ordinate (SLCPO) goodness-of-fit criterion, predictive ability of the model in terms of apparent error rate in and out of the estimation sample, as well as mixing and convergence properties of the Markov chains generated by the MCMC Bayesian algorithm.

EG045 Room CLO 102 CONTRIBUTIONS ON COMPUTATIONAL METHODS FOR REGRESSION MODELS Chair: Stefan Sperlich

EC1586: **A numerical method for the estimation of high-dimensional time-varying parameters models**
Presenter: **Stella Hadjiantoni**, Queen Mary University of London, United Kingdom
Co-authors: Erricos John Kontoghiorghes
A novel numerical method to estimate univariate and multivariate time-varying parameters (TVP) models within the context of generalised least squares (GLS) is investigated. New methods are presented to derive the Kalman filter and Kalman smoother estimates of the TVP model. The proposed method is also extended to solve the downdating problem of removing the effect of some observations from the GLS solution. Moreover, a method for rolling window estimation of the TVP model is also examined. The new methods are based on numerically accurate and computationally efficient strategies. To reduce the computational cost, the special structure of the model is exploited and previous computations are utilised. Experimental results have demonstrated the computational performance of the proposed algorithms compared to existing ones.

EC1643: **Almost unbiased variance estimation in simultaneous equation models**
Presenter: **Yongdeng Xu**, Cardiff University, United Kingdom
Co-authors: Garry Phillips
While a good deal of research in simultaneous equation models has been conducted to examine the small sample properties of coefficient estimators, there has not been a corresponding interest in the properties of estimators for the associated variances. We build on previous research and explore the biases in variance estimators. This is done for the 2SLS and the MLIML estimators. The approximations to the bias are then used to develop less biased estimators whose properties are examined and compared in a number of simulation experiments. Two bootstrap estimators are also included, one of which is found to perform especially well. The experiments also consider coverage probabilities/ test sizes and test powers where it is shown that tests based on 2SLS are generally oversized while test sizes based on MLIML are closer to nominal levels. In both cases test statistics based on the corrected variance estimates generally have a higher power than standard procedures.

EC0905: **Estimating logit models with many fixed effects**
Presenter: **Amrei Luise Stammann**, Heinrich-Heine University Duesseldorf, Germany
Co-authors: Daniel McFadden, Florian Heiss
For the panel data analysis of binary dependent variables, the fixed effect logit model is a popular specification. The conditional fixed effects logit (CFL) estimator has the drawback that it does not deliver estimates of the fixed effects or marginal effects. It is also computationally costly if the number of observations per individual T is large. The dummy variable logit (DVL) estimator is a simple logit estimator with a dummy variable for each individual. It suffers from the incidental parameters problem which causes severe biases for small T. Another problem is that with a large number of individuals N, the computational costs can be prohibitive. We suggest a pseudo-demeaning algorithm that delivers the same results as the DVL estimator without its computational burden for large N. It uses the sparsity of the Hessian and the special features of the logit model. We also discuss how to correct for the incidental parameters bias of parameters and marginal effects. Monte Carlo evidence suggests that the bias-corrected estimator has similar properties as the CFL estimator in terms of parameter estimation. Its computational burden is much lower than the CFL or the DVL estimators, especially with large N and/or T.

EC1809: **Regression estimation by local polynomial fitting for multivariate data stream**
Presenter: **Aboubacar Amiri**, Charles de Gaulle University, France
Streaming data or data streams are massive data arriving in streams, and if they are not processed immediately or stored, then they are lost forever. In many real situations such that scientific and real applications, large amount of raw data can be collected extremely easily so that experiments typically yield to a huge number of data points. In those situations, the data arrive so rapidly that it is impossible for the user to store them all in disk (as a traditional database), and then interact with them later. Consequently, to deal with such big data, the traditional nonparametric techniques rapidly require a lot of time to be computed and therefore become useless in practice if real time forecasts are expected. We are interested to the regression model and we investigate how to provide simultaneously an estimation of a nonparametric regression function and its derivatives from a locally weighted least squares fit over the data streams. We propose a sequential technique based on a multivariate counterpart of the stochastic approximation method for successive experiments for the local polynomial estimation problem. Our estimator is a recursive version of the traditional estimator based on the locally weighted fitting adapted to the multivariate data stream model.

EP001 Room Macmillan Hall and Crush Hall POSTER SESSION II Chair: Panagiotis Paoullis

EC0449: **Multiple change points detection in weakly dependent random variables using FDqV method.**
Presenter: **Mohamed Elmi**, University of Djibouti, Djibouti
Let $X = (X1, X2, \cdots, Xn)$ be a time series, that is a sequence of weakly independent of random variables indexed by the time $t = 1, 2, \cdots, n$. We assume that there exists a segmentation $(1, 2, \cdots, K)$ such that Xi is a local stationary process for all time i in $(k, k+1]$, for $k = 1, 2, \cdots, K$, where K is unknown number of changes. The simplest model is to consider that Xi are autoregressive processes with change on the mean. We estimate the instants of breakpoints and means corresponding by using the filtered derivative and false discovery rate method(FDqV). As already established in the case of independent random variables, the FDqV has two steps: the first step computes the filtered derivative (FD) and then we select the potential change points as local maxima of the FD-function reaching a threshold and the second step (false discovery rate) eliminates false alarms and keeps as possible all right change points. We give a real application with heartbeat series, because in the observed data, it exists a correlation between those.

EC1454: **The Argentine North Central Rail line in the 1920s. A time series analysis for the transported cargo**
Presenter: **Maria de las Mercedes Abril**, Universidad Nacional de Tucuman, Argentina
Co-authors: Maria Beatriz Blanco
The railroads played a key role in the transportation of goods between different regions of our country. They can be seen as an irreplaceable instrument of economic and social progress. It should be noted that during the period under consideration, the State acted by stimulating the construction of railways and intervened directly on those regions where foreign capital did not want to invest. This determination was motivated by the desire to strengthen his authority as well as the political ties with local elites. Our goal is to use time series analysis techniques in order to study the cargo, especially the one transported through the North Central Line during the 1920s, using data obtained from the Ministry of Public Works.

EP1521: **Prediction of number of trades**
Presenter: **Artur Machno**, AGH University of Science and Technology, Poland
Co-authors: Henryk Gurgul
The characterization of time between trades is one of the fundamental elements of the market microstructure. Most of models describing duration, like Autoregressive Conditional Duration, require particular properties of the data. For instance, financial data posses cyclic patterns which are difficult to impose into the model directly. Instead, patterns are measured and then data is adjusted to them. This approach reduces applicability of these models. Firstly, estimated parameters describe the adjusted data, not the observed one, thus, the obtained results are not neatly interpretable. Secondly, the estimation of an error term is challengeable, since it comes from two main sources: pattern discoveries and parameter estimations. The purpose of the presented method is forecasting number of trades of the particular asset in the particular time interval. Therefore, our approach is an original concept as is not a duration model per se, but it concerns a closely related feature and can be used in similar applications. This approach needs no statistical pre-processing and detects cyclic patterns in the duration behaviour, especially, daily patterns.

EP1530: **Distribution and neural network based fuzzy time series forecasting for accident statistics**
Presenter: **Yonca Binici**, Anadolu University, Turkey
Co-authors: Ozer Ozdemir, Sevil Senturk
Fuzzy time series models have important effect for time series forecasting nowadays. By using artificial neural networks in fuzzy time series methods, neural network based fuzzy time series methods become popular for many study areas such as statistics, computer engineering, energy and meteorology. So, we used a new distribution and neural network based fuzzy time series method for forecasting accident statistics. We used various degrees of membership in establishing fuzzy relationships with various numbers of hidden nodes. We used artificial neural network

architecture in the step of method. The time series data are used for the new method and other methods in the literature. All results are compared each other.

EP1568: **A model review and data analysis for time series of counts**
Presenter: **Glaura Franco**, Universidade Federal de Minas Gerais, Brazil
The objective is to present a methodology review and an empirical comparison between observation and parameter driven models using procedures available in the literature that are commonly employed for time series of counts, based upon the conditional Poisson distribution. We aim to point out the main differences and similarities between the two methodologies, concerning parameter estimation, model fitting and forecasting, through an extensive simulation study. The models are fitted under the Bayesian framework, using a Poisson generalized linear model with a latent AR(p) process in the mean, which accounts for autocorrelation in the data. In the parameter driven case, the INLA approach is used, while for the observation driven models addressed, an MCMC procedure is developed to estimate the parameters. Concerning the covariates, coefficient estimates turn out to be remarkably similar across the different models. On the other hand, estimates for the autoregressive coefficients and the mean function and forecasts of future values depend heavily on the underlying process which generates the data. The results also show that when the parameters of the autoregressive process approach the non-stationary region, the parameter driven models present a much better performance. An application to real data of bankruptcy in United States companies is evaluated, and the parameter driven model has presented the best performance regarding prediction of future values.

EC1702: **A comparison of the power of tests for heteroscedasticity in non-linear econometrics models**
Presenter: **Iyabode Favour Oyenuga**, The Polytechnic Ibadan, Nigeria
Co-authors: Benjamin A Oyejola
There are several methods of estimating the power of test for heteroscedasticity in a single equation econometrics model. We investigate the power of White and Breusch-Pagan test for heteroscedasticity using three non-linear models: Cobb-Douglas, constant elasticity of substitution and exponential mode. Contaminated heteroscedasticity samples were generated using uniform distribution to generate data for capital K and labour L which were used to generate the output Y.The parameter estimates for the three models were obtained. The new set of parameters were used to compute the residuals which were used as the dependent variable for the auxiliary regression. The sample sizes used for the simulation were 10, 30, 50, 100, 150 and 200 with 10,000 replications. The results show that at 5% level of significance for the White test, as the sample sizes increase the power of the test decreases for the three models. In Breusch-Pagan test, as the sample sizes increase the power of the test increases for exponential model but its power is weak for the remaining two models.

EP1767: **A comparison of biased estimation methods for predicting tourism income of Turkey**
Presenter: **Esra Polat**, Hacettepe University, Turkey
One of the problems encountered in regression models is multicollinearity. If there is a multicollinearity problem, the variance of the least squares estimator may be very large and subsequent predictions rather inaccurate. In this case, biased estimation methods could be used to overcome the problem of inaccurate predictions. Biased estimation methods such as Ridge Regression (RR), Principal Component Regression (PCR) and Partial Least Squares Regression (PLSR) are used with the consequent trade-off between increased bias and decreased variance. RR is based on adding a biasing constant k to diagonal elements of X'X. PCR and PLSR methods discard the irrelevant and unstable information and use only the most relevant part of the x-variation for regression. Tourism sector in Turkey has shown great progress since 1980s. Contribution of foreign currency, while the country was having economic problems, helped to decrease foreign debt and unemployment. The annual data set of Turkey (1985-2014) including the factors (number of foreign tourists, total bed amount of tourism facilities having tourism operation license, number of tourism agencies, exchange rate of euro and exchange rate of US dollars) affecting the tourism income, is examined. RR,PCR and PLSR methods are compared in terms of fitting to data and predictive ability. Therefore, best model, giving the best prediction of tourism income, is selected.

EP1457: **Comparison of the average kappa coefficients of two binary diagnostic tests subject to a paired design**
Presenter: **Jose Antonio Roldan Nofuentes**, University of Granada, Spain
Co-authors: Maria del Carmen Olvera Porcel
The fundamental parameters to assess and compare the performance of binary diagnostic tests are sensitivity and specificity. When considering the losses associated with an erroneous classification with the binary diagnostic test, the parameter that is used to assess the binary diagnostic test is the weighted kappa coefficient. The weighted kappa coefficient depends on the sensitivity and the specificity of the binary diagnostic test, on the disease prevalence of the population studied and on the relative loss between the false positives and the false negatives (weighting index). Another parameter to assess and compare binary diagnostic tests are the average kappa coefficients. The average kappa coefficient of a binary diagnostic test is a chance corrected index between the binary diagnostic test and the gold standard, and it depends on the sensitivity and the specificity of the diagnostic test and on the disease prevalence. Several hypothesis tests are studied to compare the average kappa coefficients of two binary diagnostic tests subject to a paired design. Simulation experiments were carried out to study the type I errors and the powers of the hypothesis tests studied. The results were applied to a real example.

EP1527: **Spatial patterns of links between temperature extremes and mortality in the Czech Republic**
Presenter: **Jan Kysely**, Institute of Atmospheric Physics AS CR, Czech Republic
Co-authors: Ales Urban
The aim is to examine spatial patterns of the effects of high temperature extremes on all-cause and cardiovascular mortality in the population of the Czech Republic, by identifying relationships of heat-related excess mortality with socioeconomic and environmental factors in 76 administrative districts. The mortality baseline for every district was determined using Generalized Additive Models, with 7 degrees of freedom for the spline function in all models. Average relative deviations of mortality from the baseline were calculated on all summer days and days exceeding a high temperature threshold (the 90th / 95th percentile), and they were correlated with selected socioeconomic and environmental variables (including climatological and land-cover characteristics) for the districts. Groups of districts with similar characteristics were eventually aggregated to bigger regions in order to evaluate different heat-related effects in i) industrial vs. predominantly agricultural areas; ii) higher- and lower-elevated regions; and iii) regions with different overall socioeconomic level. After taking into account lagged effects of temperature on mortality, we find that heat-related mortality is most pronounced in highly urbanised regions that are located in the warmest parts of the Czech Republic. In contrast to all-cause mortality, excess cardiovascular mortality is related more to areas with low socioeconomic status.

EC1547: **Comparison of means under non-normality with outliers for unbalanced data**
Presenter: **Mustafa Cavus**, Anadolu University, Turkey
Co-authors: Berna Yazici, Ahmet Sezer, Ozer Ozdemir
There are many methods to compare the means of the groups used in practise under heteroscedasticity. Most of these methods require the assumption of normality with constant variance. Nevertheless, normally distributed populations are rarely encountered in the data sets. Many cases cause the violation of the normality assumption such as outlier in the data with unbalanced groups. The effect of nonnormality sourcing outlier in the data and unbalanced groups are used to compare the means of the groups by Fishers F Test, Generalized F Test, Parametric Bootstrap Test and Welch's

F Test. These tests are compared according to their powers and type I errors in the presence of outliers in the balanced and unbalanced data. Monte Carlo simulation results are discussed to compare the power of the methods in question.

EP1554: **Modeling wind energy with ANFIS with different membership functions**
Presenter: **Gultenkin Atalik**, Anadolou University, Turkey
Co-authors: Sevil Senturk

The importance of renewable energy sources is increasing day by day due to decreasing of fossil fuels. Wind energy is a kind of green energy sources. Modelling wind energy gains importance because of global warming. There have been done many studies on wind energy. Turkey has a big wind energy potential. So, it is important to model and estimate wind energy in Turkey. Adaptive Neuro Fuzzy Inference System (ANFIS) is an adaptive technique that combines neural network with fuzzy set theory. ANFIS uses a hybrid algorithm. That algorithm combines a backpropagation algorithm and the least squares method to make an estimation. There are many membership functions in ANFIS. These membership functions affect the model performance of ANFIS. We try to model wind energy with different membership functions to determine the best model for wind energy by using ANFIS method. Finally, the best model is determined via goodness of fit criteria.

EP1574: **Extracting the major factors affecting the traffic accidents using decision trees**
Presenter: **Betul Kan Kilinc**, Anadolu, Turkey
Co-authors: Mustafa Cavus

Traffic safety is one of the most important issues in transporting systems since traffic accidents with loss of life or injuired have some negative social and economic costs in society. A case study is presented in order to identify the main factors affecting the road accidents. A real dataset of traffic accidents on an urban road has been analyzed. The nonparametric method called Classification and Regression Trees is considered to study the traffic accidents data set as it does not require any priori probabilistic knowledge about the case under consideration. The findings are displayed graphically and interpreted in details.

EP1651: **A simulation study for the CTP regression model**
Presenter: **Jose Rodriguez-Avi**, University of Jaen, Spain
Co-authors: Maria Jose Olmo-Jimenez

The Complex Triparametric Pearson (CTP) distribution with parameters a,k and γ is a discrete distribution belonging to the family of Gaussian hypergeometric distributions but with complex parameters. It is a very flexible distribution that may be used to model both, overdispersed and underdispersed count data. For this reason we use it as a residual distribution in a regression model to describe a count data variable with respect to a set of independent and explanatory variables (covariates or regressors). We carry out a simulation study with two main objectives. Firstly, analysing the properties of the CTP regression model considering two possible parametrizations: the mean depending on the covariates, whereas the variance is fixed, and both, mean and variance, depending on the covariates. Secondly, comparing the performance of the CTP regression model with regression models for overdispersed count data, such as those based on the Poisson, Negative Binomial, Generalized Waring or Complex Biparametric Pearson (CBP) distributions, and also with regression models for over and underdispersed count data, such as those based on the COM-Poisson or hyper-Poisson distributions, among others. We pay particular attention to the comparison with the CBP regression model, since the underlying distribution is a particular case of a CTP distribution when the parameter a is equal to 0.

EP1670: **Distribution of inversions and the power of the $\tau-$Kendall test**
Presenter: **Mariusz Czekala**, Wroclaw Banking School, Poland

We present firstly a known result about the distribution of the number of inversions in a sequence of random variables. Under independence, we show the exact distribution of the number of inversions in the permutation, which is equivalent to a $\tau-$Kendall distribution. The difference is in normalizing constants. Considering inversions is more convenient. The inversion or the lack of inversion occurs almost surely. The probability of inversion equals 0.5 in the case of independence. This fact can be extended. We take p, in general greater or less than 0.5, as the probability of the inversion. For $p > 0.5$ we have negative association, but for $p < 0.5$ we have positive association. In both cases the values of moments, expected value and variance, and the exact distribution for all n is given. The asymptotic theorems are presented in the (known) independent case and in the dependent case. These results allow us to compute the power of the test based on the $\tau-$Kendall test.

EP1719: **Quality of educational services, image, students satisfaction and loyalty in higher education**
Presenter: **Matilde Bini**, European University of Rome, Italy
Co-authors: Lucio Masserini, Monica Pratesi

The objective is to investigate whether the perceived Quality of Educational Services and University Image influence students overall Satisfaction with their university experience, as well as the possible consequences of these relationships on students Loyalty. For this purpose, several hypotheses were formulated and tested through a Structural Equation Model by taking the following dimensions into account as possible determinants of overall Satisfaction and Loyalty: Quality of Educational Services, such as Teaching and Lectures, Teaching and Course Organization, Educational Infrastructure, Libraries, Refectories, Students Office, and University Image. Data were collected through a web questionnaire handed out to 14870 students enrolled at the University of Pisa in 2012, with items formulated to measure students perception on a four ordered categories anchored scale. Results give a valuable insight of how students perceive quality of educational services and institutional image of the university and the influence of these dimensions on Satisfaction and Loyalty. Specifically, it is worth noting that Teaching and Lectures and Teaching and Course Organization are main determinants of Students Satisfaction and Loyalty. Moreover, it is important to emphasize the crucial role exercised by University Image both on Teaching and Lectures and Satisfaction, and of the latter on Loyalty.

EC1788: **Estimation of the AIDS incidence rate in Portugal**
Presenter: **Rita Gaio**, University of Porto, Portugal
Co-authors: Julio Silva, Alexandra Oliveira, Joaquim Costa

The Human Immunodeficiency Virus (HIV) can evolve to severe illness with the late symptoms referred to as Acquire Immune Deficiency Syndrome (AIDS). In Portugal, the epidemic is monitored by a surveillance system based on cases notification by physicians. This process suffers from reporting-delay and no-reporting, and data on incubation times is scarce. Knowledge of the incidence rate is crucial to evaluate prevention campaigns and plans of action. A lower bound for the HIV incidence rate in Portugal (2000-2012), based on national AIDS surveillance data, is provided. Correction for the reporting delays is obtained by conditional likelihood estimation assuming a Poisson distribution for the number of diagnosed cases. The estimation of the HIV incidence rate is based on the back-calculation method. The number of newly infected cases in any given month is assumed to follow a Poisson distribution. The mean number of new infections per month is estimated from the complete likelihood function, given the number of new diagnoses per month. The incubation period is assumed to be time-dependent and to follow a Weibull or a Gamma distribution. Maximization of the complete likelihood uses a combination of the expectation-maximization algorithm and linear regression. Results from the methodology are tested on simulated and real data.

EP1782: **Category prediction on short-answer question using dynamic scoring algorithm**
Presenter: **Tomoya Okubo**, The National Center for University Entarance Examinations, Japan

We show some results on evaluating a dynamic scoring algorithm that enables us to score short-answer questions efficiently. The scoring system employs natural language processing techniques in order to calibrate similarity among examinees short-answers. The scoring system is able to show raters similarity index among the short-answers and it sorts the responses based on the similarity index; therefore, raters are able to score some responses at once. Further, the system dynamically predicts probabilities for each score-category based on fixed scores given by the raters. The information of prediction also helps the raters in the rating procedures. In high-stakes tests, stakeholders do not allow us to use full-automated scoring system because it cannot be always perfect. The scoring algorithm shown can be used in high-stakes tests since it is not a full-automated system and raters give rating for all the short-answers. It is important to have accountability for test-takers in high-stake testing. The scoring algorithm is able to keep accountability for ratings but also improves efficiency of scoring.

EP1649: **Simulation results for elliptical moment based extreme value index estimators**
Presenter: **Matias Heikkila**, Aalto University School of Science, Finland
Co-authors: Yves Dominicy, Pauliina Ilmonen

Extreme value theory is concerned with atypical behavior of random phenomena. It provides theory and tools to understand and estimate probabilities of very rare, potentially disastrous, outcomes in e.g. finance, climatology and seismology. We consider extreme behavior of multivariate random variables. More specifically, we consider the problem of determining their extreme-value index: a quantity that characterizes the manner the underlying distribution decays far away from the expected value. Several estimators for this are available in the univariate case. However, the generalization to the multivariate case is non-trivial. We are interested in generalizations of moment and mixed-moment estimators to the case of multivariate elliptical distributions. The behavior of the proposed new multivariate estimators is examined in a large simulation study, where several different distributions, heavy-tailed, light-tailed and bounded, are considered.

EP1621: **Optimal design on loading frequency for a probabilistic fatigue model in concrete**
Presenter: **M Jesus Rivas-Lopez**, University of Salamanca, Spain
Co-authors: Jesus Lopez-Fidalgo, Gonzalo Ruiz, Rena C Yu

Fatigue tests are known to be time consuming, sometimes can be unachievable if not properly designed. The number of tests needed to characterize certain materials is an open issue. Optimal experimental designs are especially useful when experimentation is expensive, time consuming or difficult. A good design may save time, money and provide a better fitting of the model. Fisher Information Matrix (FIM) is used to derive the optimal location of tests to characterize fatigue performance of concrete-related materials under given loading conditions. In particular, a fatigue model based on an initial distribution is chosen as an example to carry out the optimal design process. The developed methodology, however, can be applied to any other given fatigue model.

EP1839: **Efficient algorithm for solving the cumulative link model**
Presenter: **Panagiotis Paoullis**, University of Oviedo, Spain
Co-authors: Erricos John Kontoghiorghes, Ana Colubi

By formulating the cumulative link model as an iterative generalized linear least-squares problem, the difficulties in directly computing the inverses of the variance-covariance matrix and Fisher information matrix in Fisher scoring iteration are avoided. However, this method is computationally very inefficient, for large scale problems. A computationally efficient algorithm for solving the cumulative link models is proposed. The algorithm employs the generalized QR decomposition to solve the generalized exponential dispersion models formulated as an iterative generalized linear least-squares problem. The algorithm exploits the structure of the matrices involved, and in every step, it efficiently utilizes the initial factorizations. Experimental results, which confirm the computational efficiency of the proposed algorithm, are presented.

EP1799: **Exponentiated general hazard to compare risk of death for different states of breast cancer**
Presenter: **Gleici Perdona**, USP, Brazil
Co-authors: Francisco Louzada, Hayala Cristina Cavenague de Souza, Fernanda Maris Peria

In oncology, a proportion of individuals who do not experience a poor outcome such as death recurrence in the end of follow-up have become more common. Hazard models that include such a property are called cure fraction models and the most popular type of cure rate model is the mixture model. Also, it is wished that hazard models are flexible enough to accommodate different shapes. In order to unify these properties, the Exponentiated General Hazard (EGH) model has previously been proposed, which extends several distributions widely used in the lifetime literature and allows the accommodation of non-monotone hazard function shapes, such as bathtub and unimodal. We propose to apply this model in breast cancer data, because in this formulation one parameter is directly connected with the intensity of occurrence of the event of interest (death or relapse) and this means, for a physician, a tool for interpretation and discrimination between stages of cancer. The estimation properties of particular cases of EGH model, the Exponentiated Modified Weibull (EMW) are given. The results were applied for a real breast cancer data of women treated on Hospital in Brazil.

EC1771: **Constrained clustering and anomaly detection on time series data**
Presenter: **Carlo Drago**, University of Rome Niccolo Cusano, Italy

The Anomaly Detection approach is useful to identify statistical units defined as anomalies on a definite pattern in a dataset. Various different techniques were proposed in literature in order to perform anomaly detection on various frameworks and application fields. We consider a strategy of anomaly detection for time series data. In particular, we consider a constrained clustering approach in order to detect observation which shows a deviation from the observed patterns of the series. This technique is used to identify anomalies on stock market data.

EC1584: **Applications of influence functions and robustness in statistical process control**
Presenter: **Luan Jaupi**, CNAM, France
Co-authors: Philippe Durand, Dariush Ghorbanzadeh

Applications of influence functions and robustness for process monitoring are considered and new influential control charts to monitor process mean, variability and correlation structure are proposed. To monitor the process mean the influence function of mean is proposed and to investigate process variability control charts based on influence function of eigenelements are suggested. When process parameters are unknown, control charts can be applied in a two phase procedure. In Phase I, control limits are calculated based on estimated process parameters, but in Phase II the in-control values of mean and dispersion are assumed to be known and they are used to build up control charts. The estimator used in the Phase I analysis does not necessarily have to be the same one used to construct control limits for use in Phase II. Robust estimators will be preferable for situations where outliers are present, but their benefit is primarily for Phase I applications. The proposed techniques are general, and the influence functions may be used to build up control charts relative to either nominal values or estimates. Each method is further illustrated with real datasets, from manufacturing systems.

Monday 14.12.2015 08:30 - 10:10 Parallel Session L – CFE-CMStatistics

CO492 Room MAL B20 ANALYSIS OF EXTREMES AND DEPENDENCE Chair: Artem Prokhorov

CO0189: **Fat tails and copulas: Limits of diversication revisited**
Presenter: **Rustam Ibragimov**, Imperial College London, United Kingdom
Co-authors: Artem Prokhorov
The problem of portfolio risk diversification in a Value-at-Risk framework with heavy-tailed risks and arbitrary dependence captured by a copula function is considered. Using power law distributions, we explore whether the benefits of diversification persist when the risks in consideration are allowed to have extremely heavy tails with tail indices less than one and various dependence structures, represented by different classes of copulas. We show theoretically that for asymptotically large losses with the Farlie-Gumbel-Morgenstern copula, the threshold value of tail indices, at which diversification stops being beneficial, is the same as for independent losses. In other words, diversification increases Value-at-Risk for tail indices less than one regardless of the strength of dependence between portfolio components. We further extend this result to a wider range of dependence structures which can be approximated using the FGM copula or its generalizations. This range of dependence structures includes many well-known copula families, among which there are comprehensive, Archimedian, asymmetric and tail dependent copulas. A wide set of simulations supports these theoretical results.

CO0525: **Bayesian adaptive sparse copula analysis**
Presenter: **Martin Burda**, University of Toronto, Canada
Co-authors: Artem Prokhorov
Bayesian nonparametric models based on infinite mixtures of density kernels have been recently gaining in popularity due to their flexibility and feasibility of implementation even in complicated modeling scenarios. In the social sciences, they have been particularly useful in estimating nonparametric distributions of latent variables. However, these models have been rarely applied in more than one dimension. Indeed, implementation in the multivariate case is inherently difficult due to the rapidly increasing number of parameters needed to characterize the joint dependence structure accurately. We propose a multivariate density estimation procedure based on an adaptive Bayesian copula link function on a sparse grid. We show that this procedure leads to an improvement in the precision of a multivariate density estimate relative to commonly used alternatives.

CO1270: **A new measure of vector dependence, with an application to financial contagion**
Presenter: **Ivan Medovikov**, Brock University, Canada
Co-authors: Artem Prokhorov
We propose a new nonparametric measure of association between an arbitrary number of random vectors. The measure is based on the empirical copula process for the multivariate marginals, corresponding to the vectors, and is insensitive to the dependence between the within-vector components. It is bounded by the $[0,1]$ interval, covering the entire range of dependence between vector independence and a vector version of an exact monotonic relationship. We study the properties of the new measure under several well-known copulas and provide a non-parametric estimator of the measure, along with its asymptotic theory, under fairly general assumptions. To illustrate the applicability of the new measure, we use it to assess the degree of interdependence between the financial markets in the North and South America, Europe and Asia, surrounding the financial crisis of 2008. We find strong evidence of previously unknown contagion patterns, with selected regions exhibiting little dependence before and after the crisis and a lot of dependence during the crisis period.

CO1678: **Copula by triangulation, with application to tail dependence estimation**
Presenter: **Artem Prokhorov**, U Sydney, Australia
Co-authors: Yajing Zhu, Edward Anderson
Copulas are distributions with uniform marginals. The uniformity condition is a key copula property which is hard to impose in estimation. Many nonparametric estimators suffer from poor finite sample properties due to violations of this condition. We develop a new $B-$spline estimator of a general bivariate copula based on triangulation, which ensures that marginals are uniform by construction. We look at the properties of the new estimator, both asymptotically and in finite samples, and compare it with available alternatives such as the Kantorovich-Bernstein sieve estimator, Data-Mirror and Exponential Series estimators. The new estimator dominates the alternatives in terms of MSE and computational efficiency. We also show that violations of the uniformity condition lead to severely distorted estimates of the tail dependence coefficient. These biases can be corrected by using the new copula estimator.

CO498 Room MAL 633 FINANCIAL TIME SERIES AND RISK PREMIA Chair: Jeroen Rombouts

CO0234: **Semi-Markov multi-fractal volatility models**
Presenter: **Arnaud Dufays**, Universite Catholique de Louvain, Belgium
A new Semi-Markov volatility model is specified that generalizes the Discrete Markov-switching Multi-fractal (MSM) volatility model. The process exhibits many desirable properties to fit financial time series such as the parsimony in the number of parameters and the hyperbolic decline of the autocorrelation function of the squared returns. Furthermore, this representation is not constrained by the number of latent Markov processes or by the marginal distribution of the multipliers, a limitation that arises in the standard MSM model. An empirical exercise provides evidence in favor of the new specification compared to standard volatility models as well as the MSM one.

CO0542: **A geometric treatment of time-varying volatilities**
Presenter: **Chulwoo Han**, Durham University, United Kingdom
A new framework for addressing multivariate time-varying volatilities is proposed. By employing the coordinate-free methods of differential geometry, our method respects the geometric structure of the covariance space, i.e., symmetry and positive definiteness, in a way that is independent of any local coordinate parametrization. Based on our geometric framework, we develop a fundamental time-varying volatility model. Our model preserves the symmetry and positive definiteness of the covariance matrix as it evolves, without imposing any ad hoc restrictions. Its parsimonious specification makes it particularly suitable for large dimensional systems. Simulation studies suggest that our model possesses much of the nonlinear nature of covariance dynamics observed in the market. Applied to the US and the UK stock markets, the model performs well, especially in terms of risk measurement. Its superiority over other models, however, demands further evidence. In a broad context, our framework presents a new approach treating nonlinear properties observed in the financial market, and numerous areas of application can be further considered.

CO1279: **Bounded influence filtering in general state space models**
Presenter: **Ruben Crevits**, KU Leuven, Belgium
Co-authors: Christophe Croux
In time series analysis state space models are very popular. Often it is interesting to sequentially estimate the distribution of the hidden states given all observations available. That problem is called filtering. For linear Gaussian state space models the filtering distribution can be exactly computed with the Kalman filter. For nonlinear or nongaussian models one can rely on particle methods to estimate an approximation of the filtering

distribution. Standard methods as the Kalman or bootstrap particle filter are not robust against outlying observations. In the past researchers have attempted to robustify the filters for general nonlinear nongaussian state space models, but the newly introduced Bounded Influence Filter is a filter for which the influence of a new observation on the estimated filtering density is bounded. The performance of the new filter is compared with the existing methods for an outlier contaminated and for a clean case. It turns out that the Bounded Influence Filter performs very well in the contaminated case, while still performing adequately in the clean case.

CO1411: **Nonlinear dynamics in the risk aversion coefficient of strategic investors**
Presenter: **Jose Olmo**, University of Southampton, United Kingdom
The aim is to study the portfolio decisions and risk aversion attitudes of individuals characterized by three distinguishing features. First, their optimal decisions involve a multiperiod utility function. Second, the risk aversion coefficient is dynamic and driven by a set of state variables reflecting economic conditions. Third, the model accommodates nonlinearities interpreted as differences in risk aversion to the long and short term. These nonlinearities are tested using a $p-$value transformation method. The application to a tactical portfolio reveals asymmetries in the impact of negative events on long term and short term risk aversion highlighting the differences between myopic and strategic investors.

CO462 Room MAL G15 EARLY WARNING SYSTEM AND SYSTEMIC RISK INDICATORS II Chair: Gregor von Schweinitz

CO0258: **Optimizing policymakers' loss functions in crisis prediction: Before, within or after**
Presenter: **Gregor von Schweinitz**, Halle Institute for Economic Research, Germany
Co-authors: Peter Sarlin
Early-warning models most commonly optimize signaling thresholds on crisis probabilities. The ex-post threshold optimization is based upon a loss function accounting for preferences between forecast errors, but comes with two crucial drawbacks: unstable thresholds in recursive estimations and an in-sample overfit at the expense of out-of-sample performance. We propose two alternatives for threshold setting: (i) including preferences in the estimation itself and (ii) setting thresholds ex-ante according to preferences only. We provide simulated and real-world evidence that this simplification results in stable thresholds and improves out-of-sample performance. Our solution is not restricted to binary-choice models, but directly transferable to the signaling approach and all probabilistic early-warning models.

CO0790: **Modeling sovereign risk with correlated stoschastic processes**
Presenter: **Paolo Giudici**, University of Pavia, Italy
Co-authors: Laura Parisi
We use stochastic processes and correlation networks to model systemic risk between the economies of the European monetary union, in the post crisis period. For each country, we consider a financial leverage ratio and model the time dynamic of its two components by means of a linear combination of two stochastic equations: a Eurozone systematic process and a country specific idiosyncratic process. Doing so, we model debt sustainability in terms of both the financial and the real side. We provide an estimation model for the parameters of the processes and the resulting default probabilities. Systemic risk is estimated by means of the estimated marginal and partial correlation matrix that is a function of the estimated parameters of the process.

CO1038: **An early warning model based on financial tweets data**
Presenter: **Paola Cerchiello**, University of Pavia, Italy
Co-authors: Paolo Giudici
A very important topic in systemic risk modelling is the estimation of the interrelationships between financial institutions, with the aim of establishing which of them are more central and, therefore, more contagious/subject to contagion. The aim is to develop a novel model based on factorial graphical models, that are a class of network models on the residuals from a factor analysis. A model that, in addition, does not employ information contained in financial market prices, but takes advantage of a specific kind of social network big data: financial tweets. Indeed some researchers have proved the presence of a correlation between sentiment associated to tweets and financial trends, suggesting the employment of tweets contents as a tool for alert warning in systemic risk. Our proposed model focuses on UK listed banks for which we have collected relevant tweets and analyzed their contents. The results show that the model sheds further light on the interrelationships between financial institutions and can lead to the construction of an effective early warning indicator based on social networks data.

CO1202: **Characterizing the financial cycle: Evidence from a frequency domain analysis**
Presenter: **Christian Proano**, Otto-Friedrich-University Bamberg, Germany
Co-authors: Till Strohsal, Juergen Wolters
A growing body of literature argues that the financial cycle is considerably longer in duration and larger in amplitude than the business cycle and that its distinguishing features became more pronounced over time. An empirical approach suitable to test these hypotheses is proposed. We parametrically estimate the whole spectrum of financial and real variables to obtain a complete picture of their cyclical properties. We provide strong statistical evidence for the US and slightly weaker evidence for the UK validating the hypothesized features of the financial cycle. In Germany, however, the financial cycle is, if at all, much less visible.

CO350 Room MAL 539 FINANCIAL ECONOMETRICS Chair: Niklas Ahlgren

CO0462: **Noncausality and the commodity currency hypothesis**
Presenter: **Matthijs Lof**, Aalto University, Finland
Co-authors: Henri Nyberg
New evidence on the role of exchange rates in forecasting commodity prices is provided. Consistent with previous studies, we find that commodity currencies hold out of sample predictive power for commodity prices when using standard linear predictive regressions. After we reconsider the evidence using noncausal autoregressions, which are able to accommodate the effects of nonlinearities and omitted variables, the predictive power of exchange rates disappears.

CO0518: **Finite-sample multivariate tests for ARCH in vector autoregressive models**
Presenter: **Niklas Ahlgren**, Hanken School of Economics, Finland
Co-authors: Paul Catani
Finite-sample multivariate tests for ARCH effects in vector autoregressive (VAR) models are proposed. We apply Monte Carlo (MC) testing techniques. The tests are based on Cholesky-standardised multivariate residuals. We use a parametric bootstrap to circumvent the problem that test statistics in VAR models are not free of nuisance parameters under the null hypothesis. The tests under consideration are a combined equation-by-equation LM test, the multivariate LM test and an LM test of constant error covariance matrix. The tests are evaluated in simulation experiments. The combined equation-by-equation LM test has good size and power properties both when the data are Gaussian and non-Gaussian. The test outperforms the multivariate LM test in terms of power. The LM test of constant error covariance matrix outperforms the combined LM test in

terms of power when the data are Gaussian, but suffers from serious size problems when the data are skewed or heavy-tailed. The combined equation-by-equation LM test is the preferred test for ARCH with skewed or heavy-tailed financial data.

CO1126: **The expectation hypothesis of the term structure of very short-term rates: Evidence from a new testing approach**
Presenter: **Vanessa Gunnella**, University of Bologna, Italy
The purpose is to empirically test the Expectation Hypothesis (EH) of the term structure of the US repurchasing agreements (repo) rates, considered in a Vector Autoregression (VAR) model. A multiple hypotheses approach is adopted, in order to jointly test all the statistical hypotheses implied by the EH, i.e. the long-run and short-run implications of the theory. Furthermore, the testing procedures are carried out by taking into account heteroskedasticity through bootstrap inference, White correction and rolling windows analysis. Differently from previous results, overall evidence in favor of the statistical non-rejection of the EH is found. In particular, the rolling window analysis clarifies that the EH has been rejected only during periods of turmoil of the financial/repo markets.

CO0630: **International stock market comovement outlined with a thick pen**
Presenter: **Agnieszka Jach**, Hanken School of Economics, Finland
The aim is to quantify time-varying, bi-variate and multivariate co-movement between international stock market returns, across various time scales, using a novel approach called thick pen transform. We study 11 countries and examine their comovement with respect to (non-dyadic) time scales/frequencies, development and region. We also consider all possible 2036 different combinations of 2 or more of these countries. In the bi-variate set-up with the USA, the BRIC countries, except for Brazil (especially over small time scales), offer diversification benefits, while in the multivariate one, clustering with respect to America or Europe (but not Asia) leads to homogeneous groups. Hence development and region cannot always be considered as ultimate clustering factors. Leave-one-out cross-validation shows a nuanced interplay of time scales, development and region as grouping factors for Brazil and Australia, and even more so for UK, Japan and Hong Kong.

CO400 Room MAL 532 BAYESIAN ECONOMETRICS Chair: Deborah Gefang

CO0463: **A Jacobian approach to solve the incidental parameter problem**
Presenter: **Guangjie Li**, Cardiff University, United Kingdom
A strategy to solve the incidental parameter problem is proposed. The strategy is to find the Jacobian term from the original incidental parameters in the model to the new incidental parameters motivated by information orthogonality and to use the Jacobian term as the bias removing prior. We demonstrate the strategy for linear autoregressive models of order p greater than or equal to 1 with strictly exogenous regressors and predetermined regressors. For these examples, even though information orthogonality does not exist, the Jacobian term found by our strategy can produce the posterior mode which is a consistent estimator of the common parameters.

CO0612: **Combining density forecasts under various scoring rules: An analysis of UK inflation**
Presenter: **Fabian Krueger**, Heidelberg Instititute for Theoretical Studies gGmbH, Germany
Density forecast combinations are often found to perform better than individual forecasts under the log score criterion. Focusing on UK inflation, we analyze whether this result carries over to a range of other scoring rules. These scoring rules express the utility derived from a density forecast, given a realization. We consider both traditional choices and rules that emphasize the Bank of England's inflation target. In addition to the Bank's two-piece-normal density, our combined model contains two Bayesian vector autoregressive and two Bayesian Markov switching specifications.

CO1073: **On the implications of time variation in reduced rank econometric models for Bayesian forecasting**
Presenter: **Nalan Basturk**, Maastricht University, Netherlands
Co-authors: Lennart Hoogerheide, Herman van Dijk
Three well-known classes of econometric models - instrumental variable, vector auto regressive and factor models - have a similar structural shape of the likelihood and posterior, which is important for identification analysis. We show existence conditions for posterior moments of the parameters of interest under weak prior information and propose some identification strategies through alternative priors. Given that these shapes may vary over time, we explore this feature, using a smoothly mixing model combination approach, for two forecasting problems: time varying effects of education on income in economics and time varying patterns in momentum strategies in finance.

CO1313: **Asymmetric volatility spillovers between UK regional worker flows and vacancies**
Presenter: **Deborah Gefang**, University of Leicester, United Kingdom
Volatility spillovers between UK regional job finding, job separation and vacancy rates are investigated. Employing a logistic smooth transition vector autoregression (VAR) to model the large nonlinear dynamic system, we use previous methods to decompose the forecast error variances. Our approach is Bayesian. More specifically, we extend doubly adaptive elastic-netLasso (DAELasso) methods for VAR parameter shrinkage into a nonlinear framework to allow for the possible regime changes. We find that for each variable, both the volatility spillovers to and from other variables are high, providing clear evidence for the close interdependence between UK regional labour markets. The pivotal role of London in generating and spreading changes in volatility is highlighted. Analysis of net spillovers shows that, in general, shocks to job separation rates tend to spread into job finding and vacancy rates. By contrast,vacancy rates are usually at the receiving ends of shocks transmitted from the job separation and finding rates. We further examine the shock propagation mechanism in more detail, such as the differences in spillovers between regions within the same regime, and that of the same region but in different regimes. Finally, we draw inferences that are of economic and policy importance.

CO372 Room MAL 402 FINANCIAL NETWORKS: SYSTEMIC COEVOLUTION OF FINANCIAL SYSTEMS Chair: Fabio Caccioli

CO0649: **Application of filtered networks to Finance**
Presenter: **Tiziana Di Matteo**, Kings College London, United Kingdom
We are witnessing interesting times rich of information, readily available for us all. Using, understanding and filtering such information has become a major activity across science, industry and society at large. It is important to have tools that can analyze this information and that can provide ways to reduce complexity and dimensionality. We propose network-theoretic tools to filter information in large-scale datasets and we show that applications to financial data-sets can meaningfully identify industrial activities and structural market changes. Planar filtered graphs are powerful tools to study complex datasets: a clustering can be extracted allowing dimensionality reduction in a deterministic manner without the use of any prior information. They can also be used to diversify financial risk by building a well-diversified portfolio that effectively reduces investment risk. However, the algorithm so far proposed is numerically costly and cannot be applied to large-scale data. There is therefore scope to search for novel algorithms that can provide, in a numerically efficient way, such a reduction to planar filtered graphs. We introduce a new algorithm, the Triangulated Maximally Filtered Graph, that is scalable to very large datasets and can take advantage of parallel and GPUs computing. The method is adaptable allowing online updating and learning with continuous insertion and deletion of new data as well changes in the strength of the similarity measure.

CO0820: **Systemic liquidity risk in the interbank network**
Presenter: **Gerardo Ferrara**, Bank of England, United Kingdom
Co-authors: Sam Langfield, Zijun Liu, Tomohiro Ota
UK banks' systemic liquidity risk is studies by using a unique dataset on banks' daily cash flows and short-term interbank funding. We identify banks that fall short of liquidity owing to idiosyncratic stress (individually illiquid banks) and banks that fall short of liquidity because their bank counterparties fail to repay debts (systemically illiquid banks). To do this, we simulate a dynamic model in which illiquidity externalities propagate in the interbank network over multiple days. This modelling approach allows us to identify liquidity SIFIs the failure of which would have a significant impact on other banks through liquidity contagion. We also calculate the optimal distribution of liquid assets between banks when the total amount of liquid assets available is limited, and show that the impact of a liquidity crisis will be reduced when liquidity contagion risk is taken into consideration.

CO0793: **Complexity driven collapse of economic equilibria**
Presenter: **Giacomo Livan**, University College London, United Kingdom
Co-authors: Marco Bardoscia, Matteo Marsili
Our economies have evolved to remarkable degrees of complexity. In the pursuit of ever increasing efficiency and growth, with nested production processes feeding each other in order to create products of greater sophistication from less sophisticated ones, down to raw materials. The engine of such an expansion have been competitive markets, whose functioning is described by the mathematical framework of General Equilibrium Theory (GET), which indeed shows that, under specific conditions, competitive markets achieve efficient allocation. Yet, we know from the study of heterogeneous systems in other disciplines, that complexity may bring about unintended consequences that are hard to anticipate, as they are non-trivial emergent properties. Such are the algorithmic and freezing phase transitions in high-dimensional random systems. We show that a non-trivial freezing transition also occurs in GET of complex economies, modelled as large random systems. Specifically we show that when the fraction of non-primary goods, i.e. goods that result as an output of a production process, exceeds a critical threshold, the economy freezes in a state where all production processes collapse.

CO0866: **Proximity-based networks and filtered networks in economic and financial systems**
Presenter: **Rosario Mantegna**, Central European University, Hungary
Several years ago we have shown that part of the information stored in a proximity matrix can be efficiently summarized in a network. Examples of proximity based networks are minimum spanning trees, planar graph networks and partial correlation networks. The topology and structure of these networks are highly informative. The most typical set of complex network comprises networks where links are observed between two nodes when a relationship or an event is occurring between them. When a complex network presents deviations form the corresponding configuration model, a technique able to statistically validate the over- or the under-expression of a given relationships with respect to a null hypothesis is highly informative. We briefly discuss the concept of statistically validated networks and present some applications.

CO412 Room MAL 541 MIXTURE MODELS, IDENTIFICATION, AND INDEPENDENT COMPONENTS Chair: Markus Haas

CO0887: **Normal mixture variance states for structural identification of FX market shocks**
Presenter: **Sebastian Mueller**, Christian-Albrechts-University Kiel, Germany
Co-authors: Markus Haas
In multivariate models the interacting dynamics of several endogeneous variables can be estimated and projected to the future. When it comes to a structural perspective these models need to borrow from possibly competing economic theories. Identifying assumptions are imposed to reveal contemporaneous causalities, since the reduced form model does not provide sufficient information. Historically, extra information is introduced through restrictions on the instantaneous parameters, e.g., through a simple Cholesky decomposition. Recent statistical alternatives make use of information from higher-order moments. The identification problem can be solved by exploiting normal mixture properties of the considered data. A normal mixture variance structural vector-autoregressive model (NMH- SVAR) is employed to investigate the contemporaneous causal transmissions of structural shocks between the real and financial economy, the central bank and the Brazilian Real/ US-Dollar FX market. Additionally, an estimation procedure in closed-form is applied and further helpful properties of the normal mixture approach are highlighted. Subsequently, results are compared to (competing) economic theories like information-to-prices vs. feedback-trading and information-signalling between different groups of FX market participants.

CO0752: **Getting out of the COMFORT Zone: The MEXI distribution for asset returns**
Presenter: **Jeffrey Naef**, University of Zurich, Switzerland
Co-authors: Marc Paolella, Pawel Polak, Ronald W Butler
Most models for multivariate asset return data assume equal tail indexes for each asset. This can result in non-optimal asset allocation and have potentially detrimental implications for risk management. A new class of very flexible distributions is proposed which nest various models currently in use, and such that, among other things, each margin can have its own tail index. Estimation methods and empirical performance of some of the new structures are presented.

CO0486: **Simulation-based method for portfolio optimization for copula models**
Presenter: **Jochen Krause**, University of Zurich, Switzerland
Co-authors: Marc Paolella, Pawel Polak
A simple t-copula with noncentral Student's t APARCH margins is used as a model for asset returns and portfolio optimization. Using a non-standard method of estimation, the model can be estimated nearly instantaneously, and its performance enhanced using shrinkage methods. The portfolio distribution is obtained via a combination of simulation and parametric approximation for speed enhancement. An approximate method for mean-expected shortfall portfolio optimization is presented, and its performance on two data sets and comparison to competing models is presented.

CO0726: **Stochastic dominance criteria for normal mixture distributions**
Presenter: **Markus Haas**, University of Kiel, Germany
If markets are subject to stochastic regime changes with Gaussian distributions within each regime, then the overall distribution of portfolio returns is a mixture of normals. To facilitate the application to the ranking of portfolios in markets characterized by regime-switching, recently established stochastic dominance criteria for Gaussian mixture distributions are considerably extended and simplified. E.g., the distributions of the portfolios in the comparison may have different regime probabilities, or even a different number of regimes. Moreover, both second- and fourth-order stochastic dominance criteria for normal mixture distributions are considered, where the latter is shown to fit nicely into recent definitions of higher-order risk attitudes (in particular, temperance) as preferences over particular lottery pairs. Extending the scope to fourth- order stochastic dominance allows the comparison of portfolios with the same overall variance, and, in case a risk-free asset is available, turns out to be equivalent to a comparison of the regime-specific and overall Sharpe ratios (excess return per unit of standard deviation) of the portfolios involved in the comparison.

CO582 Room MAL 421 RASTANEWS SPECIAL SESSION ON VOLATILITY AND TIME SERIES MODELING Chair: Christian Conrad

CO1193: **Time-varying volatility persistence and the business cycle**
Presenter: **Karin Loch**, Heidelberg University, Germany
A new GARCH model that allows for a time-varying volatility persistence is presented. Information on the macroeconomic environment is directly incorporated into a time varying GARCH parameter via a MIDAS type filter. We find significant time variation in the GARCH parameter over the business cycle with elevated volatility persistence during times of financial distress and recession periods. Volatility forecasts from the new models are evaluated in an application to value at risk and expected shortfall measurements.

CO1250: **Determination of vector error correction models in higher dimensions**
Presenter: **Melanie Schienle**, Karlsruhe Institute of Technology, Germany
Co-authors: Chong Liang
We provide a shrinkage type methodology which allows for simultaneous model selection and estimation of vector error correction models (VECM) in one step. Model determination is treated as a joint selection problem of cointegrating rank and autoregressive lags. We show consistency of the selection mechanism by the resulting Lasso-VECM estimator under sparsity in lags and cointegration relations. In contrast to existing two-step approaches based on information criteria, we also derive the asymptotic properties of the final estimator. Moreover, with only linear computational complexity, the procedure remains computationally tractable also for higher dimensions. We demonstrate the effectiveness of the proposed approach by a simulation study and an empirical application to recent CDS data after the financial crisis.

CC1168: **Inference for impulse response coefficients from multivariate fractionally integrated processes**
Presenter: **Richard Baillie**, Michigan State University, United States
Co-authors: George Kapetanios, Fotis Papailias
A multivariate system of fractionally integrated time series is considered and the most appropriate way for estimating Impulse Response (IRs) coefficients and their associated confidence intervals are investigated. It extends a recent univariate analysis, and uses a semi parametric, time domain estimator, based on a vector autoregression (VAR) approximation. Results are also derived for the orthogonalized estimated IRs which are generally more practically relevant. Simulation evidence strongly indicates the desirability of applying the Kilian small sample bias correction, which is found to improve the coverage accuracy of confidence intervals for IRs. The most appropriate order of the VAR turns out to be relevant for the lag length of the IR being estimated.

CO1660: **A global perspective on the Great Moderation-Great Recession interconnection September 2015**
Presenter: **Claudio Morana**, Universita di Milano-Bicocca, Italy
Co-authors: Fabio C Bagliano
A large-scale model of the global economy is used to investigate the structural determinants of the Great Moderation and the transition to the Great Recession (1986-2010). Beside the global economy perspective, the model presents the novel feature of a broad range of included financial variables and risk factors measures. The results point to the relevance of various mechanisms related to the global monetary policy stance (Great Deviation), financial institutions' risk taking behavior (Great Leveraging) and global imbalances (savings glut), in shaping aggregate fluctuations. There is contribution to the literature on early warning indicators, assessing the information content of risk factor innovations for the prediction of the timing and depth of the Great Recession and the early phase of the Euro area sovereign debt crisis.

CO366 Room MAL 632 ENERGY AND MACROECONOMICS Chair: Hilde Bjornland

CO1413: **Country-specific oil supply shocks and the global economy: A counterfactual analysis**
Presenter: **Kamiar Mohaddes**, University of Cambridge, United Kingdom
Co-authors: Hashem Pesaran
The global macroeconomic consequences of country-specific oil-supply shocks are investigated. Our contribution is both theoretical and empirical. On the theoretical side, we develop a model for the global oil market and integrate this within a compact quarterly model of the global economy to illustrate how our multi-country approach to modelling oil markets can be used to identify country-specific oil-supply shocks. On the empirical side, estimating the GVAR-Oil model for 27 countries/regions over the period 1979Q2 to 2013Q1, we show that the global economic implications of oil-supply shocks (due to, for instance, sanctions, wars, or natural disasters) vary considerably depending on which country is subject to the shock. In particular, we find that adverse shocks to Iranian oil output are neutralized in terms of their effects on the global economy (real outputs and financial markets) mainly due to an increase in Saudi Arabian oil production. In contrast, a negative shock to oil supply in Saudi Arabia leads to an immediate and permanent increase in oil prices, given that the loss in Saudi Arabian production is not compensated for by the other oil producers. As a result, a Saudi Arabian oil supply shock has significant adverse effects for the global economy with real GDP falling in both advanced and emerging economies, and large losses in real equity prices worldwide.

CO1412: **Supply elasticity and drilling technology: Evidence from North Dakota**
Presenter: **Frode Martin Nordvik**, BI Norwegian Business School, Norway
Co-authors: Hilde Bjornland
Hydraulic fracturing technology has led to an extensive boom in US oil production, and in recent months oil prices have plunged partly as a result of an oil market supply glut. We ask whether oil production from the new fractured wells are more responsive to oil price changes than output from conventional wells. Using monthly production data for more than 16,000 crude oil wells in North Dakota, we find that production from fracked wells responds more than twice as strongly to oil price changes as conventional wells. Furthermore, the supply response of oil wells in general depends positively on the size of the well's operator firm. The results suggest that, as the share of unconventional oil grows, oil production could become more price sensitive in the short run.

CO1417: **Oil news shocks, OPEC response and the macroeconomy**
Presenter: **Knut Are Aastveit**, Norges Bank, Norway
Co-authors: Rabah Arezki , Akito Matsumoto
In the fall of 2014 oil prices dropped with 20USD in 7 days. It has been emphasized that changes in the demand for oil have been the main contributor for large oil price movements the last two decades, as opposed to the 1970s where changes on the supply side of the market played a large role. The recent large fluctuation in oil prices seems to suggest that we have entered a new phase where supply factors are again becoming more important. We explore the main drivers of oil prices using a SVAR model identifying 5 structural shocks to the global oil market. The novelty of our study is twofold. First, we distinguish between supply shocks stemming from OPEC and non-OPEC producers. This deviates from the standard assumption in the literature, focusing on the global production of oil as a whole, and thus assuming perfect competition in the oil market. The market power of OPEC clearly introduces a distortion in the oil market that potentially may affect oil prices and the global macroeconomy. In our model, OPEC is reacting to market conditions by systematically adjusting its production in order to affect the price of oil by utilizing market power. Second, while the current literature has focused on contemporaneous shocks to the oil market we aim to enrich our understanding by

allowing future prices or future conditions in the oil market to play a role. We distinguish between shocks to the spot price of oil and shocks that contain news about future conditions in the oil market.

CO1422: **Forecasting the Brent oil price: Addressing time-variation in forecast performance**
Presenter: **Ine Van Robays**, European Central Bank, Germany
In a policy environment, it is crucial to rely on forecast tools that generate projections whose forecast performance is stable over time. With the goal of finding such a tool for Brent oil price forecasting, a range of different forecast methods are explored, and their forecast performance relative to oil futures and the random walk is analyzed. Remarkably, none of the considered methods manages to outperform either benchmark consistently over time or across forecast horizons. To address this instability in forecast performance, we propose a forecast combination approach for predicting quarterly real Brent oil prices. We find that a four-model combination - consisting of futures, risk-adjusted futures, a Bayesian VAR and a DGSE model of the oil market - predicts Brent oil prices more accurately compared to all methods evaluated up to 3 years ahead on average, and generates a forecast whose performance is more robust over time. Moreover, the model combination reduces the forecast bias and predicts the direction of the oil price change more accurately. Compared to the gains documented in the literature, the improvements are noticeable.

CG019 Room MAL 540 CONTRIBUTIONS ON BOOTSTRAP INFERENCE Chair: Russell Davidson

CC0882: **On the validity of the pairs bootstrap for lasso estimators**
Presenter: **Lorenzo Camponovo**, University of St Gallen, Switzerland
We study the validity of the pairs bootstrap for lasso estimators in linear regression models with random covariates and heteroscedastic error terms. We show that the naive pairs bootstrap does not provide a valid method for approximating the distribution of the lasso estimator. To overcome this problem, we introduce a modified pairs bootstrap procedure and prove its consistency. Finally, we consider also the adaptive lasso, and show that the modified pairs bootstrap consistently estimates the distribution of the adaptive lasso estimator.

CC1409: **Bootstrapping integrated covariance matrix estimators in noisy jump-diffusion models with non-synchronous trading**
Presenter: **Ulrich Hounyo**, Aarhus University, Denmark
We propose a bootstrap method for estimating the distribution (and functionals of it such as the variance) of various integrated covariance matrix estimators. We show the first-order asymptotic validity of the wild blocks of blocks bootstrap in the multivariate context with a potential presence of jumps, dependent microstructure noise, irregularly spaced and non-synchronous data. Our results justify using the bootstrap to estimate the covariance matrix of a broad class of covolatility estimators. The bootstrap variance estimator is positive semi-definite by construction, an appealing feature that is not always shared by existing variance estimators of the integrated covariance estimator. As an application of our results, we also consider the bootstrap for regression coefficients. We show that the wild blocks of blocks bootstrap, appropriately centered, is able to mimic both the dependence and heterogeneity of the scores. We provide a proof of construction of bootstrap percentile and percentile-*t* intervals as well as variance estimates in this context. This contrasts with the traditional pairs bootstrap which is not able to mimic the score heterogeneity even in the simple case where no microstructure noise is present. Our Monte Carlo simulations show that the wild blocks of blocks bootstrap improves the finite sample properties of the alternative approach based on the Gaussian approximation. We illustrate its practical use on high-frequency equity data.

CC1627: **Validity of Edgeworth expansions for realized volatility estimators**
Presenter: **Bezirgen Veliyev**, Aarhus University, Denmark
Co-authors: Ulrich Hounyo
The main contribution is to establish the formal validity of Edgeworth expansions for realized volatility estimators. First, in the context of no microstructure effects, our results rigorously justify the Edgeworth expansion for realized volatility that was previously done in the literature. Second, we show that the validity of the Edgeworth expansions for realized volatility may not cover an optimal two-point distribution wild bootstrap previously proposed. Then, we propose a new optimal nonlattice distribution which ensures the second-order correctness of the bootstrap. Third, in the presence of microstructure noise, based on our Edgeworth expansions, we show that the new optimal choice proposed in the absence of noise is still valid in noisy data for the pre-averaged realized volatility estimator. Finally, we show how confidence intervals for integrated volatility can be constructed using these Edgeworth expansions for noisy data. Our Monte Carlo simulations show that the intervals based on the Edgeworth corrections have improved the finite sample properties relatively to the conventional intervals based on the normal approximation.

CC1031: **Generating prediction bands for path forecasts from SETAR models**
Presenter: **Daniel Grabowski**, Universitaet Giessen, Germany
Co-authors: Peter Winker, Anna Staszewska-Bystrova
SETAR models are increasingly used in time series analysis and forecasting. Prediction bands are usually generated point-wise, i.e. separately for each horizon, by bootstrap methods. Methods for constructing joint prediction bands for forecast paths obtained from SETAR models are proposed. Both approaches based on statistical theory and explicit sequential and global optimization methods are considered. Monte Carlo simulation is used to assess the performance of the proposed methods. The comparison is done with regard to the actual coverage of the constructed prediction bands for full forecast paths as well as with regard to the size of the bands. An empirical application demonstrates the relevance of the choice of bands for indicating the uncertainty of forecast paths in nonlinear models.

EO044 Room Holden BIG DATA ANALYSIS: STATISTICAL METHODOLOGIES AND APPLICATIONS Chair: Abdulkadir Hussein

EO0169: **Shrinkage estimation in SAR spatial models**
Presenter: **Abdulkadir Hussein**, University of Windsor, Canada
The spatial simultaneous autoregressive models (SAR) have been applied in many scientific areas such as plant species richness, cancer mortality rates, image processing, analysis of fMRI data and many more. We propose some efficient estimators for the regression coefficients of the SAR model based on combination of estimators from full and reduced models via shrinkage strategies and on absolute penalty estimators. We compare the array of the proposed estimators in terms of their asymptotic and simulated efficiency. We apply the proposed methods to Boston housing prices data and illustrate how a bootstrapping approach can be employed to compute prediction errors of the estimators.

EO0228: **Post-model selection robust inference using empirical divergence statistics in ultra-high dimensional problems**
Presenter: **Anand Vidyashankar**, George Mason University, United States
High dimensional data are ubiquitous in contemporary science and regression type models are typically used to address related inferential questions. A strategy, typically adopted by practicing statisticians is to first perform "exploratory analyses" and use model selection criteria such as BIC/GCV to select an appropriate model. The chosen model then gets treated as a "true model" and further inferences are performed. More recently, methods such as LASSO/ALASSO/SCAD/MCP and their variants also get used towards simultaneous variable selection and inference. It is now folklore that such a strategy towards inference may lead to inaccurate standard errors for the estimates of the regression parameters, potentially leading to erroneous decision making; also, additional complications arise if the underlying statistical model is misspecified. We provide a new framework,

using divergences and empirical likelihood, to assess model selection variability and identify two important sub-components: namely, intrinsic and extrinsic uncertainty. We evaluate the effect of these sub-components on the robustness and efficiency of inference.

EO0282: **Gateaux differential-based variable selection for analysis of gene-gene interactions**
Presenter: **Yi Li**, University of Michigan, United States

Identifying gene-gene interactions is fundamentally important to clarify genetic pathways. Three fundamental principles can be applied: 1) symmetric hierarchy; when an interaction is selected, the lower-order effect should also be present, 2) asymmetric hierarchy; in order to select an interaction, at least one of its parent lower-order effects should also be in the model, and 3) epistasis; interactions can be selected without the presence of their main effects. For complex diseases such as cancers, a variety of genetic alterations interact with each other, and important interactions may not pass lower-order examination. Therefore, asymmetric hierarchy or epistasis is often desirable. However, as the number of potential interactions is ultrahigh, most existing statistical methods lack sufficient computation power. To address this problem, we first proposed a grouped Gateaux differential-based boosting which can be applied to adapt asymmetric hierarchy. The proposed method has the advantage that the candidate set for variable selection is dynamic, which enforces hierarchy. To add more flexibility, we generalized our proposed method with Gateaux differential-based screening to consider epistasis. The proposed algorithm can be shown to take the group LARS as special cases. It is more flexible in that it can be applied to loss functions that are more complex than linear least squares and avoids some limitations inherited with the existing constrained optimization methods.

EO0360: **Use of latent class models to identify an early placebo responder subgroup with matrix-valued EEG measurements**
Presenter: **Bei Jiang**, University of Alberta, Canada

Latent class models are widely used to identify latent subgroups based upon one or more manifest variables. The probability of belonging to each subgroup can be simultaneously related to a set of measured covariates. We extend existing latent class models to incorporate matrix covariates. This research is motivated by a placebo-controlled depression clinical trial. The proposed method is built upon a low rank CANDECOMP/PARAFAC (CP) decomposition to express the target coefficient matrix through low-dimensional latent variables, which effectively reduces the model dimensionality, and utilizes a Bayesian hierarchical modeling approach to estimating these latent variables, which provides a flexible way to incorporate prior knowledge on the patterns of covariate effect heterogeneity and provides a data-driven method of regularization. We develop an efficient Markov chain Monte Carlo algorithm for posterior computation and study the properties of the method via simulation studies. Finally in our motivating example, we show that the proposed method allows us to extract useful information from baseline EEG measurements that explains the likelihood of belonging to the placebo responder subgroup.

EO320 Room Senate INFLUENCE AND ROBUSTNESS IN STATISTICAL MODELING Chair: Andreas Alfons

EO0179: **Influence measures and stability for graphical models**
Presenter: **Jean-Michel Poggi**, University Paris-Sud Orsay, France
Co-authors: Avner Bar-Hen

Graphical models allow to represent a set of random variables together with their probabilistic conditional dependencies. Various algorithms have been proposed to estimate such models from data. Our first focus is on individual observations diagnosis issues. The use of an influence measure is a classical diagnostic method to measure the perturbation induced by a single element, in other terms we consider stability issue through jackknife. For a given graphical model, we provide tools to perform diagnosis on observations. In a second step we propose a filtering of the dataset to obtain a stable network. Finally we propose to use our influential measure to assess the relevance of the partition induced by a clustering. An application to a gene expression dataset illustrates the proposals.

EO1212: **Robust regression in high dimensions: Sparse S- and MM-estimation**
Presenter: **Andreas Alfons**, Erasmus University Rotterdam, Netherlands
Co-authors: Viktoria Oellerer, Christophe Croux

The S-estimator and the MM-estimator are frequently used regression estimators that are robust against the presence of outliers. While both estimators are highly robust, the latter also attains high efficiency. A drawback of those methods is that they cannot be applied to high-dimensional data, i.e., data with more variables than observations. By adding an L_1 penalty on the coefficient estimates to the respective objective functions, the sparse S-estimator and the sparse MM-estimator are introduced. These new estimators combine robust regression with sparse model estimation. In addition to deriving the breakdown point and the influence function, the performance of sparse S and sparse MM is assessed by means of a simulation study.

EC1751: **Influence functions for penalized M-estimators**
Presenter: **Marco Avella Medina**, University of Geneva, Switzerland

We study the local robustness properties of general non-differentiable penalized M-estimators via the influence function. More precisely, we propose a framework that allows us to define rigorously the influence function as the limiting influence function of a sequence of approximating M-estimators. We show that it can be used to characterize the robustness properties of a wide range of sparse estimators and we derive its form for general penalized M-estimators including lasso and adaptive lasso type estimators. We prove that our influence function is equivalent to a derivative in the sense of distribution theory.

EC0566: **Robust lasso regression using the Tukey's biweight criterion**
Presenter: **Le Chang**, the Australian National University, Australia
Co-authors: Steven Roberts, Alan Welsh

The adaptive lasso is a common means of performing simultaneous parameter estimation and variable selection. The adaptive weights used in its penalty term mean that the adaptive lasso achieves the desirable oracle property. We propose an extension of the adaptive lasso called the Tukey-lasso. By using Tukey's biweight criterion, instead of squared loss, the Tukey-lasso is resistant to outliers in both the response and covariates. Importantly, like the adaptive lasso, the Tukey-lasso also enjoys the oracle property. The Tukey-lasso is compared to various competitors and extensive simulation studies show that it offers significant improvements in performance compared to the traditional adaptive lasso and other robust implementations of the lasso in the presence of outliers. Real data examples further demonstrate the utility of the Tukey-lasso.

EO164 Room Court BAYESIAN SEMI- AND NONPARAMETRIC MODELLING III Chair: Antonio Lijoi

EO0180: **Bayesian inference for intra-tumor heterogeneity in mutations and copy number variation**
Presenter: **Juhee Lee**, University of California Santa Cruz, United States
Co-authors: Peter Mueller, Yuan Ji, Subhajit Sengupta, Kamalakar Gulukota
Tissue samples from the same tumor are heterogeneous. They consist of different subclones that can be characterized by differences in DNA nucleotide sequences and copy numbers on multiple loci. Inference on tumor heterogeneity involves the identification of the subclonal copy number and single nucleotide mutations at a selected set of loci. We estimate such tumor heterogeneity on the basis of a Bayesian feature allocation model. We jointly model subclonal copy numbers and the corresponding allele sequences for the same loci. The proposed method utilizes three random matrices, L, Z and w to represent subclonal copy numbers (L), the number of subclonal variant alleles (Z) and the cellular fractions (w) of subclones in one or more tumor samples, respectively. The unknown number of subclones implies a random number of columns for these matrices and having more than one subclone indicates tumor heterogeneity. We estimate the subclonal structures through inference on these three matrices, using next-generation sequencing data. Using simulation studies and a real data analysis, we demonstrate how posterior inference on the subclonal structure is enhanced with the joint modeling of both structure and sequencing variants on subclonal genomes. R package is available at http://cran.rproject.org/web/packages/ BayClone2/index.html.

EO0927: **Bayesian ordination of species sampling from dependent populations with applications to microbiome studies**
Presenter: **Sergio Bacallado**, Cambridge University, United States
Co-authors: Boyu Ren, Lorenzo Trippa, Stefano Favaro, Susan Holmes
We introduce a Bayesian nonparametric prior for sampling from many discrete distributions. The distributions are correlated through continuous latent factors, whose dimensionality is shrunk by the prior. These factors can be used in exploratory data analysis, and we discuss the visualization of credible regions through multivariate analysis methods. It is also possible to include factors for fixed effects if the distributions are characterized by covariates. The marginal prior on each distribution is a normalized generalised Gamma process, a well-known nonparametric prior. We apply Markov chain Monte Carlo and variational Bayesian inference methods. The procedure is applied to the analysis of microbiome studies, which produce discrete samples from the distribution of bacterial species in several environments. The goal of these studies is to understand the variation of species abundance across environments and their relation to covariates. The analysis deals naturally with the uncertainty in the distribution of species, which is important when the sampling depth is low.

EO1096: **Multi-resolution scanning for distributional variation via the k-group design**
Presenter: **Li Ma**, Duke University, United States
Co-authors: Jacopo Soriano
An inference task that holds key to numerous applications is the comparison of multiple data sets to identify the underlying difference. The most common setting is the k-sample problem, in which the samples are collected under k different experimental settings or treatment/response statuses. A fundamental challenge in modern k-sample problems is the presence of many potential confounders, or extraneous sources of variation, that can contribute to the difference across the sampling distributions, but are difficult to control explicitly. This results in numerous irreplicable findings in many modern applications such as functional genomics. A simple experiment design that allows the intrinsic (i.e. scientifically interesting) variation to be identified from the extraneous ones is the k-group design, under which replicate data sets are collected under each experiment setting. A fully probabilistic multi-resolution scanning framework is presented based on local hierarchical Binomial testing for analyzing k-group studies that effectively identifies actual cross-group variation while adjusting for extraneous ones. It takes into account the spatial heterogeneity of cross-sample variation and is particularly effective for finding highly local cross-group variation. It is also highly computationally efficient, suitable for real-time applications.

EO0758: **On some exchangeable increment processes derived from Bayesian ideas**
Presenter: **Ramses Mena**, UNAM, Mexico
In several applied areas, such as finance and risk theory, the use of stochastic models with the property of having independent increments is predominant. This is due perhaps to the mathematical tractability such a property implies. However, some phenomena require other kind of dependence structure and thus models relaxing such assumption are required. We discuss some models with exchangeable increments that, while relaxing the independent increments assumption, remain mathematically and computationally tractable. Our proposals borrow some Bayesian ideas for their construction. In particular, some applications in risk theory and finance will be presented.

EO330 Room Chancellor's Hall COPULA MODELS AND APPLICATIONS Chair: Claudia Tarantola

EO0237: **Bayesian copula modelling in the presence of covariates**
Presenter: **Julian Stander**, Plymouth University, United Kingdom
Co-authors: Luciana Dalla Valle, Charlotte Taglioni, Mario Cortina Borja
Copula models separate the dependence structure in a multivariate distribution from its univariate marginals, so overcoming many of the issues associated with commonly used statistical modelling methods by allowing, for example, different complex asymmetric dependencies and tail behaviours to be modelled. We discuss the modelling of bivariate data using copulas, of which there are now a rich choice. The parameter or parameters of the copula density are modelled as a function of a covariate using a natural cubic spline. Working in the Bayesian framework, we perform inference on the natural cubic spline and an associated smoothing parameter. We also discuss the choice of the copula density itself. We illustrate our approach using data from finance and medicine. We outline the extension of our methodology to more than one covariate and to multivariate data.

EO0494: **Tree copula mixture distribution for multivariate dependence analysis: An application to energy data**
Presenter: **Federico Bassetti**, Univeristy of Pavia, Italy
Co-authors: Maria Elena De Giuli, Enrica Nicolino, Claudia Tarantola
Motivated by the study of the energy market, a Bayesian analysis for a panel AR(p) model is proposed where the multivariate distribution of the innovations is described by a mixture of tree copula distributions. We assume that the innovations of the AR(p) model have normal marginal distributions, without assuming their joint normality. The use of the copulas allows us to split the dependence structure from the marginal distributions. Hence, we can represent a multivariate distribution through its univariate marginal distributions and a copula that captures the dependence structure. Although in the multivariate case the Gaussian and Student$-t$ copulas are the most used, they often are not flexible enough to represent the dependence structure of the data. A possible solution is the use of tree copula: particular types of Markov tree distributions that allow to represent the multivariate joint density through a suitable set of bivariate densities. In particular, in a tree copula the bivariate distributions over the edges of the underlying tree structure are specified by bivariate copulas. In order to develop a more flexible model and to take into account more complex dependence structure, we consider finite mixture of tree copulas. We use a MCMC algorithm to obtain posterior distributions.

EO0543: **Systemic default risk attribution in EU area**
Presenter: **Gianluca Farina**, Bergamo, Italy
Co-authors: Maria Elena De Giuli, Rosella Giacometti
The aim is to provide a model for systemic default risk attribution in order to disentangle the different components. Systemic default risk is related to the possibility of a simultaneous default of multiple institutions. This risk has caused great concern in recent past, however its measure is not a trivial subject. We introduce a multivariate copula for all the countries in EU zone, providing an integrated analysis. We consider a Multivariate Generalized Marshall-Olkin copula to get the joint dynamic of the countries and the financial sectors in each country. The marginal probability of default of each country and each bank depends on default intensities hierarchically decomposed in EU-wide, country-specific and idiosyncratic effects. The model is calibrated using market data of sovereign and Banks CDSs. The systemic risk attribution among countries and banks is finally proposed.

EC1558: **Modelling the dependence structure between international financial returns using regime switching vine copula**
Presenter: **Alex Donov**, University of Essex, United Kingdom
The dependence structure in international financial returns is modeled using second-order regime switching vine copula. Gaussian copula is used as a main building block in the vine specification where the dependence parameters are held constant within each regime. The model is applied to returns of the S&P500, FTSE100 and DAX stock indices. The standard errors of the estimates are computed using Godambe information matrix. Empirically, regimes of high dependence are identified. The model is then compared against the benchmark model in which regime variable follows first-order process. Information criteria such as AIC and BIC suggest that the second-order regime switching model may be a good choice.

EO066 Room CLO B01 NON- AND SEMI-PARAMETRIC FUNCTIONAL STATISTICS I Chair: Enea Bongiorno

EO0257: **A longitudinal functional data analysis for change of daily physical activity patterns**
Presenter: **Oliver Yibing Chen**, The Johns Hopkins Bloomberg School of Public Health, United States
Co-authors: Luo Xiao, Martin Lindquist, Brian Caffo, Jennifer Schrack, Luigi Ferrucci, Ciprian Crainiceanu
Objective measurement of physical activity using wearable devices such as accelerometers provides tantalizing new insights into the association between physical activity and health outcomes. Accelerometers record quasi-continuous activity information for many days and for hundreds of individuals. For example, in the Baltimore Longitudinal Study on Aging, daily physical activity was recorded for about 300 adults during each visit for several days. Each subject has two to four visits. An interesting problem that naturally arises is how to quantify daily physical activity patterns change with age, gender, body mass index, among other covariates. We propose a longitudinal functional data model where the parameters of interest are bivariate functions of time and age. To deal with the complex correlation structure in the data, we use a GEE-type approach for model estimation. For efficient parameters and covariance estimation, we introduce a two-step procedure. Our results reveal several interesting and previously unknown daily activity patterns associated with human aging.

EO0689: **Tests for separability in nonparametric covariance operators of random surfaces**
Presenter: **Shahin Tavakoli**, University of Cambridge, United Kingdom
Co-authors: John Aston, Davide Pigoli
The assumption of separability of the covariance operator for a random image or hypersurface can be of substantial use in applications, especially in situations where the accurate estimation of the full covariance structure is unfeasible, either for computational reasons or due to a small sample size. However, inferential tools to verify this assumption are somewhat lacking in high-dimensional or functional settings where this assumption is most relevant. We propose to test separability by focusing on K-dimensional projections of the difference between the covariance operator and its nonparametric separable approximation. The subspace we project onto is one generated by the eigenfunctions estimated under the separability hypothesis, negating the need to ever estimate the full non-separable covariance. We show that the rescaled difference of the sample covariance operator with its separable approximation is asymptotically Gaussian. As a by-product of this result, we derive asymptotically pivotal tests under Gaussian assumptions, and propose bootstrap methods for approximating the distribution of the test statistics when multiple eigendirections are taken into account. We probe the finite sample performance through simulations studies, and present an application to log-spectrogram images from a phonetic linguistics dataset.

EO1116: **Regression models with categorical functional data**
Presenter: **Cristian Preda**, University of Lille, France
Regression models based on RKHS methods are used to estimate the regression function for scalar response and categorical functional predictor. A simulation study based on paths of a Markov jump process with finite set of states will illustrate the proposed methodology.

EO1447: **Empirical evolution equations**
Presenter: **Susan Wei**, EPFL, Switzerland
Co-authors: Victor Panaretos
Noisy vector fields arise in many disciplines of study. One of the most fundamental operations on a vector field is its integration. We derive conditions under which the integral curve of a noisy vector field is asymptotically normal. Results in this vein have previously been obtained in the specific contexts of diffusion tensor imaging in neuroimaging and filament estimation in astrophysics. The theory we build decouples the analysis from the specific vector field estimator used. Numerous illustrative examples will be given to demonstrate the wide-ranging applicability of our results.

EO162 Room Beveridge Hall ANALYSIS OF TOROIDAL AND CYLINDRICAL DATA Chair: Christophe Ley

EO0323: **Adapting the Bernstein copula to modeling cylindrical and spherical data with applications**
Presenter: **Michael Wiper**, Universidad Carlos III de Madrid, Spain
Co-authors: Concepcion Ausin
A nonparametric approach is proposed for estimating the dependence relationships between circular variables and other circular or linear variables using copulas. The proposed method is based on the use of Bernstein copulas which allow for the approximation of any kind of dependence structure, including asymmetric relationships. We present a simple procedure to adapt Bernstein copulas to the circular framework and guarantee that the constructed bivariate distributions are strictly continuous. Applications to wind direction data are provided.

EO0402: **A five-parameter bivariate wrapped Cauchy model for toroidal data**
Presenter: **Arthur Pewsey**, University of Extremadura, Spain
Toroidal data, consisting of pairs of circular observations, arise in numerous scientific disciplines. We propose a five-parameter bivariate wrapped Cauchy model which is unimodal, pointwise symmetric and has a closed-form expression for the mode. Two of its parameters position the mode of the distribution, and a single dependence parameter controls the relationship between the two component circular variables with concentrations regulated by the remaining two parameters. The distribution arises as a submodel of a six-parameter model obtained by applying Mobius transfor-

mation to an existing bivariate circular distribution. Two other derivations of it are available: one, as a maximum entropy distribution, and the other as a special case of the popular Wehrly-Johnson construction. Its trigonometric moments can be expressed in closed form, as can three correlation coefficients specifically designed for use with toroidal data. Method of moments and likelihood-based inference are both computationally very efficient, and bootstrap tests for independence and goodness-of-fit are available. The model also provides the basis for a new Markov process for circular data and a distance measure on the torus. An analysis involving dihedral angles of the proteinogenic amino acid Tyrosine illustrates the application of the model and the inferential techniques developed for it.

EO0467: **Testing parametric models in linear-directional regression**
Presenter: **Ingrid Van Keilegom**, Universite catholique de Louvain, Belgium
Co-authors: Eduardo Garcia-Portugues, Wenceslao Gonzalez-Manteiga, Rosa Crujeiras
The aim is to present a goodness-of-fit test for parametric regression models with scalar response and directional predictor, that is, vectors in a sphere of arbitrary dimension. The testing procedure is based on the weighted squared distance between a smooth and a parametric regression estimator, where the smooth regression estimator is obtained by a projected local approach. Asymptotic behavior of the test statistic under the null hypothesis and local alternatives is provided, jointly with a consistent bootstrap algorithm for application in practice. A simulation study illustrates the performance of the test in finite samples. The procedure is also applied to a real data example from text mining.

EO0592: **Length dependent models for cylindrical data through sine skewed wrapped Cauchy distribution**
Presenter: **Toshihiro Abe**, Nanzan University, Japan
There has been an increased interest towards to the cylindrical distributions in recent years. The combination of direction and length, such as a pair of wind direction and its strength, can be regarded as the cylindrical data. Although some cylindrical models are well known, there are a few cylindrical models compared with the circular models. We propose another class of cylindrical distributions using circular distributions with certain dependence structure. The model has the simple normalizing constant. As a special case, we propose a cylindrical distribution whose conditional distribution is the sine-skewed wrapped Cauchy distribution. The advantages of such the selection are a general form of the trigonometric moments, simple generation of random number and explicit form of the mode.

EO172 Room Woburn STATISTICAL INFERENCE AND APPLICATION IN FUNCTIONAL ANALYSIS Chair: Catherine Liu

EO0373: **Functional factor analysis of spatial-temporal data**
Presenter: **Chi Tim Ng**, Chonnam National University, Korea, South
The aim is to investigate the likelihood inferences for functional factor analysis of high-dimensional multivariate time series data. A matrix decomposition technique is developed to obtain expressions of the likelihood functions and its derivatives. With such expressions, the traditional delta method that relies heavily on score function and Hessian matrix can be extended to high dimensional cases. Moreover, fast computational algorithms are developed for estimation. Applications to climate data and financial data are discussed.

EO0375: **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 nonivertibility 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 the asymptotically perfect classification can be achieved under certain circumstances. The performance of this new approach is further demonstrated with numerical examples.

EO0901: **Testing equality of covariance operators/matrices for functional/high-dimensional data**
Presenter: **Catherine Liu**, The Hong Kong polytechnic University, China
Co-authors: Jin Yang, Tao Zhang
The purpose is twofold. The first goal is to propose a unified methodology for testing equality of covariance operators of two functional samples whatever the type of the functional data is dense or sparse, balanced or irregularly spaced. A two-step procedure is developed which leads to a global testing statistic. The second goal is, from the insight of functional data analysis, to present a novel method to test equality of covariance matrices of two high-dimensional samples. It might be an inchoate but inspiring trying to apply the idea of functional data analysis into high-dimensional data study. Under null and alternative hypotheses, asymptotic distributions of the testing statistics have been derived for afore two types of data. Extensive simulation experiments have been conducted demonstrating consistency of the tests for both data scenarios. In finite sample numerical analysis, the proposed approaches outperform existing work in terms of both size and power for equality testing problems for both functional data and high-dimensional data. Two applications are provided: Air pollution data in Southwestern area of China is analyzed to illustrate our procedure of testing equality of covariance operators for functional data samples; Mitochondrial calcium concentration data is analyzed to demonstrate how our proposed method can be applied to test equality of covariance matrices for high-dimensional data.

EO0489: **Frechet regression**
Presenter: **Alexander Petersen**, University of California Davis, United States
Co-authors: Hans-Georg Mueller
The regression of a response variable Y on a vector-valued predictor X is considered, where Y is situated in a metric space with metric d that is not necessarily a linear space. A global regression relationship is defined as a direct generalization of ordinary linear regression. This global regression can be expressed in terms of weighted Frechet means with respect to the metric d. The convergence of the corresponding sample estimates is studied and a generalized R^2 coefficient is proposed as a measure of the strength of the relationship that also serves as a test statistic for testing the null hypothesis of no relationship between predictor and response. The methods are illustrated with some nonlinear functional data from brain imaging.

EO194 Room Bedford POPULATION MODELS: METHODS AND COMPUTATIONS I Chair: Manuel Molina

EO0551: **Robust Bayesian estimation on controlled branching processes: The disparity approach**
Presenter: **Carmen Minuesa Abril**, University of Extremadura, Spain
Co-authors: Miguel Gonzalez Velasco, Ines Maria del Puerto Garcia

Controlled branching processes are suitable probabilistic models for the description of population dynamics in which the number of individuals with reproductive capacity in each generation is controlled by a random control function. The probabilistic theory of these processes has been extensively developed, mainly focused on the extinction problem and the limiting behaviour. Nowadays, an important issue is to study the inferential problems arising from this model. The purpose is to develop a robust procedure to obtain estimators for the main parameters of the model from a Bayesian outlook assuming that the offspring distribution belongs to a very general one-dimensional parametric family. We provide robust Bayesian estimators of the offspring parameter through the use of disparities by considering different samples. For each sample, the method consists in replacing the log likelihood with an appropriately scaled disparity in the expression of the posterior distribution. We study the asymptotic and robustness properties of the estimators associated with the resulting distributions. Finally, we illustrate the accuracy of the proposed methods by simulated examples developed with the statistical software R.

EO0736: **Simulating and forecasting human population with general branching process**
Presenter: **Plamen Trayanov**, Sofia University, Bulgaria

The theory of General Branching Processes (GBP) presents a continuous time model in which every woman has random life length and gives birth to children in random intervals of time. This flexibility of the GBP makes it very useful for modelling and forecasting human population. An application of the GBP is presented in forecasting the population age structure of several European countries. Confidence intervals of the forecasts are calculated using simulations which reflect both the stochastic nature of the birth and death laws and the branching process itself. Another application of simulations is to determine the main sources of risk to the forecast. The GBP model allows us to obtain knowledge on the processes happening in those countries and the simulations allow us to obtain insight for their future.

EO0947: **Algorithm for modeling of branching processes in continuous time as models of mutations**
Presenter: **Maroussia Slavtchova-Bojkova**, Sofia University, Bulgaria
Co-authors: Plamen Trayanov, Dimitrov Stoyan

Branching processes in continuous time are both powerful tools for studying the mutations in cell populations and useful models of the dynamics of the number of different types of cells, which due to a small reproductive ratio are fated to become extinct. However, mutations occurring during the reproduction process, may lead to the appearance of a new type of cells that may escape extinction. We are deriving theoretically the numbers of mutations of the escape type and their moments. A cell of the mutation type, which leads possibly to the beginning of a lineage, that will never become extinct is called successful mutant. Using the results about the probability generating function of the single type branching processes, the distribution of the waiting time to produce a successful mutant in continuous time setting is obtained and the time of attaining high levels is studied. Simulation code is developed for deriving the distribution of these quantities. A numerical approach for obtaining the distribution of the time to appearance of successful mutant is developed.

EO0699: **Quantifying stochastic introgression processes with hazard rates**
Presenter: **Maria Conceicao Serra**, Minho Universitu, Portugal

Introgression is the permanent incorporation of genes from one population into another through hybridization and backcrossing. It is currently of particular concern as a possible mechanism for the spread of modified crop genes to wild populations. The hazard rate is the probability, per time unit, that such an escape takes place, given that it has not happened before. It is a quantitative measure of introgression risk that takes the stochastic elements inherent in introgression processes into account. We use multitype branching processes to model the evolution of populations that are exposed to introgression events. We present a methodology to calculate the hazard rate for situations with: i) constant; ii) time-varying; iii) random; gene flow from a crop to a large recipient wild population. We analyze the effect that conditions i)-iii) have on the behaviour of the hazard rate, namely on the long term behaviour. Finally, we discuss the risk of introgression in meta-populations.

EO621 Room Gordon COVARIANCE, VOLATILITY, AND EFFICIENT INFERENCE Chair: Roy Welsch

EO0563: **Linear response methods for accurate covariance estimates from mean field variational Bayes**
Presenter: **Tamara Broderick**, MIT, United States
Co-authors: Ryan Giordano, Michael Jordan

Mean field variational Bayes (MFVB) is a popular posterior approximation method due to its fast runtime on large-scale data sets. However, it is well known that a major failing of MFVB is that it underestimates the uncertainty of model variables (sometimes severely) and provides no information about model variable covariance. We generalize linear response methods from statistical physics to deliver accurate uncertainty estimates for model variables - both for individual variables and coherently across variables. We call our method linear response variational Bayes (LRVB). When the MFVB posterior approximation is in the exponential family, LRVB has a simple, analytic form, even for non-conjugate models. Indeed, we make no assumptions about the form of the true posterior. We demonstrate the accuracy and scalability of our method on a range of models for both simulated and real data.

EO0745: **Measuring correlation in the presence of spikes**
Presenter: **Debbie Dupuis**, HEC Montreal, Canada

For electricity markets with multiple zones and local pricing, an understanding of electricity price dynamics is crucial for long-term planning. But electricity prices are notorious for the frequent occurrence of large price spikes. Looking at one time series of electricity price data, some have sought to identify spikes, forecast spikes, or explain spikes. We rather examine multiple time series and seek to estimate the correlation between two series in the presence of these spikes. More robust estimators than the usual Pearson correlation exist, e.g. Spearman's rho or Kendall's tau. The time-varying nature of the correlations must be considered however. We calculate Spearman's rho over moving windows and apply a recently developed exponentially-weighted Kendall's tau. We develop and investigate a robust exponentially weighted correlation estimator that performs well in the presence of electricity price spikes and can better track a rapidly-changing time-varying correlation. A simulation study shows how traditional correlation estimators fail in the presence of spikes and compares all the robust alternatives in the context of time-varying correlation. An application to real electricity markets price data demonstrates the need for the new robust estimator.

EO1201: **Fast inference with submodular functions**
Presenter: **Stefanie Jegelka**, MIT, United States

Submodularity - also known as a discrete analogue of convexity - has been an important concept in areas like graph theory, game theory, and combinatorial optimization. Submodular functions also arise in various forms in statistics in machine learning. They may, for example, define probabilities in discrete graphical models and point processes, or structured norms for M-estimation. Submodularity will be introduced and recent

work will be highlighted on exploiting properties of submodular functions and related polyhedra to solve inference problems, focusing in particular on the case of having a sum of submodular functions.

EO1106: **Structural dynamic analysis of systematic risk**
Presenter: **Veronika Czellar**, EDHEC, France
Co-authors: Laurent Calvet, Christian Gourieroux
A structural dynamic factor model (SDFM) for stock prices is introduced. Compared to standard linear factor models, structural modeling accounts for nonlinear effects of common factors when the distance-to-default is small. Such a SDFM has a rather complicated form, especially since the underlying factors are unobservable. We develop appropriate methods to estimate such models and use them for prediction and filtering purposes, such as indirect inference or Approximate Bayesian Computation (ABC) filtering. These SDFM are applied to the analysis of systematic risk of financial institutions and intermediaries.

EO292 Room Montague DYNAMIC FACTOR MODELS AND SUFFICIENT DIMENSION REDUCTION Chair: Efstathia Bura

EO0730: **Envelopes and partial least squares regression**
Presenter: **Zhihua Su**, University of Florida, United States
Co-authors: Dennis Cook, Inge Helland
The aim is to build connections between envelopes, a recently proposed context for efficient estimation in multivariate statistics, and multivariate partial least squares (PLS) regression. In particular, we establish an envelope as the nucleus of both univariate and multivariate PLS, which opens the door to pursuing the same goals as PLS but using different envelope estimators. It is argued that a likelihood based envelope estimator is less sensitive to the number of PLS components selected and that it outperforms PLS in prediction and estimation.

EO1148: **On the level, slope, curvature effect in yield curves and eventual total positivity**
Presenter: **Carlos Tolmasky**, University of Minnesota, United States
Principal components analysis has become widely used in a variety of fields, e.g., in finance and, more specifically, in the theory of interest rate derivative modeling. The key finding in this area was that a few components explain most of the variance of treasury zero-coupon rates and that the first three eigenvectors represent level, slope and curvature (LSC) changes on the curve. This result has been, since then, observed in various markets. Over the years, there have been several attempts at modeling correlation matrices displaying the observed effects as well as trying to understand what properties of the those matrices are responsible for the effect. Using recent results of the theory of total positiveness we characterize these matrices and, as an application, we shed light on the critique to the methodology previously raised.

EO1488: **Nowcasting business cycles: A Bayesian approach to dynamic heterogeneous factor models**
Presenter: **Antonello DAgostino**, European Stability Mechanism, Luxembourg
Co-authors: Domenico Giannone, Michele Lenza, Michele Modugno
A framework for measuring and monitoring business cycles in real time is developed. Following a long tradition in macroeconometrics, inference is based on a variety of indicators of economic activity, treated as imperfect measures of an underlying index of business cycle conditions. We extend existing approaches by permitting for heterogenous lead-lag patterns of the various indicators along the business cycles. The framework is well suited for high-frequency monitoring of current economic conditions in real time - nowcasting - since inference can be conducted in presence of mixed frequency data and irregular patterns of data availability. Our assessment of the underlying index of business cycle conditions is accurate and more timely than popular alternatives, including the Chicago Fed National Activity Index (CFNAI). A formal real-time forecasting evaluation shows that the framework produces well-calibrated probability nowcasts that resemble the consensus assessment of the Survey of Professional Forecasters.

EC1764: **Forecasting with sufficient dimension reductions**
Presenter: **Efstathia Bura**, The George Washington University, United States
Co-authors: Alessandro Barbarino
Factor models have been successfully employed in summarizing large data sets with few underlying latent factors and in building time series forecasting models for economic variables. When the objective is to forecast a target variable y with a large set of predictors x, the construction of the summary of the x's should be driven by how informative on y it is. Most existing methods first reduce the predictors and then forecast y in independent phases of the modeling process. We present a potentially more attractive alternative: summarizing x as it relates to y, so that all the information in the conditional distribution of y given x is preserved. These y-targeted reductions of the predictors are obtained using Sufficient Dimension Reduction techniques. We show in simulations and real data analysis that forecasting models based on sufficient reductions have the potential of significantly improved performance.

EO629 Room Jessel MIXTURE MODEL AND VARIABLE SELECTION Chair: Ryan Browne

EO1128: **Implicit versus explicit variable selection**
Presenter: **Paul McNicholas**, McMaster University, Canada
Many approaches have been put forth for variable selection in mixture model-based clustering. Broadly, these approaches can be categorized as implicit or explicit, with the categories differing on whether or not variables are explicitly selected. After some discussion of popular implicit and explicit approaches, the two categories are compared and contrasted.

EO0886: **Variable selection for clustering**
Presenter: **Jeffrey Andrews**, University of British Columbia, Canada
Variable selection under a clustering paradigm is discussed, highlighting recent advances in the literature and available software. The VSCC (variable selection for clustering and classification) method looks for variables which minimize the within group variation, while simultaneously ensuring that redundant features are removed. The technique and R software are introduced, focusing on some recent updates, and then comparisons with state of the art methods on a variety of data sets, both real and simulated, are shown.

EO0987: **Variable selection for latent class analysis**
Presenter: **Michael Fop**, University College Dublin, Ireland
A stepwise variable selection method for latent class cluster analysis is proposed. Given a set of already selected clustering variables, the usefulness of a variable is assessed comparing two models. In one model the variable adds further information about the clusters, and in the other model it does not but it can be related to the clustering variables. A stepwise greedy algorithm is used to conduct the search over the model space and the comparison between models is performed using BIC. The framework allows to discard both non-informative and redundant variables and to perform a parsimonious selection. An application on data related to low back pain disorders is considered and the method recovers the true group structure with a small number of variables.

EO1105: **Identification of key match-related variables in women's rugby sevens**
Presenter: **Paula Murray**, McMaster University, Canada
Co-authors: Paul McNicholas, Ryan Browne
In 2016, rugby sevens will make its debut as an Olympic sport at the summer Olympics in Rio de Janeiro, Brazil. Using data from the Canadian national women's rugby sevens team, we attempt to identify key variables related to winning and losing rugby sevens matches. We develop a variable selection method and present our algorithm for selection of match variables recorded by the International Rugby Board. We demonstrate our method on data from the 2014-2015 World Rugby Women's Sevens Series and compare our results with established variable selection techniques.

EO056 Room Bloomsbury SURVEY SAMPLING FOR FUNCTIONAL DATA AND/OR BIG DATA Chair: Patrice Bertail

EO1229: **SGD algorithms based on incomplete U-statistics: Large-scale minimization of empirical risk**
Presenter: **Guillaume Papa**, Telecom Paristech, France
Co-authors: Stephan Clemencon
In many learning problems, ranging from clustering to ranking through metric learning, empirical estimates of the risk functional consist of an average over tuples of observations. We argue that in the large-scale setting, the inductive principle of stochastic approximation for risk minimization should be implemented in a very specific manner. Precisely, the gradient estimates should be obtained by sampling tuples of data points with replacement (incomplete U-statistics) instead of sampling data points without replacement (complete U-statistics based on subsamples). We develop a theoretical framework accounting for the considerable impact of this strategy on the generalization ability of the prediction model returned by the Stochastic Gradient Descent (SGD) algorithm. It reveals that the method we promote achieves a much better trade-off between statistical accuracy and computational cost. Beyond the rate bound analysis, numerical experiments on AUC maximization and metric learning provide strong empirical evidence of the superiority of the proposed approach.

EO1305: **Survey schemes for stochastic gradient descent with applications to M-estimation**
Presenter: **Emilie Chautru**, MINES ParisTech, France
Co-authors: Stephan Clemencon, Patrice Bertail, Guillaume Papa
The main purpose is to investigate the impact of survey sampling with unequal inclusion probabilities on stochastic gradient descent-based M-estimation methods in large-scale statistical and machine-learning problems. Precisely, when possible, we propose to take advantage of some auxiliary information to increase asymptotic accuracy of the estimation. The method is discussed in the specific context of big data.

EO1819: **Scaling by subsampling for big data**
Presenter: **Melanie Zetlaoui**, Paris West University, France
The increasing capacity to collect data has improved much faster than our ability to process and analyze Big Datasets. The availability of massive information in the Big Data era suggests to use subsampling techniques as a remedy to the apparent intractability of learning from datasets of explosive size in order to break the current computational barriers. The aim is to recall some basic methods which were developed earlier (and not directed to big data), to actually show that it is possible to drop the hypothesis concerning the explicit knowledge of standardisation in order to use subsampling for the construction of confidence regions or predictors at a larger scale. We show the benefit of such methods in terms of computational cost and algorithmic complexity by using simulations.

EO1263: **Using sampling methods to estimate rare stats on Twitter's graph**
Presenter: **Antoine Rebecq**, Universite Paris X, France
Many computer science or social science studies about Twitter focus on the analysis of tweets alone, without knowing the number or characteristics of accounts who wrote the tweets. In fact, in many cases gaining access to the whole Twitter graph is very costly. The alternatives are the "Rest API", which allows only a few queries per hour and the "Streaming API", which only outputs a fraction (1 percent) of the tweets published in real time that match a certain search query. Many reasearchers choose the latter, mostly because the number of tweets output is much larger than when using the former. However, this suffers an additional drawback: the sampling method used to select the 1 percent tweets output is not disclosed by Twitter, which means classic unbiased estimators cannot be used. We propose to use the "Rest API" along with an adaptive sampling method that focuses on the estimation of rare quantities. This suits well the problem of estimating stats about accounts that produced a set of tweets matching a certain query because the Twitter graph is so large that most search queries will only be met by a very small fraction of vertices. We use our method to assess the number of accounts behind the 1 million tweets posted in less than 3 hours about the Pluto flyby of June 14th, 2015.

EO218 Room SH349 ANALYSIS OF SURVEY DATA FOR UNPLANNED DOMAIN Chair: Monica Pratesi

EO1589: **Time series small area estimation for unemployment rates using latent Markov models**
Presenter: **Maria Giovanna Ranalli**, University of Perugia, Italy
Co-authors: Gaia Bertarelli, Francesco Bartolucci
In Italy, the Labour Force Survey (LFS) is conducted quarterly by the National Statistical Institute (ISTAT) to produce estimates of the labour force status of the population, at national, regional (NUTS2) and province (LAU1) levels. In addition, ISTAT also disseminates yearly LFS estimates of employed and unemployed counts and rates at a finer level given by Local Labour Market Areas (LLMAs). LLMAs are aggregations of municipalities and are defined at every census in terms of daily working commuting flows. In contrast with the NUTS3 and LAU1 levels, LLMAs are unplanned domains. The continuous nature of LFS allows us to borrow strength not only from other areas but also over time. We develop a new area-level SAE method using latent Markov (LM) models in a Bayesian setting. LM models allow for the analysis of longitudinal data when the response variable(s) measure common characteristics of interest that are not directly observable. In these models the characteristics of interest, and their evolution in time, are represented by a latent process that follows a Markov chain, so that statistical units are allowed to move between the latent states during the observation. Estimation is conducted using a Gibbs sampler with data augmentation and the proposed model is applied to estimate annual employment and unemployment rates for the Italian LLMAs using data from 2004 to 2013.

EO1681: **Two-level M-quantile model for poverty estimation**
Presenter: **Stefano Marchetti**, University of Pisa, Italy
Co-authors: Nicola Salvati
Small area estimation (SAE) aims to allow efficient estimation of population characteristics of domains with a small sample size that produce unreliable estimates. In the last decade there has been a rising interest in poverty estimation where the ELL (a simulation-based synthetic poverty mapping methodology) approach is the de-facto industry standard for small area estimation applied to poverty assessment. Alternatives to the ELL approach take into account area and unit heterogeneity, but implicitly assume homogeneity for clusters within area level (e.g. PSU). We propose a two-level M-quantile linear model that should be able to capture variability at area/domain and cluster level. The model in its simplest form can mimic a two-nested-error model. Using a Monte Carlo approach we can obtain poverty estimates via the two- level M-quantile model. Performance of poverty estimates will be shown by means of Monte Carlo simulation where we tried to build a realistic poverty estimation scenario.

EO1795: **Bayesian M-quantile regression in small area estimation**
Presenter: **Giovanni Riccardi**, University of Pisa, Italy
Co-authors: Carlo Trivisano, Enrico Fabrizi
Mixed effects models are widely used in SAE in order to estimate the effect for a particular area. However, such models depend on parametric and distributional assumptions as well as requiring specification of the random part of the model. An alternative approach to this regression estimation is the M-quantile regression. The M-quantile regression has been previously applied as basis of their small area estimation method. A number of papers on M-quantile regression applied to SAE has been published since then, concerning theoretical developments, extension to non-linear models and various applications. But it has not been proposed any Bayesian approach. The main objective is to propose a Bayesian M-quantile regression model specifying an asymmetric likelihood function based on Generalized Asymmetric Least Informative distribution. Furthermore, the Bayesian M-quantile regression will be applied to SAE in order to obtain small area estimators and their standard errors.

EO1798: **Geoadditive models for unplanned geographical domains**
Presenter: **Alessandra Petrucci**, University of Firenze, Italy
Co-authors: Chiara Bocci, Emilia Rocco
The complexity of spatial data and the intrinsic spatial relationships limit the usefulness of conventional techniques for extracting spatial patterns. Therefore, the area definition and the assignment of the data to appropriate areas can pose problems in the estimation process. Semi-parametric models have been proposed to simultaneously incorporate the spatial distribution of the study variable and the other covariate effects. Geoadditive models, in particular, merge an additive model, that accounts for the relationship between the variables, and a kriging model, that accounts for the spatial distribution, under the linear mixed model framework. The small area estimates are usually based on a combination of sample surveys and administrative data. Moreover, variables can be skewed, thus the relationship between the response variable and the auxiliary variables may not be linear in the original scale, but can be linear in a transformed scale e.g. the logarithm scale. In such case, small area estimation (SAE) methods based on log-transformed models are required. Combining the need of small area estimation methods for skewed variables with the flexibility of a semi-parametric model, we discuss a recent approach to identify and include the spatial pattern in small area estimation using a model-based direct estimator - MBDE.

EO106 Room Torrington ADVANCES IN FUZZY CLUSTERING Chair: Maria Brigida Ferraro

EO0454: **A possibilistic and fuzzy clustering algorithm for imprecise data**
Presenter: **Maria Brigida Ferraro**, Sapienza University of Rome, Italy
Co-authors: Paolo Giordani
The problem of clustering data affected by imprecision is considered. The imprecision is managed in terms of fuzzy sets, in particular, LR fuzzy data. The clustering process is addressed according to the fuzzy and possibilistic approaches. In both the approaches the objects are assigned to the clusters by means of membership degrees. In fuzzy clustering the membership degrees express the degrees of sharing of the objects to the clusters whereas, in possibilistic clustering, these give the degrees of compatibility. These two sources of information are not exclusive because the former helps to discover the best fuzzy partition of the objects and the latter gives how well the objects are described by the cluster prototypes. In order to exploit the benefits of both the approaches a new possibilistic fuzzy clustering algorithm for LR fuzzy data is proposed. This is done by hybridizing the fuzzy and possibilistic k-means algorithms for fuzzy data. The effectiveness of the proposal is investigated by means of real-case examples.

EO0914: **A fuzzy cluster scaling model**
Presenter: **Mika Ilic**, University of Tsukuba, Japan
A fuzzy cluster scaling model is proposed in which we assume that clusters vary simultaneously with respect to several dimensions. In fuzzy clustering, the result of the clustering is represented by degree of belongingness of objects to several fuzzy clusters. By using the degree as continuous values from 0 to 1, the result can allow the overlapping belongingness of objects to several clusters and show more feasible solutions as a clustering result when considering real world data situation. However, fuzzy clustering has a problem of clarity when interpreting the result of clustering. The main reason for this is that the result of the fuzzy clustering does not have the scale for indicating the difference between the belonging statuses of a pair of objects. Therefore, it is necessary to include the scaling to the fuzzy clustering result. Classical metric multidimensional scaling (MDS) is a method to obtain the latent structure of a data in a lower dimensional Euclidean space in order to summarize the data features and the visualization of the data structure. By exploiting the idea of MDS to the fuzzy clusters, the fuzzy cluster scaling model can obtain measurable fuzzy clusters in Euclidean space.

EO1036: **Advantages and limits of fuzzy clustering**
Presenter: **Frank Klawonn**, Ostfalia University of Applied Sciences, Germany
Co-authors: Roland Winkler
Fuzzy clustering (FC) relaxes the constraint of assigning an object to a single cluster by admitting membership degrees. Compared to mixture models using probabilities, FC is less rigorous in terms of a model but allows at for more flexible cluster shapes. While membership degrees to clusters convey more information than crisp ones, their computation could still be detained until the end of the clustering process. However, incorporating membership degrees in the clustering iteration scheme can level out local minima of the underlying objective function which lead to undesired clustering results. Clustering algorithms for more flexible cluster shapes were published as FC techniques although they could also work with crisp membership degrees. But since complex cluster shapes tend to introduce more local minima, they would make it difficult for crisp clustering to find the global minimum. Nevertheless membership degrees can also introduce new local minima, especially for high-dimensional data. To evaluate the capability of clustering algorithms to handle high-dimensional data sets, a library to create artificial data sets with various properties has been implemented and fuzzy FC algorithms and other techniques like k-means or Gaussian mixtures have been tested with these data sets. There is no best algorithm always outperforming the others. Their performance depends very much on the characteristics and the dimensionality of the data sets.

EG111 Room Athlone CONTRIBUTIONS ON LASSO Chair: Jacob Bien

EC1532: **Using the Lasso for gene selection in bladder cancer data**
Presenter: **Christophe Guyeux**, Universite de Franche Comte, France
Co-authors: Stephane Chretien
Given a gene expression data array of a list of bladder cancer patients with their tumor states, it may be difficult to determine which genes can operate as disease markers when the array is large and possibly contains outliers and missing data. An additional difficulty is that observations (tumor states) in the regression problem are discrete ones. We solve these problems on concrete data using first a clustering approach, followed by Least Absolute Shrinkage and Selection Operator (LASSO) estimators in a nonlinear regression problem involving discrete variables, as described in the brand-new research work of Plan and Vershynin. Gene markers of the most severe tumor state are finally provided using the proposed approach.

EC1376: **Forecasting using sparse cointegration**
Presenter: **Ines Wilms**, KU Leuven, Belgium
Co-authors: Christophe Croux
Cointegration analysis is used to estimate the long-run equilibrium relations between several time series. The coefficients of these long-run equilibrium relations are the cointegrating vectors. We provide a sparse estimator of the cointegrating vectors. Sparsity means that some elements of the cointegrating vectors are estimated as exactly zero, improving interpretability. The sparse estimator is applicable in high-dimensional settings, where the time series length is short relative to the number of time series. Our method achieves better estimation accuracy than the traditional Johansen method in sparse and/or high-dimensional settings. We use the sparse method for interest rate growth forecasting and consumption growth forecasting. The sparse cointegration method leads to important gains in forecast accuracy compared to the Johansen method.

EC0236: **Oracle inequalities for ranking with Lasso penalty**
Presenter: **Wojciech Rejchel**, Warsaw University, Poland
The goal of ranking (ordinal regression) is to guess or to predict the ordering between objects on the basis of their observed predictors. We assume that the number of predictors is large, it can be considerably greater than the sample size. Ranking estimators, that we look for, should predict accurately the ordering between objects as well as be able to select significant predictors in the model. To obtain such estimators we minimize the convex empirical risk, that is a U-process, with the Lasso penalty in the family of linear combinations of pre-specified base functions. We investigate properties of such estimators by proving probabilistic bounds for their risks. These results confirm that estimators behave almost like the oracle that knows which parameters should be included in the model by balancing the approximation risk and sparseness. Moreover, we obtain bounds for the l_1-distance between the estimator and the oracle. Basing on this bound one can prove model selection consistency while considering two-stage procedures. Theoretical results are completed by experiments on simulated and real data sets.

EC1482: **Mixed scale joint graphical lasso**
Presenter: **Eugen Pircalabelu**, Katholieke Universiteit Leuven, Belgium
Co-authors: Gerda Claeskens, Lourens Waldorp
We estimate brain networks from fMRI datasets that do not all contain measurements on the same set of regions. For certain datasets some of the regions have been split in smaller subregions, while others have not been split. This gives rise to the framework of mixed scale measurements. We estimate sparse undirected graphical models that combine information from several subjects using the data of all coarseness levels. This overcomes the problem of having different coarseness levels and takes into account that dependencies exist between a coarse scale node and its finer scale nodes. We use two algebraic operators to *expand* and *reduce* the matrices. The procedure uses a combination of *fused* or *group* graphical lasso penalties and an ℓ_1 penalty. Graphical models are estimated for each subject and coarseness level, referred to as *within* level edges, and identifies connections between a large region and its subregions, the *between* level edges. The method extends beyond fMRI data, to other areas where data on different scales are observed and where the joint estimation of similar undirected graphs is desired. Moreover, the method avoids the tedious task of selecting one coarseness level for the analysis and produces interpretable results at all available coarseness levels.

EG113 Room G21A CONTRIBUTIONS IN TIME SERIES ANALYSIS I Chair: Gregory Rice

EC1466: **Mixtures of distributions and volatility**
Presenter: **Juan Carlos Abril**, Universidad Nacional de Tucuman, Argentina
Co-authors: Maria de las Mercedes Abril, Carlos Ismael Martinez
Using Monte Carlo methods we generate time series with the following features: a) series with distributions that are the mix of two normal distributions with different variances, b) series that satisfy volatility models, c) series that satisfy an AR(1) model but with contaminated errors that follow the same distribution as the mixes given in a) and d) series that follow the same distribution as the mixes given in a) but with conditional heterocedasticity. From the analysis we see that it is difficult to identify in practical situations the real generating process of the series. In fact, the processes that come from distribution mixes have many similar characteristics to the ones that satisfy the volatility scheme. We use the corresponding theoretical considerations and also the usual tools in the identifying process of any time series; that is, series graphs, histograms, the corresponding sampling distributions, correlograms and partial correlograms.

EC1644: **Some advances about the probability structure of the SETAR process**
Presenter: **Marcella Niglio**, University of Salerno, Italy
Co-authors: Francesco Giordano, Cosimo Damiano Vitale
The study of probability properties (such as ergodicity and stationarity) of nonlinear time series models is not always an easy task and in most cases the results are incomplete or related to some simplified parameterizations of the models. In this domain we focus the attention on the Threshold Autoregressive model represented as SETAR model. In particular, we get an ergodic region on the parametric space wider than that proposed in the literature. Among the innovations of our contribution, differently to what usually done in the literature, we face the ergodicity of the SETAR process by investigating the dynamic structure of the indicator process that regulates the switching among regimes. In this way, we use a kind of dynamic approach which allows us to overcome the main theoretical issues. The previous results can be useful in different contexts such as the presence of unit roots in the SETAR model and the estimation of the parameters in the model under minimal conditions.

EC1754: **Deterministic optimal control helps estimating sloppy ordinary differential equations**
Presenter: **Nicolas Brunel**, ENSIIE, France
Co-authors: Quentin Clairon
The estimation of an ordinary differential equation (ODE) $\dot{x} = f(t,x,\theta)$ is a particular case of nonlinear parametric regression problem. The complex and implicit dependence of the trajectories in the parameters can give rise to sloppy parameters, meaning that the sensitivity matrix is badly conditionned, and uncertainty occurs in some directions of the parameter space. We propose to regularize the inverse problem and to reduce the variance estimation by adding a forcing function u in the original ODE. The objective is then to find the parameter θ that needs the smallest perturbation u in order to be close to the data. The derived procedure is close to the popular generalized profiling approach, but does not suffer from the same limitations. The estimation of the pertubation u is obtained by solving an optimal control problem and not a series expansion, hence avoiding a delicate tuning of the approximation error. In the case of linear ODEs, the criterion has a closed form expression thanks to the optimality criteria in function spaces. We show that the estimators are consistent, with optimal rate, and asymptotically normal. Moreover, experiments on simulated and real sloppy models show that our estimator improves on nonlinear least squares and generalized profiling fo sloppy models.

EC1571: **Long memory parameter estimation: Signal or noise**
Presenter: **Grace Yap**, The University of Nottingham Malaysia Campus, Malaysia
Co-authors: Wen Cheong Chin
Long-memory parameter estimation using bias-reduced log-periodogram regression (BRLP) $\hat{d}_r$ is proven efficient as it eliminates the first and higher order of biases of the log-periodogram model. Nonetheless, its performance relies largely on the frequency bandwidth m and the order of

estimation r. Literature suggests a data-dependent plug-in method for selecting the frequency bandwidth that minimizes the asymptotic mean-squared error (MSE). However, this choice of m significantly increases the MSE's over the finite sample minimum MSEs due to the non-parametric estimation problem in the unknown term within the plug-in method. In a long-memory time series with mild short range contamination, a simple approach to determine the bandwidth size is suggested based on the spectral analysis. Monte Carlo simulation results for stationary ARFIMA $(1,d,0)$ and ARFIMA $(0,d,1)$ processes show that with the proper order of estimation, the proposed bandwidth selection performs better than that of the MSE optimal choice.

Monday 14.12.2015 10:40 - 11:55 Parallel Session M – CFE-CMStatistics

CO546 Room Holden TIME-VARYING PARAMETERS Chair: Matei Demetrescu

CO0201: **The responsiveness of monetary policy to financial stress: A panel data analysis**
Presenter: **Danvee Floro**, University of Kiel, Germany
Co-authors: Bjoern van Roye
The aim is to test for regime-switching changes in monetary policy's response to increases in financial stress across a panel of advanced and emerging economy central banks. We employ a Factor-augmented dynamic panel threshold regression in order to address the issue of cross section dependence inherent in the panel setup. We find evidence of an asymmetric monetary policy response from advanced economy central banks in different financial stress regimes. The exceptional circumstances surrounding the global financial crisis exacerbated the easing effect of financial stress on advanced economy central monetary policy settings. Emerging market central banks react to financial stress by raising short-term interest rates up to a certain threshold, beyond which the interest rate response is insignificant. Our results are specific to the type of financial stress considered in the analysis. Stock market and banking stresses are driving advanced economy central banks' interest rate decisions to loosen monetary policy in the face of high financial instability. By contrast, emerging market central banks adjust interest rates positively in response to the overall financial stress, banking and foreign exchange stress subcomponents only when financial stress is low or moderate, and appear to be more sensitive to stock market stress.

CO0434: **Forecasting methods for functional time series**
Presenter: **Nazarii Salish**, BGSE and University of Cologne, Germany
A statistical theory is developed for forecasting functional times series. Despite the large number of potential applications, the statistical theory of functional times series is still in its infancy. First, to tackle the central issue of time dependence we introduce the notion of functional dependence through scores of the Karhunen-Loeve expansion. This approach allows us to adapt various notions of dependence available in the classical time series literature to the functional context. Second, we investigate the impact of time dependence thus quantified on the estimation of the functional principal components. The rate of mean squared convergence of the estimator of the covariance operator is derived under the so called long memory property of the series. Third, to provide forecasting methods applicable to linear and nonlinear functional time series we study the statistical properties of the functional autoregressive model and the functional k-nearest neighbors method in detail. In particular, the asymptotic behavior of the predictors is derived for both models. Finally, Monte Carlo simulations are presented to assess the accuracy of the proposed methods in different (linear and nonlinear) setups. Empirical relevance of the theory is illustrated through applications to electricity demand in the U.S. and MRI brain scans.

CO1357: **Testing long memory under time-varying short memory parameters**
Presenter: **Matei Demetrescu**, CAU Kiel, Germany
Distinguishing long from short memory is a challenging and important task of applied time series analysis. The typical set of assumptions underlying long memory tests requires weak stationarity, if not strong. Hence smooth or abrupt changes in the short memory distort limiting null distributions of long memory tests and are thus likely to generate spurious findings. A robust version of a previous test statistic is provided. To this end, we recast the test in a regression framework, which makes the use of heteroskedasticity and autocorrelation-consistent standard errors straightforward. Exploiting some nice properties of HAC estimators under locally stationary data generating processes, we then show that the limiting distribution of the modified test statistic is standard normal, even under time-varying dynamics and variances. We illustrate the behavior of our test in Monte Carlo simulations.

CO440 Room Senate ANALYSIS OF HIGH-DIMENSIONAL TIME SERIES II Chair: Marc Hallin

CO0209: **Non-stationary dynamic factor models for large macroeconomic databases**
Presenter: **Matteo Luciani**, Universite libre de Bruxelles, Belgium
Co-authors: Matteo Barigozzi, Marco Lippi
The Non-Stationary Dynamic Factor Model for Large Macroeconomic Databases is presented, which is an extension of the Dynamic Factor Model that is nowadays commonly used in Central Banks and Policy Institutions for both structural analysis and forecasting. The model is built in such a way that it both matches standard stylized facts of macroeconomic databases, and is consistent with commonly accepted macroeconomic theory. We discuss consistent estimation of the model as well as how to determine the number of common trends in large panels. Finally, we also show an application of our model to a large panel of US quarterly data to study the effects of monetary policy shocks.

CO0253: **Asymptotically UMP tests for unit roots in cross-sectionally dependent panels**
Presenter: **Gaia Becheri**, Delft University of Technology, Netherlands
Co-authors: Ramon Van den Akker
Testing for unit roots is a crucial aspect of (time series and) panel data analysis. The presence of unit roots not only determines how to proceed for a correct statistical inference but may also have relevant policy implications. For a correct analysis of cross-country data, the presence of heterogeneity and correlation between units should be taken into account. We consider large panels with cross-sectional dependence generated by dynamic common factors and derive the (asymptotic) power envelope for testing for a unit root. We demonstrate that some of the most popular second-generation tests are asymptotically equivalent and only attain the power envelope in case the long-run variances of the idiosyncratic components of the innovations are homogeneous. We propose new tests that are asymptotically UMP, i.e. whose power functions attain the envelope irrespective of possible heterogeneity in the long-run variances. We consider an asymptotic scheme in which n (the number of units) and T (the time series dimension) tend to infinity jointly, which is the common when studying macroeconomic panels with large n and T. Finally, we assess the finite-sample performances of the new tests via a Monte Carlo study.

CO1239: **Dynamic factor models with infinite-dimensional factor space: Forecasting**
Presenter: **Marco Lippi**, Universita di Roma La Sapienza, Italy
Co-authors: Mario Forni, Alessandro Giovannelli, Stefano Soccorsi
The aim is to compare the pseudo real-time forecasting performance of three different previous factor models. The standard US monthly large dataset of macroeconomic and financial time series, which includes the Great Moderation, the Great Recession and the subsequent recovery, is employed. In a ten-years rolling window framework, we find that the first two methods, based on spectral estimation, outperform the third on average. Moreover, applying a test for relative forecasting performance recently introduced, we find that the spectral-estimation methods significantly prevail in the Great Recession. We extend a previous method by introducing forecast combinations and we find that the prediction accuracy for the industrial production and inflation is greatly enhanced for some forecasting horizons.

CO472 Room SH349 NONREGULAR PANEL DATA AND TIME SERIES MODELS Chair: Carla Moreira

CO0221: **Asymptotic analysis and efficiency of large dynamic panel models**
Presenter: **Jose Diogo Barbosa**, EPGE-FGV, Brazil
Co-authors: Marcelo J Moreira
The efficient estimation of a dynamic panel model with fixed effects, time-series heteroskedasticity, and covariates is considered. This model allows natural rotational invariance conditions, such as re-ordering the individuals in the sample. An invariance principle determines a maximal invariant statistic. Its distribution yields an estimator for the structural parameters that is consistent as the number of cross-section and time-series observations increase with the sample size. Most estimators for this model are simply GMM estimators. Their respective moment conditions are found to be partial linear combinations of the first moment of the maximal invariant statistic. Consequently, this general framework allows us to obtain efficiency gains for these estimators in this simple dynamic panel data model.

CO0251: **Nonparametric estimation of non-exchangeable latent-variable models**
Presenter: **Koen Jochmans**, Sciences Po, France
The aim is to propose a two-step method to nonparametrically estimate multivariate models where the outcome variables are independent conditional on discrete latent variables. Applications include microeconometric models with discrete unobserved types of agents, regime-switching models, and models with misclassification error. In the first step, we estimate weights that transform moments of the marginal distribution of the data into moments of the conditional distribution of the data for given values of the latent variable. In the second step, these conditional moments are estimated as weighted sample averages. We illustrate the method by estimating a model of wages with time-invariant unobserved heterogeneity on PSID data.

CO0474: **Identification and inference in moments based analysis of linear dynamic panel data models**
Presenter: **Frank Kleibergen**, University of Amsterdam, Netherlands
Co-authors: Maurice Bun
The aim is to show that Difference, Level or Non-Linear moment conditions do not separately identify the parameters of a first-order autoregressive panel data model when the autoregressive parameter is close to one and the variance of the initial observations is large. We construct a new set of (robust) moment conditions that identify the autoregressive parameter irrespective of the variance of the initial observations. These robust moment conditions are (non-linear) combinations of the System (Sys) and Non-linear (AS) moment conditions. We use them to determine the maximal attainable power under the worst case setting which results in a quartic root convergence rate. It is identical for the AS and Sys moment conditions so assuming mean stationarity does not improve power in worst case settings. We compare the maximal attainable power under the worst case setting with the lower envelopes of power curves of different GMM test procedures. These power envelopes show the smallest rejection frequencies of these test procedures. The power envelope of the robust Lagrange Multiplier statistic coincides with the maximal attainable power curve under the worst case setting so the robust Lagrange multiplier statistic is efficient when the autoregressive parameter equals one which it also is for smaller values of the autoregressive parameter.

CO643 Room Athlone EARLY WARNING SYSTEM AND SYSTEMIC RISK INDICATORS III Chair: Benjamin Klaus

CO0268: **Leading indicators of financial stress: New evidence**
Presenter: **Robert Vermeulen**, De Nederlandsche Bank, Netherlands
The aim is to examine which variables have predictive power for financial stress in a sample of 25 OECD countries, using a recently constructed Financial Stress Index (FSI). First, we employ Bayesian model averaging to identify leading indicators of our FSI. Next, we use those indicators as explanatory variables in a panel model for all our countries and in models at the individual country level. It turns out that panel models can hardly explain FSI dynamics. Although better results are achieved in models estimated at the country level, our findings suggest that (increases in) financial stress is (are) hard to predict out-of-sample.

CO0296: **Dating systemic financial stress episodes in the EU countries**
Presenter: **Benjamin Klaus**, European Central Bank, Germany
Co-authors: Thibaut Duprey, Tuomas Peltonen
The aim is to identify in a transparent and objective way the financial stress events that are associated with a negative impact on the real economy. This is a first attempt to build a chronology of systemic financial stress beyond the expert-selected stress events available so far. As such it is a prerequisite for further macroprudential analyses. First we build on the financial stress literature and construct a very simple monthly country-specific index for EU-27 countries starting as early as 1964 for core European countries. Second we use the Markov Switching method widely used in the business cycle dating literature in order to endogenously distinguish low from high financial stress periods. Third, we use a simple algorithm to select the episodes of financial stress that are associated with a significant negative impact on the real economy. Our model-implied crisis dates are (a) consistent with many expert-detected stress periods, and (b) robust to the issue of event reclassification once new data become available.

CO1225: **A practical approach to financial crisis indicators based on random matrices**
Presenter: **Antoine Kornprobst**, University Paris Sorbonne, France
Co-authors: Raphael Douady
The aim is to build financial crisis indicators based on market data and our basic financial intuition is that correlation and volatility govern the financial market: when correlations between asset prices increase or develop abnormal patterns, when volatility starts to increase, then a crisis might be forthcoming. The first type of indicators relies on the Hellinger distance, computed between the spectral distribution of the empirical covariance matrix and the spectral distribution of a reference covariance matrix. The idea behind this approach is that when the empirical spectral distribution is deviating from the reference, then a crisis may be forthcoming. Indicators of the second type study the spectral radius and the trace of the covariance and correlation matrices as a mean to examine the volatility and correlations. The idea behind this approach is the fact that large eigenvalues are indicative of dynamic instability.

CO488 Room Gordon FILTERS WAVELETS AND SIGNALS Chair: Stephen Pollock

CO0762: **Testing for constant parameters in nonlinear models**
Presenter: **Carlos Rivero**, Complutense University of Madrid, Spain
Co-authors: Juan del Hoyo, Guillermo Llorente

There is an increasing interest in testing for structural breaks in economic and financial models. Estimation and diagnostic tests in nonlinear models is computationally more complex than in the linear ones. Testing for parameter instability in these models is also quite involving, because most of the tests require some kind of recursive/sequential estimation. Heavy computational load and low small-sample precision are two important reasons that limit the use of structural breaks tests in nonlinear models. A two steps method is proposed to test for parameters instability -structural change- in nonlinear models. The procedure is easy to implement and it has a low computational cost. The asymptotic distribution and the consistency of the procedure are derived. Monte Carlo simulations support the relevance of the proposed method, evaluate the performance of the procedure compared to a previous test, and highlight its small computational load.

CO0860: **Mixing data of different sampling frequencies in the frequency domain**
Presenter: **Marc Wildi**, Zurich University, Switzerland

Synthesizing information from data sampled at different frequencies - say daily, weekly, monthly or quarterly data - has a strong appeal in the context of real-time economic monitoring, since indicators can be up-dated continuously, as new information drops in. We propose a new mixed-frequency approach entirely designed in the frequency-domain. Specifically, we tackle the problem of integrating the various (unequally sampled) spectra within a common multivariate framework (MDFA: Multivariate Direct Filter Approach) and we derive closed-form solutions for various practically relevant optimization principles, including minimal (mean-square) revision errors. We discuss some of the advantages of the frequency-domain over the time-domain in the context of mixed-frequency data. Finally, the methodology is applied to a high-frequency indicator of US-GDP, combining quarterly and monthly macro-, weekly employment- and daily market-data.

CO1261: **The predictive content of business survey indicators: Evidence from SIGE**
Presenter: **Tatiana Cesaroni**, Bank of Italy, Italy

Business surveys indicators represent an important tool in economic analysis and forecasting practices. While there is wide consensus on the coincident properties of such data, there is mixed evidence on their ability to predict macroeconomic developments in the short term. We extend the previous research on business surveys predictive content by examining the leading properties of the main business survey indicators coming from the Italian Survey on Inflation and Growth Expectations (SIGE). To this end we provide a complete characterization of the business cycle properties of survey data (volatility, stationarity, turning points etc.) and we compare them with National Accounts reference series. We further analyze the forecast ability of the SIGE indicators to detect turning points using both discrete and continuous dynamic single equation models against their benchmark (B) ARIMA models. Overall the results indicate that SIGE business indicators are able to early detect turning points of their corresponding national account reference series. These findings are very important from a policy making point of view.

CO486 Room Montague SEQUENTIAL MONTE CARLO METHODS IN ECONOMETRICS Chair: Michael Pitt

CO1344: **Bond risk premia in consumption-based models**
Presenter: **Drew Creal**, University of Chicago, United States
Co-authors: Jing Cynthia Wu

The literature on recursive preference attributes all the time variation in bond risk premia to stochastic volatility. We introduce another source: time-varying prices of risk that co-move with inflation and consumption growth through a preference shock. We find that a time-varying price of risk driven by inflation dominates stochastic volatility in contributing to time variation in term premia. Once preference shocks are present, term premia are economically the same with or without stochastic volatility.

CO1359: **Bayesian modeling and forecasting of high-frequency volatility**
Presenter: **Jonathan Stroud**, Georgetown University, United States
Co-authors: Michael Johannes

Models of high-frequency index futures returns are estimated by using around-the-clock 5-minute returns that incorporate the following key features: multiple persistent stochastic volatility factors, jumps in prices and volatilities, seasonal components capturing time of the day patterns, correlations between return and volatility shocks, and announcement effects. We develop an integrated MCMC approach to estimate interday and intraday parameters and states using high-frequency data without resorting to various aggregation measures like realized volatility. We provide a case study using financial crisis data from 2007 to 2009, and use particle filters to construct likelihood functions for model comparison and out-of-sample forecasting from 2009 to 2012. We show that our approach improves realized volatility forecasts by up to 50% over existing benchmarks and is also useful for risk management and trading applications.

CO1806: **An extended space approach for particle Markov chain Monte Carlo methods**
Presenter: **Robert Kohn**, University of New South Wales, Australia
Co-authors: Eduardo Mendes, Christopher K Carter

We consider fully Bayesian inference in general state space models. Existing particle Markov chain Monte Carlo (MCMC) algorithms use an augmented model that takes into account all the variable sampled in a sequential Monte Carlo algorithm. An approach is described that also uses sequential Monte Carlo to construct an approximation to the state space, but generates extra states using MCMC runs at each time point. We construct an augmented model for our extended space with the marginal distribution of the sampled states matching the posterior distribution of the state vector. We show how our method may be combined with particle independent Metropolis-Hastings or particle Gibbs steps to obtain a smoothing algorithm. All the Metropolis acceptance probabilities are identical to those obtained in existing approaches, so there is no extra cost in term of Metropolis-Hastings rejections when using our approach. The number of MCMC iterates at each time point is chosen by the used and our augmented model collapses back to a previous model when the number of MCMC iterations reduces. We show empirically that our approach works well on applied examples and can outperform existing methods. The methodology will be illustrated with examples from financial econometrics.

CO474 Room Torrington DATA SCIENCE AND PLATFORM DESIGN Chair: Denis Nekipelov

CO1605: **Mechanism design for data science**
Presenter: **Denis Nekipelov**, University of Virginia, United States
Co-authors: Jason Hartline, Shuchi Chawla
Good economic mechanisms depend on the preferences of participants in the mechanism. For example, the revenue-optimal auction for selling an item is parameterized by a reserve price, and the appropriate reserve price depends on how much the bidders are willing to pay. A mechanism designer can potentially learn about the participants' preferences by observing historical data from the mechanism; the designer could then update the mechanism in response to learned preferences to improve its performance. However, in the context of large markets both the inference of participants' preferences and the simulation of their response to the new mechanism is computationally challenging due to both the data volume and the overall number of participants. We discuss a set of results that allow robust yet computationally simple approach for market analysis when the object of interest is an aggregate auction outcome such as revenue or welfare. In the setting where data is generated by a Bayes-Nash equilibrium it becomes possible making point predictions of counterfactual revenues and welfare of a new mechanism which are expressed as weighted order statistics of the observed bids. The lead illustration comes from the analysis of A/B testing for online auction platforms.

CO1608: **Simultaneous first-price auctions with preferences over combinations: Identification, estimation and application**
Presenter: **Tatiana Komarova**, London School of Economics and Political Science, United Kingdom
Co-authors: Matthew Gentry, Pasquale Schiraldi
Motivated by the empirical prevalence of simultaneous bidding across a wide range of auction markets, we develop and estimate a structural model of strategic interaction in simultaneous first-price auctions when objects are heterogeneous and bidders have preferences over combinations. We begin by proposing a general theoretical model of bidding in simultaneous first price auctions, exploring properties of best responses and existence of equilibrium within this environment. We then specialize this model to an empirical framework in which bidders have stochastic private valuations for each object and stable incremental preferences over combinations; this immediately reduces to the standard separable model when incremental preferences over combinations are zero. We establish non-parametric identification of the resulting model under standard exclusion restrictions, thereby providing a basis for both testing on and estimation of preferences over combinations. We then apply our model to data on Michigan Department of Transportation highway procurement auctions, we quantify the magnitude of cost synergies and assess possible efficiency losses arising from simultaneous bidding in this market.

CC0621: **Robust stochastic optimisation with indistinguishable models**
Presenter: **Anne Balter**, Maastricht University, Netherlands
Co-authors: Antoon Pelsser
Models can be wrong and recognising their limitations is important in financial and economic decision making under uncertainty. In asset pricing, model uncertainty has implications for the valuation of derivatives and long-dated contracts. We develop a method that provides a credible set of models to use in robust decision making. The choice of the specific size of the uncertainty region is what we will focus on. We use the Neyman-Pearson Lemma to characterise a set of models that cannot be distinguished statistically from a central model. Both deterministic and time-consistent stochastic deviations are proven to have maximal power for a log normal Radon-Nikodym derivative with bounded volatility. Therefore the set of indistinguishable models can explicitly be obtained ex ante, for a given Type I and II error. The quantification of uncertainty has applications for robust optimisation problems.

CO659 Room Bloomsbury BEHAVIOURAL AND EMOTIONAL FINANCE: THEORY AND EVIDENCE II Chair: Richard Fairchild

CO1158: **The role of immediate emotions in investment decisions**
Presenter: **Muhamed Alsharman**, Bath University, United Kingdom
Co-authors: Richard Fairchild
A huge debate exists over the impact of risk-preferences and emotions on decision-making and performance in financial markets. Many scholars argue that emotions interfere with effective decision-making: however, other researchers argue that emotions and cognition are interlinked, and therefore emotions promote good decision-making. We explore the relationship between emotions, risk preferences, and financial market trading and performance. We developed and conducted a series of neuro-economic trading experiments, employing questionnaires (containing both quantitative and qualitative analysis) and neuro-economic techniques (such as GSR, heart rate monitoring, and eye-tracking equipment) whilst subjects engaged in a computerised stock-trading game (which resembled a bear market). We find key and striking results relating to the interaction between risk preferences, emotions and trading behaviour and performance.

CO1209: **From behavioural to emotional corporate finance: A new research direction**
Presenter: **Richard Fairchild**, University of Bath, United Kingdom
Behavioural finance and behavioural corporate finance analyse the effects of psychological biases, heuristics, and emotions on investors and managers decision-making and performance. A major paradigm shift has been previously proposed by introducing a new field of research, namely, Emotional Finance. This ground-breaking approach employs Freud's theory of phantastic objects to analyse the effect of unconscious, infantile, emotions on investors decisions. We extend that paradigm by proposing a new development, namely, emotional corporate finance. We argue that, just as investors may view investments as phantastic objects, managers may view their projects similarly. We develop a formal approach that considers the effects of managerial phantasy on the investment appraisal decision, project performance, and managerial entrapment in a losing project. Our main results are as follows: a) Managerial project-phantasy may induce a manager to mistakenly invest in value-reducing projects. b) Phantasy may lead to volatility of managerial emotions, and hence volatility of project performance. c) Phantasy may lead to project entrapment, but may result in project abandonment if the managers project-phantasy turns to project hatred. We conclude by considering whether managerial phantasy could explain managements entrapment in the Concorde project.

CO0635: **On corporate bond market liquidity**
Presenter: **Xiaohua Chen**, Bank of England, United Kingdom
Investors complain that post-crisis regulatory reforms have reduced market liquidity. They say that dealers are less willing to trade corporate bonds and other fixed-income securities with them because they face additional costs in holding the resulting positions on their balance sheets. They fear that if asset managers, whose bond holdings have increased significantly in recent years, begin to sell then prices could fall sharply. Focusing on high-yield corporate bonds, we use an econometric model to investigate whether the typical responses of dealer inventories and market prices to falls in asset manager demand have changed in recent years. We find that dealer holdings act less as a shock absorber than they did around a decade ago. Instead, bond spreads rise more. We also find that greater declines in issuance result from these shocks.

CO052 Room Jessel HIGH DIMENSIONAL MODELS AND NETWORKS IN MACROECONOMICS AND FINANCE Chair: Matteo Barigozzi

CO0487: **Bank credit risk networks: Evidence from the Eurozone**
Presenter: **Christian Brownlees**, UPF, Spain
Co-authors: Christina Hanse, Eulalia Nualart
The credit risk of large financial institutions is highly interdependent as a result of a number of linkages between financial entities such as exposure to common asset classes and counterparty risk. We propose a novel methodology to study credit risk interconnectedness in large panels of financial institutions. Building upon the standard reduced form framework for credit risk, we introduce a model for European financial institutions in which defaults can be triggered by systematic global and country shocks as well as idiosyncratic bank specific shocks. The idiosyncratic shocks are assumed to have a sparse conditional dependence structure that we call the bank credit risk network. We then develop an estimation strategy based on Lasso regression that allows to detect and estimate network linkages from CDS data. We apply this technique to analyse the interdependence of large European financial institutions between 2006 and 2013. Results show that the credit risk network captures a substantial amount of dependence in addition to what is explained by systematic factors.

CO1581: **Sparse adjustment for spatial weight matrix and applications in spatial econometrics**
Presenter: **Clifford Lam**, London School of Economics and Political Science, United Kingdom
Co-authors: Pedro Souza
Spatial econometric models allow for the effect of spatial interactions among variables through the specification of a spatial weight matrix. While practitioners often face the risk of misspecification of such a matrix, in many problems they often have a number of potential specifications, for instance, specification through geographic distances or through various economic quantities among variables. To facilitate choosing a good spatial weight matrix for modeling, we propose to estimate such by a linear combination of spatial weight matrix specifications, added with a potential sparse adjustment matrix. The coefficients in the linear combination, together with the sparse adjustment matrix, are subjected to variable selection through the adaptive LASSO. Rate of convergence of all proposed estimators are spelt out, with simulations and a real data application demonstrating the performance of our procedure.

CO0379: **Supply chain disruptions: Evidence from the Great East Japan Earthquake**
Presenter: **Vasco Carvalho**, Cambridge and CREI, United Kingdom
The purpose is to examine whether propagation of idiosyncratic, firm-level shocks through input-output linkages can lead to sizable fluctuations at the aggregate level. Using a large-scale dataset on supply chain linkages among Japanese firms together with information on firm-level exposures to a large, but localized, natural-disaster — the Great East Japan Earthquake in 2011 — we quantify the earthquake's impact on firms that were (directly or indirectly) linked to affected firms. We find that having a supplier in the earthquake-hit region led to a 3% loss in terms of sales growth compared to firms with no such suppliers. We also find evidence for smaller but nevertheless significant upstream propagation from affected firms to their suppliers. Furthermore, we show that these losses do not remain confined to the disrupted firms' immediate customers and suppliers. Rather, firms that were only indirectly related to the firms in the affected areas (such as their customers' customers) were also negatively impacted. Even though our results suggest that such cascade effects decay with supply chain distance, the number of firms affected is large enough for this localized disruption to have a meaningful macroeconomic impact: the propagation of the earthquake shock over input-output linkages led to a 1% drop in Japan's aggregate output in the year following the earthquake.

CG555 Room Bedford CONTRIBUTIONS ON GARCH MODELS Chair: Roderick McCrorie

CC1492: **Mixed-frequency multivariate GARCH**
Presenter: **Jianbin Wu**, KU Leuven, Belgium
The aim is to introduce mixed-frequency multivariate GARCH models for forecasting low-frequency (weekly or monthly) multivariate volatility based on high-frequency intraday returns (at five-minute intervals) and on the overnight returns. The squared high-frequency returns enter the GARCH model through a parametrically specified mixed-data sampling (MIDAS) weight function. The proposed models are evaluated using five-minute and overnight return data on four DJIA stocks (AXP, GE, HD, and IBM) from January 1988 to November 2014. The mixed-frequency GARCH models are found to systematically dominate the low-frequency GARCH model in terms of in-sample fit and out-of-sample weekly forecasting accuracy. They also exhibit much lower low-frequency volatility persistence than the low-frequency GARCH model. Within the mixed-frequency models, the low-frequency persistence estimates decrease as the high- frequency increases from daily to five-minute frequency, and as overnight returns are included. Among the other findings are that the overnight volatility component exhibits more persistence than the intraday component; and that MIDAS weighting performs better than not weighting at all (i.e., than realized volatility).

CC1471: **The volatility of conditional correlation in multivariate GARCH models**
Presenter: **Cristiana Tudor**, Bucharest University of Economics, Romania
As the co-movements of asset returns in a portfolio constitute an important element in both portfolio and risk management, a multivariate GARCH volatility model should be more dependable than independent univariate models. Thus, three multivariate GARCH models are examined, namely the Dynamic Conditional Correlation GARCH model, the Generalized Orthogonal GARCH model and the Generalized Orthogonal GARCH model on empirical data from a sample of seven blue-chip companies listed at the Bucharest Stock Exchange. We estimate the conditional correlation of three stocks portfolios developed within our dataset and we are interested in the variability of conditional correlations estimates. If large variability is found between the alternate models, then the problem of the right choice among different multivariate CC models arises.

CC1597: **On volatility persistence and structural breaks**
Presenter: **Yuhan Zhang**, University of Leicester, United Kingdom
Co-authors: Alexandra Dias
Being one of the stylized features of financial markets, volatility persistence has been extensively studied. Recent research has conjectured that structural breaks in the volatility of financial assets could be induced by significant events, especially around a financial crisis. Moreover, not incorporating such structural breaks in GARCH type models could lead to an upward bias in volatility persistence. In order to address this problem, a modified ICSS algorithm is employed to identify the occurrence of structural breaks around the 2008 financial crisis in stock return series in the UK and China. Fifteen points are found in the return series in the UK, while five break points in China, indicating a relative stable volatility structure in the Chinese market over the sample period. Furthermore, after incorporating the detected structural breaks using dummy variables in a GARCH model, the volatility persistence of the return series drops in both stock markets, especially to a greater degree in the UK, suggesting a more severe size distortion of volatility persistence in this developed stock market than the Chinese emerging market. Observed from the data, the Chinese market experienced a more persistent volatility than the UK market, both before and after accommodating the structural breaks.

EO649 Room Beveridge Hall OBJECT ORIENTED DATA ANALYSIS II Chair: John Kent

EO0351: **Anisotropic distributions on manifolds, diffusion processes, and most probable paths**
Presenter: **Stefan Sommer**, University of Copenhagen, Denmark
Using the frame bundle of a differentiable manifold, stochastic development provides a mapping of sample paths of Euclidean space diffusion processes to paths on the manifold. We will use stochastic development to construct a family of probability distributions on manifolds that generalize Euclidean space normal distributions. This family will be used for likelihood estimation of mean and covariance structure from manifold valued data. The procedure avoids the commonly used linearization that arises when first estimating a mean before performing statistics in the tangent space of the mean. The focus will furthermore be on the notion of most probable paths reaching sampled data points. A lifted sub-Riemannian metric in the frame bundle will be used for deriving flow equations of the most probable paths.

EO1029: **Models and statistics for projective shape analysis**
Presenter: **John Kent**, University of Leeds, United Kingdom
The projective shape of a geometric object is the information that is invariant under projective transformations. The main application is to camera images, where the choice of projective transformation, or pose, corresponds to the camera view of the object. The simplest example of a projective shape is the cross ratio for a set of four collinear points. The way in which measurement errors in a camera image of an object affect the observed projective shape depends on the pose of the object. Thus the statistical analyis for a collection of images involves an interplay between the underlying projective shapes and the estimated poses.

EO0321: **Statistical tests for large tree-structured data**
Presenter: **Karthik Bharath**, University of Nottingham, United Kingdom
The Continuum Random Tree (CRT) proposed by Aldous arises as the (invariant) continuous limit, as the number of vertices grow without bound, for a general class of probability models for tree-structured data. We propose powerful goodness-of-fit tests for data which allow for hierarchical, tree-like representations using two different characterizations of the CRT, relating to a Brownian excursion and a special class of subtrees. Appropriateness of the tests on binary trees obtained from hierarchical clustering algorithms will be discussed, and applied to a dataset of tumour images with the objective of detecting heterogeneity of tumour.

EO280 Room MAL 633 NETWORK INFERENCE Chair: Ernst Wit

EO0497: **Exponential random graphical models**
Presenter: **Ernst Wit**, University of Groningen, Netherlands
Most statistical models are defined as a probability measure on some observable outcome. Clearly, this definition is rarely helpful directly to analyze real data. In fact, modern data tend to be rather complex. For example, genomic data comes from large monitoring systems with no prior screening. However, in most of these systems, these interactions are rather structured and the actual set of relationships, therefore, tends to be sparse. A graph is one possible way to describe complex relationships between many actors, such as for example genes and psychiatric symptoms. Graphical models present an appealing and insightful way to describe graph-based dependencies between the random variables. Although potentially still interesting, the main aim of inference is not the precise estimation of the parameters in the graphical model, but the underlying structure of the graph. Combining graphical models with exponential random graph models is an interesting new way to model the underlying topology of such non-observed graphs.

EO0522: **Nonparametrics and network summarization**
Presenter: **Sofia Olhede**, University College London, United Kingdom
Co-authors: Patrick Wolfe, Pierre-Andre Maugis
Relational data have become an important component of modern statistics. Networks are ubiquitous in modern applications such as disease dynamics, food webs and financial contagion. The inference of networks is harder, in parts because the measure placed on the observables need to satisfy sets of permutation invariances, and most networks are very sparse. The purpose is to explore how to best construct nonparametric summaries of such objects, in such a way that the underlying statistical model of the observations is well described, and any estimators computable with scalable algorithms.

EO0534: **Making sense of large genomic networks**
Presenter: **Veronica Vinciotti**, Brunel University London, United Kingdom
Inference of gene-regulatory networks from genomic data is an active research area and methods that estimate the inverse covariance matrix under sparsity constraints have received particular attention in this area. We present two applications of these approaches: the first aims to infer the dynamic regulatory network of the bacterium Neisseria meningitidis from microarray data, the second aims to detect differences in the regulatory mechanisms between the high and low glucose conditions from DeepSAGE data. The focus is on extracting knowledge from these high dimensional networks using appropriately developed enrichment analysis methods.

EO344 Room MAL 539 SEMI-NON-PARAMETRIC STATISTICS WITH HIGH-DIMENSIONALITY Chair: Rui Song

EO0560: **Genetic risk prediction using support vector machines**
Presenter: **Yair Goldberg**, University of Haifa, Israel
Co-authors: Malka Gorfine
Complex traits are the product of many genes but also reflect the influence of environmental and lifestyle factors. Risk prediction models estimate the risk of developing future outcomes based on both genes and environmental factors. We present a novel SVM methodology for risk prediction that can handle not only dichotomous traits but also continuous traits and censored data. We discuss the theoretical properties of the proposed methodology. We also present a computationally fast implementation of the algorithm using stochastic sub-gradient approach.

EO1329: **Variable selection in kernel regression using measurement error selection likelihoods**
Presenter: **Yichao Wu**, North Carolina State University, United States
We present a nonparametric shrinkage and selection estimator via a measurement error selection likelihood approach recently proposed. The Measurement Error Kernel Regression Operator (MEKRO) has the same form as the Nadaraya-Watson kernel estimator, but optimizes a measurement error model selection likelihood to estimate the kernel bandwidths. Much like LASSOor COSSO solution paths, MEKRO results in solution paths depending on a tuning parameter that controls shrinkage and selection via a bound on the harmonic mean of the pseudo-measurement error standard deviations.

EO1559: **Semi-varying coefficient multinomial logistic regression for disease progression risk prediction**
Presenter: **Bo Fu**, University College London, United Kingdom

A risk prediction model is proposed by using a semi-varying coefficient multinomial logistic regression. The model can be used to improve predictive modelling when non-linear interactions between predictors are present. We use a penalized local likelihood method to do the model selection and estimate both functional and constant coefficients in the selected model. The proposed method is used to classify the patients with early rheumatoid arthritis at baseline into different risk groups in future disease progression and we also propose a recalibration method to evaluate the reliability of the proposed risk prediction model.

EO270 Room MAL 541 STATISTICAL ANALYSIS OF TEXT Chair: Mattias Villani

EO0586: **Statistical analysis of the text of computer programs**
Presenter: **Charles Sutton**, University of Edinburgh, United Kingdom

Billions of lines of source code have been written, many of which are freely available on the Internet. The text of this code can be analysed statistically like any other textual corpus, with the goal of identifying textual patterns that characterise software systems that are more reliable and whose code is easy to read. We describe three new tools based on statistical textual analysis that are designed to help software developers write better programs. First, Naturalize is a system that suggests more descriptive names for local variables and functions based on Markov models. Second, TASSAL is a system that summarizes code by automatically hiding regions of code that are least informative according to a latent Dirichlet model. Finally, HAGGIS is a system that learns textual patterns that have syntactic structure, such as for-loops that iterate over vectors. HAGGIS accomplishes this using a nonparametric Bayesian probabilistic grammar.

EO0691: **Processing text data with latent-variable grammars**
Presenter: **Shay Cohen**, University of Edinburgh, United Kingdom

Enormous growth in the amount of information available from various data resources is being faced. This growth is even more notable when it comes to text data; the number of pages on the internet, for example, is expected to double itself every five years, with billions of multilingual webpages already available. In order to make use of this information, much of it needs to be parsed into natural language structures, such as syntactic trees or semantic graphs. In natural language processing, two important tools are available for this kind of structured prediction: probabilistic grammar formalisms and latent-variable modeling. Probabilistic grammar formalisms are a family of statistical models that give a principled way to process textual data and predict various types of structures for it. Latent-variable modelling, on the other hand, helps to discover patterns in data that are hard to manually annotate. We describe some work that combines these two ideas. We present an algorithm for estimating latent-variable grammars that stands in stark contrast to algorithms that have been used insofar for such estimation, such as the expectation-maximization algorithm. We simplify the algorithm, demonstrating a fast, simple algorithm for doing latent-variable grammar learning. We also describe three applications for these latent-variabe grammars that analyze text: syntactic parsing, machine translation and analysis of text in online forums.

EO0744: **Parallelizing LDA using partially collapsed Gibbs sampling**
Presenter: **Mans Magnusson**, Linkoping University Sweden, Sweden
Co-authors: Mattias Villani, Leif Jonsson, David Broman

Latent dirichlet allocation (LDA) is a model widely used for unsupervised probabilistic modeling of text and images. MCMC sampling from the posterior distribution is typically performed using a collapsed Gibbs sampler that integrates out all model parameters except the topic indicators for each word. The topic indicators are Gibbs sampled iteratively by drawing each topic from its conditional posterior. The popularity of this sampler stems from its balanced combination of simplicity and efficiency, but its inherently sequential nature is an obstacle for parallel implementations. Growing corpus sizes and increasing model complexity are making inference in LDA models computationally infeasible without parallel sampling. We propose a parallel implementation of LDA that only collapses over the topic proportions in each document and therefore allows independent sampling of the topic indicators in different documents. We develop several modifications of the basic algorithm that exploits sparsity and structure to further improve the performance of the partially collapsed sampler. Contrary to other parallel LDA implementations, the partially collapsed sampler guarantees convergence to the true posterior. We show on several well-known corpora that the expected increase in statistical inefficiency from only partial collapsing is smaller than commonly assumed, and can be more than compensated by the speed-up from parallelization for larger corpora.

EO296 Room MAL 532 ROBUSTNESS AND MISSING DATA Chair: Graciela Boente

EO0631: **Robust semiparametric estimation with missing responses**
Presenter: **Francesco Bravo**, University of York, United Kingdom

Estimation in a class of semiparametric models is considered when some of the responses are missing at random and outliers are present. A general estimator is proposed for the unknown infinite dimensional parameter using inverse probability weigthing and robust local linear estimation. An asymptotically equivalent one step version of the estimator that is computationally attractive is also proposed. The result is illustrated with two examples: a robust semiparametric quasi-likelihood and a robust semiparametric regression estimator. A simulation study shows that the proposed estimators have competitive finite sample properties

EO0643: **Robustness against model misspecifications in missing data analysis**
Presenter: **Peisong Han**, University of Waterloo, Canada

Methods that are robust against model misspecifications are highly desired. In missing data analysis, doubly robust methods have received wide attention due to their double protection on estimation consistency. Doubly robust estimators are consistent if either the model for selection probability or the model for data distribution is correctly specified. We propose a method that exhibits a further improved robustness. This method can simultaneously account for multiple models for both selection probability and data distribution. The resulting estimators are consistent if any one model is correctly specified, without knowing exactly which one it is. When both selection probability and data distribution are correctly modeled, the resulting estimators achieve maximum possible efficiency, again without knowing which models are the correct ones. This new method is based on the calibration idea in sampling survey literature, and has a strong connection to empirical likelihood. Another superior property of the multiply robust estimators is that, unlike many existing ones, they are not sensitive to near-zero values of estimated selection probabilities. Simulation evidence will also be presented to demonstrate the excellent numerical performance of the new method.

EO0657: **Robust tests in nonlinear regression models with possible missing responses**
Presenter: **Ana Maria Bianco**, Universidad de Buenos Aires and CONICET, Argentina
Co-authors: Paula Spano

The linear model has been widely used, but sometimes, the nature of the data or some other considerations suggest that the relationship between the response and the covariates is not linear in the unknown parameters. In addition, in many situations, just by chance or by design, there is missing data. We present a robust test based on a Wald-type statistic to check hypotheses that involve the parameter of a nonlinear regression model when there are responses missing at random. The tests statistic is based on a weighted $MM-$estimator of the parameter. The asymptotic behaviour of the

proposed testing procedure is studied under the null hypothesis and under contiguous alternatives. The influence function of the test is also derived. The proposal can be also used in complete samples, obtaining the full data results as a corollary of the missing data case. The performance of the proposed test is investigated through simulations, under different missing probability schemes and both, in contaminated and uncontaminated samples. The procedure is illustrated on a real data example.

EO619 Room Chancellor's Hall RISK MANAGEMENT AND DEPENDENCE MODELLING Chair: Irene Gijbels

EO0693: **Dependence uncertainty for aggregate risk: Examples and simple bounds**
Presenter: **Edgars Jakobsons**, ETH Zurich, Switzerland

Over the recent years, numerous results have been derived in order to assess the properties of regulatory risk measures (in particular VaR and ES) under dependence uncertainty. We complement this mainly methodological research by providing several numerical examples for both homogeneous as well as inhomogeneous portfolios. In particular, we investigate under which circumstances the so-called worst-case VaR can be well approximated by the worst-case (i.e. comonotonic) ES. We also study best-case values and simple lower bounds.

EO1079: **Understanding and visualizing complex dependencies for risk assessment**
Presenter: **Roberta Pappada**, University of Trieste, Italy

In the analysis of financial risks, a main task is the study of the interdependencies among the involved random variables (or individual risk factors), moving away from simplified assumptions (e.g., independence) and misleading numerical quantities (e.g., linear correlation coefficients). This may allow to identify risks that seem to be highly unlikely to occur but could have a major impact on the estimation of the global risk. Different approaches for a preliminary analysis of the extremal joint behaviour of the positions in a portfolio are presented. In particular, various copula-based measures of tail dependence are evaluated in order to detect the presence of homogeneous clusters of random variables from a multivariate dataset. The visualization needs to provide an efficient way for the user to gain insight into risky dependency scenarios. From another perspective, these graphical copula tools may supply valuable indications for a preliminary overview of the tail features of different copula families helpful in the choice of a suitable parametric model. This latter aspect may be particularly important during extreme market downturns, in order to construct better risk-mitigating mechanisms. The usefulness of the proposed tools will be discussed and the application to real data will be illustrated.

EO0881: **The impact of varying dependence structures on sums of random variables and implications for portfolio selection.**
Presenter: **Klaus Herrmann**, KU Leuven, Belgium
Co-authors: Irene Gijbels

We study sums of random variables $Z = \sum_{i=1}^{d} \omega_i X_i$ that are derived from a random vector $\mathbf{X} = (X_1, \ldots, X_d)$ and given weights $(\omega_i)_{i=1}^{d}$. Aside from classical results based on independence or joint normality of $\mathbf{X}$ there have been recent developments to study properties of Z more generally using copulas to model the dependence structure on $\mathbf{X}$. While current approaches focus on random vectors with components that are almost surely bounded from below we alleviate this constraint, allowing for unbounded marginal distributions. In this new formulation the influence of the margins is clearly separated from the influence of the dependence structure, a key feature of this copula based approach. This leads to new numerical algorithms to compute the density, distribution and quantile function of Z. Due to the importance of sums of random variables in actuarial and financial applications we highlight the impact of changing dependence structures on the properties of Z. Our main application is the portfolio choice problem, where the weights $(\omega_i)_{i=1}^{d}$ are now treated as decision variables. In this context we discuss an optimal choice of weights that minimize the expected shortfall of the associated portfolio.

EO346 Room MAL 540 CLUSTERWISE METHODS Chair: Gilbert Saporta

EO0767: **Clusterwise multiblock PLS regression**
Presenter: **Ndeye Niang**, CNAM, France
Co-authors: Stephanie Bougeard, Gilbert Saporta

Clusterwise regression methods aim at partitioning data sets into clusters characterized by their specific coefficients in the regression model. Usually, one dependent variable is linearly related to independent variables which are in a single data table. We present clusterwise multiblock PLS: an extension of clusterwise PLS regression to multiresponse variables and independent variables organized in meaningful blocks. This block structure is taken into account through a set of weights based on the importance of the block on the response prediction. This new method provides a partition of the data such as each of its cluster is associated with its own PLS model, which is then used to improve the overall fit of the prediction step. To do so, a new observation is first assigned to the relevant cluster minimizing a specific distance measure or maximizing the class membership probability. The prediction is then performed using the associated local model or using model averaging strategies. This general approach is based on a clear criterion to minimize and can be directly extended to other multiblock regression methods. The clusterwise multiblock PLS regression will be illustrated on both synthetic and real data.

EO1121: **Clustering categorical functional data**
Presenter: **Cristina Preda**, INRIA Lille, France
Co-authors: Cristian Preda, Vincent Vandewalle

Categorical functional data represented by paths of a stochastic jump process are considered for clustering. For paths of the same length, the extension of the multiple correspondence analysis allows the use of well-known methods for clustering finite dimensional data. When the paths are of different lengths, the analysis is more complex. In this case, for Markov models we propose an EM algorithm to estimate a mixture of Markov processes. A simulation study as well as a real application on hospital stays will be presented.

EO0517: **Calibration by stupid k-means and other aspects of cluster validation**
Presenter: **Christian Hennig**, UCL, United Kingdom

There are various measurements for cluster validity. Often these are used in such a way that the validity of the whole clustering is measured by a single number such as the Average Silhouette Width. But the quality of a clustering has various aspects such as within-cluster homogeneity, between-cluster separation, representation of cluster members by a centroid object, stability or within-cluster normal distribution shape, and what is most important depends on the aim of clustering. Furthermore, in many clusterings, various aspects of cluster validity differ between clusters. We discuss a number of measurements of different aspects of cluster validity, partly to be evaluated for every single cluster, including some plots to summarise the measurements. A key aspect is calibration, i.e., making different measurements comparable, so that they can be used, for example, to compare different numbers of clusters. The proposed approach is to explore the variation of the index over several clusterings of the same dataset that can be generated by random clustering methods called "stupid k-means" (i.e., assigning points to a random set of centroids) or "stupid nearest neighbour" (i.e., adding nearest neighbours starting from random points).

EO250 Room MAL G15 COMBINATORIAL OPTIMIZATION FOR GRAPHICAL MODEL SELECTION I Chair: Cassio Polpo de Campos

EO0843: **Polyhedral aspects of score equivalence in Bayesian network structure learning**
Presenter: **James Cussens**, University of York, United Kingdom
Co-authors: Milan Studeny, David Haws
The focus is on the faces and facets of the family-variable polytope and characteristic-imset polytope, which are special polytopes used in integer linear programming approaches to learn Bayesian network structure. A common form of linear objective to be maximised in this area leads to the concept of score equivalence (SE), both for linear objectives and for faces of the family-variable polytope. We characterize the linear space of SE objectives and establish a one-to-one correspondence between SE faces of the family-variable polytope, the faces of the characteristic-imset polytope, and standardized supermodular functions. The characterization of SE facets in terms of extremality of the corresponding supermodular function gives an elegant method to verify whether an inequality is SE-facet-defining for the family-variable polytope. We also show that when maximizing an SE objective one can eliminate linear constraints of the family-variable polytope that correspond to non-SE facets. However, we show that solely considering SE facets is not enough as a counter-example shows; one has to consider the linear inequality constraints that correspond to facets of the characteristic-imset polytope despite the fact that they may not define facets in the family-variable mode.

EO0598: **Towards ILP-based learning decomposable models by means of characteristic imsets**
Presenter: **Milan Studeny**, Institute of Information Theory and Automation of ASCR, Czech Republic
Co-authors: James Cussens
One of promising approaches to structural learning decomposable graphical models is the integer linear programming (ILP) approach based on the idea to represent these particular graphical models by means of special zero-one vectors, the so-called characteristic imsets of the corresponding Bayesian network model. This leads to the study of a special polytope, defined as the convex hull of all characteristic imsets for decomposable graphs we name the decomposable characteristic imset polytope. Some of the valid inequalities for this polytope and its LP relaxations have been studied previously. Recently, we succeeded to obtain all facet-defining inequalities for this polytope for a small number of variables (= nodes). Surprisingly, this leads to a sensible conjecture how these inequalities look in general and how to classify the facets of the polytope.

EO0628: **Optimal triangulation of Bayesian networks via integer linear programming**
Presenter: **Robert Cowell**, City University London, United Kingdom
There has been much work recently on optimal learning (in the sense of maximizing a decomposable scoring function) Bayesian networks from a dataset of observations using Integer Linear Programming techniques. To carry out efficient exact inference on such a learned network, or indeed any Bayesian network, usually requires finding an associated junction tree for message propagation. There are typically many possible junction trees for a given Bayesian network, and finding an optimal junction (in the sense, for example, of minimizing the sum of the state spaces of the cliques of the junction tree) is an NP hard problem. An algorithm is presented for finding optimal junction trees using on Integer Linear Programming.

EO298 Room Court RECENT DEVELOPMENTS IN BAYESIAN QUANTILE REGRESSION Chair: Mauro Bernardi

EO0902: **Bayesian inference for generalised quantiles**
Presenter: **Valeria Bignozzi**, University of Rome Sapienza, Italy
Co-authors: Mauro Bernardi, Lea Petrella
We investigate analytical properties and financial meaning of generalised quantiles used as risk measures. We compare them with classical examples of risk measures such as Value-at-Risk (VaR), Expected Shortfall and Conditional Tail-Variance. Thanks to their elicitability property we are able to extend the concepts of quantile and asymmetric least square regression used for VaR and expectiles to generalised quantiles and provide a link with the kernel of the Asymmetric Exponential Power Distribution (AEPD). This distribution can then be used to perform Bayesian inference for the estimation of generalised quantiles.

EO0959: **Bayesian binary quantile regression for the analysis of bachelor-master transition**
Presenter: **Cristina Mollica**, Sapienza Universita di Roma, Italy
Co-authors: Lea Petrella
The multi-cycle organization of the modern university systems stimulates the interest in studying the progression to higher level degree courses during the academic career. In particular, after the achievement of the first level qualification (Bachelor degree), students have to decide whether to continue their university studies, by enrolling in a second level (Master) programme, or to conclude their training experience. We propose a binary quantile regression approach to analyse the Bachelor-Master transition adopting the Bayesian inferential perspective. Quantile regression represents a well-established and useful device to gain a more in-depth understanding of the relation between the outcome of interest and the explanatory variables. By using the data augmentation strategy, quantile regression modeling for continuous responses has been recently extended for the treatment of binary response variables. We illustrate the utility of the Bayesian binary quantile regression approach to characterize the non-continuation decision with an application to administrative data of Bachelor graduates at the Faculty of Economics of "Sapienza" University of Rome.

EO1342: **Bayesian factor–augmented dynamic quantile VAR**
Presenter: **Mauro Bernardi**, University of Padova, Italy
A novel Bayesian model to estimate multi-quantiles in a dynamic framework is introduced. The main innovation relies on the assumption that the τ-th quantile of a vector of response variables depends on macroeconomic variables as well as on latent factors having their on stochastic dynamics. The proposed framework can be conveniently thought as a factor-augmented vector autoregressive extension of traditional univariate quantile models. We develop Bayesian methods that rely on state space methods and data augmentation approaches that efficiently deal with the estimation of model parameters and the signal extraction from latent variables. We estimate the model using a large panel of US equity market returns and macroeconomic variables to analyse the dynamic evolution of spillovers in individual Value-at-Risks.

EO062 Room G21A HIGH-DIMENSIONAL LATENT VARIABLE MODELS Chair: Veronika Rockova

EO1103: **Bayesian nonparametric inference on latent graphs**
Presenter: **James Johndrow**, Duke University, United States
Co-authors: Anirban Bhattacharya, David Dunson
Graphical models are central to machine learning and multivariate statistics. The number of parameters in these models grows at a super exponential rate in the number of nodes. The literature mainly focuses on cases where the graph is very sparse and consistent estimation is possible at the ratio of sample size to number of parameters typically encountered. In many application areas, such as genetics, epidemiology, and neuroscience, these extreme sparsity conditions are considered scientifically unrealistic. In these cases, a more realistic model in one in which the joint distribution of the data can be explained by a much lower-dimensional latent graph, or what in machine learning is referred to as a deep network. These latent graphical models can be very low-dimensional yet correspond to traditional graphical models that are non-sparse. We propose a fully Bayesian approach to inference on latent graph structure that uses a nonparametric model of dependence between latent nodes. Some theoretical properties of the model are derived, and the model is applied to learning the dependence structure between scientifically meaningful groups of variables from a genetic epidemiological study and an electrophysiology dataset.

EO1069: **Robust Bayesian inference via coarsening**
Presenter: **Jeff Miller**, Duke University, United States
Co-authors: David Dunson
The standard approach to Bayesian inference is based on the assumption that the distribution of the data belongs to the chosen model class. However, for large and/or high-dimensional datasets, even a small violation of this assumption can have a large impact on the outcome of a Bayesian procedure. We introduce a simple, coherent approach to Bayesian inference that improves robustness to perturbations of the model: rather than condition on the data exactly, one conditions on a neighborhood of the empirical distribution. In certain cases, inference is easily implemented using standard methods. We illustrate with real and simulated data, and provide theoretical results.

EO1227: **Hierarchical infinite divisibility for multiscale shrinkage**
Presenter: **Vinayak Rao**, Purdue University, United States
We describe a new shrinkage-based construction for an $n-$dimensional compressible vector x, for cases in which the components are naturally associated with a tree structure. Important examples are when x corresponds to the coefficients of a wavelet or block-DCT representation of data. The method we consider in detail, and for which numerical results are presented, is based on the gamma distribution. The gamma distribution is a heavy-tailed distribution that is infinitely divisible, and these characteristics are leveraged within the model. We further demonstrate that the general framework is appropriate for many other types of infinitely-divisible heavy-tailed distributions. Bayesian inference is carried out by approximating the posterior with samples from an MCMC algorithm, as well as by constructing a variational approximation to the posterior. We also consider expectation-maximization (EM) for a MAP (point) solution. State-of-the-art results are manifested for compressive sensing and denoising applications, the latter with spiky (non-Gaussian) noise.

EO655 Room MAL 421 POPULATION MODELS: METHODS AND COMPUTATIONS II Chair: Maria Conceicao Serra

EO1149: **EM algorithm for the multivariate power series offspring distributions**
Presenter: **Vessela Stoimenova**, Sofia University, Bulgaria
Co-authors: Dimitar Atanasov
The two-type discrete-time branching process is considered supposing that the individual distributions of both particle types belong to the bivariate power series distribution family. The focus of the study is the parametric estimation of the offspring distribution parameters. Based on the maximum likelihood estimators over the entire family tree EM estimators are found when data on the generation sizes over one trajectory is collected. Simulation and computation examples are carried out.

EO0616: **A test of homogeneity for age-dependent branching processes with immigration**
Presenter: **Ollivier Hyrien**, University of Rochester, United States
Co-authors: Nikolay Yanev
A novel procedure is presented to test whether the immigration process of an age-dependent branching process with immigration is time-homogeneous. The construction of the test is motivated by the behavior of the coefficient of variation of the population size. When immigration is time-homogeneous, we find that the coefficient of variation converges to a constant, whereas when immigration is time-inhomogeneous we find that it is time-dependent, at least transiently. Thus, we test the assumption that the immigration process is time-homogeneous by verifying that the sample coefficient of variation does not vary signicantly over time. The test is simple to implement and does not require specication or fitting any branching process to the data. Simulations and an application to real data are presented to illustrate the approach.

EO0156: **Some methodological and computational contributions to the class of discrete-time two-sex branching processes**
Presenter: **Manuel Molina**, University of Extremadura, Spain
Co-authors: Manuel Mota, Alfonso Ramos
In the general setting of statistical modelling, branching processes deal with populations whose size evolves over time, due to random births and deaths. It is an active research area of theoretical and practical interest with applicability to such fields as biology, demography, ecology, epidemiology, genetics, population dynamics, and others. Most animal species reproduce sexually, which requires the involvement of females and males in the population. Two important biological phases are carried out: mating and reproduction. Mating is the pairing of females and males, usually for the main purpose of reproduction. In the reproduction phase, the couples produce new female and male descendants. Recently, in order to describe the demographic dynamics of animal populations with sexual reproduction, some discrete-time two-sex branching processes, changing over is focused in such stochastic models. Several contributions and applications are presented.

EO300 Room MAL 402 DEPENDENCE MODELS IN BIOSTATISTICS AND BIOINFORMATICS Chair: Simone Vantini

EO1790: **Bayesian inference from multiple sources to inform infectious disease health policy**
Presenter: **Daniela de Angelis**, University of Cambridge, United Kingdom
Health related policy decision making for epidemic control is increasingly evidence-based, relying on the use of defendable models that realistically approximate the processes of interest and, crucially, incorporate all available information. From a statistical point of view, assimilation of information from a variety of heterogeneous, incomplete and biased sources poses a number of problems. We describe how a Bayesian approach to such evidence synthesis can accommodate all information in a single coherent probabilistic model and give examples to illustrate current challenges in this area.

EO1254: **Variable selection in covariate dependent random partition models: An application to urinary tract infection**
Presenter: **Maria De Iorio**, UCL, United Kingdom

Lower urinary tract symptoms (LUTS) can indicate the presence of urinary tract infection (UTI), a condition that if it becomes chronic requires expensive and time consuming care as well as leading to reduced quality of life. Detecting the presence and gravity of an infection from the earliest symptoms is then highly valuable. Typically, white blood cell count (WBC) measured in a sample of urine is used to assess UTI. We consider clinical data from 1341 patients at their first visit in which UTI (i.e. WBC 1) is diagnosed. In addition, for each patient, a clinical profile of 34 symptoms was recorded. We propose a Bayesian nonparametric regression model based on the Dirichlet Process (DP) prior aimed at providing the clinicians with a meaningful clustering of the patients based on both the WBC (response variable) and possible patterns within the symptoms profiles (covariates). This is achieved by assuming a probability model for the symptoms as well as for the response variable. To identify the symptoms most associated to UTI, we specify a spike and slab base measure for the regression coefficients: this induces dependence of symptoms selection on cluster assignment. Posterior inference is performed through Markov Chain Monte Carlo methods.

EO0561: **Flexible copula models for mixed binary-continuous data**
Presenter: **Rosalba Radice**, Birkbeck University of London, United Kingdom
Co-authors: Giampiero Marra

The focus is on regression models for associated mixed binary and continuous outcomes constructed using copulae. The approach entails specifying marginal regression models for the outcomes, and combining them via a copula function to form a joint model. Specifically, the framework allows for Gaussian and non Gaussian dependencies and for the mean, higher order moments and association parameters to be heterogeneous by incorporating flexible linear predictor structures. The utilization of penalized regression splines and Gaussian Markov random fields allows one to account for non-linear covariate effects and for geographic clustering. A ridge penalty avoids convergence failures, even when the parameters of a highly collinear variable are not fully identified. The theoretical background and software for straightforward implementation of this approach are provided. The approach is illustrated by fitting interpretable models of different complexity on different data-sets.

EO242 Room CLO B01 ADVANCED FUNCTIONAL DATA ANALYSIS Chair: Jian Qing Shi

EO0388: **Models for sparse and spatially correlated functional data**
Presenter: **Surajit Ray**, University of Glasgow, United Kingdom
Co-authors: Giles Hooker, Chong Liu

The analysis of geo-spatially correlated functional data is considered. The between-curve correlation is modeled by correlating functional principal component scores of the functional data. We propose a Spatial Principal Analysis by Conditional Expectation framework to explicitly estimate spatial correlations and reconstruct individual curves. This approach works even when the observed data per curve are sparse. Assuming spatial stationarity, empirical spatial correlations are calculated as the ratio of eigenvalues of the smoothed covariance surface $Cov(X_i(s),X_i(t))$ and cross-covariance surface $Cov(X_i(s),X_j(t))$ at locations indexed by i and j. Then a anisotropy Matern spatial correlation model is fit to empirical correlations. Finally, principal component scores are estimated to reconstruct the sparsely observed curves. This framework can naturally accommodate arbitrary covariance structures, but there is an enormous reduction in computation if one can assume the separability of temporal and spatial components. We propose hypothesis tests to examine the separability as well as the isotropy effect of spatial correlation. Simulation studies and applications of empirical data show improvements in the curve reconstruction using our framework over the method where curves are assumed to be independent.

EO0748: **Bayesian functional data analysis using Gaussian process priors**
Presenter: **Taeryon Choi**, Korea University, Korea, South

Some aspects of Bayesian functional data analysis are provided using Gaussian process priors in theory and methods. For theoretical aspects, we consider the asymptotic properties of Bayesian functional linear regression models where the response is a scalar and the predictor is a random function. We discuss a theoretical study with the goal of understanding the posterior contraction rate in the Bayesian functional linear regression. For methodological aspects, we consider a nonparametric Bayesian approach to smooth all functional observations simultaneously and nonparametrically. In the proposed approach, we assume that the functional observations are independent Gaussian processes subject to a common level of measurement errors, in a hierarchical way. Empirical analysis demonstrates that, in comparison with alternative methods, the proposed Bayesian approach achieves better results.

EO0772: **Curves registration in function-valued quantitative genetics**
Presenter: **Davide Pigoli**, University of Cambridge, United Kingdom
Co-authors: John Aston, Patrick Carter

Quantitative genetics explores inheritance patterns of continuous traits (phenotypes) that can be observed in a population of organisms. A quantity of interest is heritability, which is the proportion of the variability in the observed phenotypes that is due to genetic effects. Some of these phenotypes happen to be function-valued, examples include growth trajectories, mortality or fertility curves and shape functions. The covariance function of a function-valued phenotype needs therefore to be decomposed in its genetic and environmental components, but it is essential to address the problem of phase misalignment across individuals, each of which may have their own biological clock. However, it is often not straightforward to determine the correct phase-amplitude decomposition and this choice may impact the results. Moreover, the time warping function cannot be treated simply as a nuisance parameter, often being correlated with the shape of the functional data across individuals. We investigate the robustness of the heritability estimates with respect to time warping, using many well-established alternative methods for curve registration. Finally, we consider how to directly incorporate phase variability in the inheritance analysis. We demonstrate the proposed approach on growth curves from larval Tribolium castaneum (flour beetles).

EO284 Room Woburn RECENT DEVELOPMENTS IN SEMIPARAMETRIC ANALYSIS OF SURVIVAL DATA Chair: Liming Xiang

EO1428: **Variable selection for the Cox proportional hazard frailty model**
Presenter: **Ioanna Pelagia**, The University of Manchester, United Kingdom
Co-authors: Jianxin Pan

Extending the standard Cox Proportional Hazard (PH) model by adding frailty terms, can handle clustered data. However this might increase the dimension of variable components and become a challenging task in terms of estimation and significance of parameters. The proposed method is to overcome the problem of high dimension under the Cox PH Frailty model, by considering a simultaneous variable selection of both fixed effects and frailty components through penalty functions, such as Least Absolute Shrinkage and Selection Operator (LASSO) and Smoothly Clipped Absolute Deviation (SCAD). The proposed method has been examined through simulation studies and applied on real data of Kidney disease.

EO1478: **Semiparametric mixture modeling for multiphase survival data**

Presenter: **Liming Xiang**, Nanyang Technological University, Singapore

We propose a flexible regression approach for multiphase survival data using semiparametric mixture modeling. Our attention focuses on inference concerning the effects of covariates on both the mixing proportion and conditional hazard of each component, whereby the baseline hazard functions are completely unspecified. This approach is related to mixture models for competing risks data, but differs in that the group membership of each subject is unknown. We establish the identifiability of the proposed model. An expectation-conditional maximization (ECM) algorithm is developed for maximum likelihood estimation of parameters. The effectiveness of the proposed approach is demonstrated through simulation studies and a real data analysis.

EO1481: **Frailty modelling approaches for various survival data via hierarchical likelihood**

Presenter: **Il Do Ha**, Pukyong National University, Korea, South

Recently, frailty models have been widely studied to various survival data from clustered or competing-risks designs with censoring (left, right or interval) and truncation. We show the usefulness of semiparametric frailty models via hierarchical likelihood (or $h-$likelihood) for such various survival data. The $h-$likelihood obviates the need for intractable integrations over the frailty terms, whereas marginal likelihood usually does not. The proposed method is illustrated using time-to-event data from clinical studies.

Monday 14.12.2015 14:30 - 15:50 Parallel Session O – CFE-CMStatistics

CG507 Room Woburn CONTRIBUTIONS ON VOLATILITY Chair: Gernot Mueller

CC0184: **Modelling intraday volatility in European bond market**
Presenter: **Hanyu Zhang**, University of Reading, United Kingdom
Co-authors: Alfonso Dufour
Intraday volatility of European government bonds is studied under the framework of the multiplicative component GARCH model developed by Engle. High-frequency return volatility is specified as the product of daily volatility, intraday seasonality, and a unit GARCH process. Daily conditional variance is captured by the simple GARCH(1,1) and is shown to have a moderate correlation with high-frequency quoting activity. The usual assumption that conditional volatility reverts to a fixed value is relaxed by the intraday seasonal pattern, which is characterized by an exponential linear spline of time interval. The model is applied to ten-year government bonds of seven European countries and the sample period covers the sovereign debt crisis. The dynamics of intraday volatility vary significantly across countries. The periodic component of Spain and Italy is notably different from the ones of other sample countries. We observe large transitory intraday volatility of mid-quote prices often due to illiquidity effects and outliers. The estimation results are found to be strongly influenced by outliers and thus we highlight the importance of data cleaning. We suggest a flexible and effective procedure for jointly filtering mid-quote prices and estimating volatility models.

CC1603: **Volatility forecasting using global stochastic financial trends extracted from non-synchronous data**
Presenter: **Lyudmila Grigoryeva**, University of Konstanz, Germany
Co-authors: Juan-Pablo Ortega, Anatoly Peresetsky
We introduce a method based on the application of various linear and nonlinear state space models and that uses non-synchronous data to extract global stochastic financial trends (GST). These models are specifically constructed to take advantage of the intraday arrival of closing information coming from different international markets in order to improve the quality of volatility description and forecasting performances. A set of three major asynchronous international stock market indices is used in order to empirically show that this forecasting scheme is capable of significant performance improvements when compared with those obtained with standard models like the dynamic conditional correlation (DCC) family.

CC1531: **Information flows between the US and China's agricultural commodity futures markets-based on VAR-BEKK-skew-t model**
Presenter: **Qian Chen**, Peking University Shenzhen Campus, China
Co-authors: Xin Weng
A parametric approach is adopted to estimate and examine the price linkage and volatility spillover effects of corn, wheat and soybeans futures between the US and China's markets. Specifically, a VAR-BEKK-MGARCH model with multivariate skew-t error terms is proposed to capture the potential skewness and heavy-tails in the futures return series. The model is illustrated via application to real futures returns between 2005 and 2014. The empirical results indicate that the US market plays a major role in terms of mean and volatility information transmission over China's market, while China's market has increasing volatility spillovers to the US market in some highly marketised commodities after 2010. The results also imply that No.2 Soybeans in China may not be a good hedging vehicle for international risks.

CC1708: **Modelling conditional densities with asymmetric tails**
Presenter: **Steve Thiele**, Queensland University of Technology, Australia
A new time varying density model is proposed based on the DCS/GAS approach that allows for time varying volatility and different upper/lower tail thickness. The asymptotic properties of the model are presented together with simulations showing that such distributional features can be estimated with practical sample sizes. The importance of tail asymmetry for a range of equity index returns is first demonstrated in-sample relative to existing heavy tailed/skewed models. Application to risk management is then considered, with out-of-sample results for volatility, Value-at-Risk and expected shortfall forecasting shows strong performance relative to leading models from the literature.

CG287 Room Bedford CONTRIBUTIONS IN TIME SERIES AND APPLICATIONS Chair: Ansgar Belke

CC0343: **A new approach to measure the risk of a financial time series**
Presenter: **Tharanga Wickramarachchi**, Georgia southern University, United States
Co-authors: Colin Gallagher, Robert Lund
The necessity of more trustworthy methods for measuring the risk (volatility) of financial assets has come to the surface with the global market downturn. Nowadays investors are more vigilant when investments are made on markets. Therefore it is more of a requirement to figure out companies or sectors they should put on money so that the risk is minimized. We propose the arc length as a tool of quantifying the risk of a financial time series. As the main result, we prove the functional central limit theorem for the sample arc length of a multivariate time series under finite second moment conditions. The second moment conditions play a significant role since empirical evidence suggests that most of the asset returns have finite second moments, but infinite fourth moments. We show that the limit theory holds for a variety of popular models of log returns such as ARMA, GARCH and stochastic volatility model families. As an application, changepoints in the volatility of the Dow Jones Index is investigated using the CUSUM statistic based on the sample arc lengths. The simulation studies show that the arc length outperforms squared returns, which holds the functional central limit theorem only under finite fourth moment conditions and performs in a relatively similar manner as absolute returns. Comparison of time series in terms of volatility is also done as another application.

CC1418: **Short-term salmon price forecasting**
Presenter: **Daumantas Bloznelis**, Norwegian University of Life Sciences, Norway
Price forecasting has a fundamental value in the highly volatile market of farmed Atlantic salmon. Accurate price predictions could reduce uncertainty and aid planning decisions of salmon farmers, processors and other market participants. We forecast salmon spot prices by univariate and multivariate time series models, artificial neural networks and the $k-$nearest neighbours method using weekly data for 2007-2014. A naive seasonal forecast reduces the mean absolute error of a naive no-change forecast by up to 5%, and predicts the direction of the price change correctly up to 65% of the time for 1-5 weeks ahead. However, no method seems to consistently improve upon the naive seasonal forecast, lending support to a hypothesis of weak form efficiency of the salmon spot market.

CC1295: **On the applicability of advanced forecasting techniques to developing economies: A case of Sri Lanka**
Presenter: **Sujeetha Jegajeevan**, Queen Mary University of London, United Kingdom
Forecasting inflation and output has always been a key research area in macroeconomics. Empirical studies for advanced economies have a long history and the forecasting methodology is improved constantly in the recent years. These advanced forecasting techniques were successful in forecasting key economic variables precisely in advanced economies, at least before the recent financial crisis. Studies covering emerging and developing countries are just a handful and are only at the evolutionary stage. It is worthwhile applying these methods to a country like Sri Lanka to check their applicability to economies in transition. No study has been carried out to forecast output and inflation in Sri Lanka in the past. The aim is to fill the gap in the empirical literature. We employ 6 main forecasting techniques that are widely used for advanced economies in the

recent times. They are BVAR, Large-BVAR, Time Varying Parameter BVAR (TVP-BAVR), Mixed Frequency BVAR (MF-BVAR), DSGE and Unobserved Component model. The sample covers a period between 1996 and 2014. The forecasting exercise not only includes point forecasts but also density forecasts. Forecast accuracy of point forecasts is assessed by Root Mean Square Error (RMSE) statistic and that of density forecast is assessed by Log Score.

CC1377: **Adaptive forecasting in the presence of structural change and long memory persistence**
Presenter: **Mohaimen Mansur**, The University of Manchester, United Kingdom
Spurious relationship and consequent confusion between long memory and structural change have been widely documented in econometrics literature and pose critical challenge for forecasting time series that are subject to them. Failure to appropriately account for one or both the features that are truly present in the data can lead to large forecast errors. Given that it is difficult to distinguish between the two it is desirable to develop forecasting strategies that are robust to presence of both structural change and long range dependence. Such a strategy is proposed where future forecasts are weighted average of past data with the data down-weighting parameter selected adaptively through cross-validation. Detailed theoretical proofs of asymptotic optimality of such forecasts have been provided by considering discrete and continuous structural changes and both weak and strong long memory. Practical usefulness of the adaptive forecasts has been illustrated through an extensive Monte Carlo study and an application to many economic and financial time series of the UK.

CG531 Room Bloomsbury CONTRIBUTIONS IN PANEL DATA ECONOMETRICS Chair: Martin Wagner

CC0709: **Improving GMM efficiency in dynamic models for panel data with mean stationarity**
Presenter: **Giorgio Calzolari**, University of Firenze, Italy
Co-authors: Laura Magazzini, Giorgio Calzolari
Estimation of dynamic panel data models largely relies on GMM methods, and adopted sets of moment conditions exploit information up to the second moment of the variables. However, in many microeconomic applications the variables of interest are skewed, as for example in the analysis of individual wages, sizes of the firms, number of employees, etc.; therefore third moments might provide useful information for the estimation process. We propose a moment condition, to be added to the set of conditions customarily exploited in GMM estimation of dynamic panel data models, that is based on third moments. The moment condition we propose spans from the observation that, under mean stationarity, y_{i0} can be written as the long run mean of the dependent variable and a randomly distributed error term. As a result, the individual effect α_i enters the first observation y_{i0} multiplied by the same factor for each i (equal to $1/(1-\beta)$). This data generating process for y_{i0} is always adopted in the Monte Carlo simulations on dynamic panel data models that assume mean stationarity but not explicitly exploited. Monte Carlo experiments show remarkable efficiency improvements when the distribution of individual effects, and thus of y_{i0}, is indeed skewed.

CC1718: **Consistent estimation in dynamic panel data models with individual fixed effects**
Presenter: **Daniel Becker**, University of Bonn, Germany
We consider the estimation of dynamic panel data models in the presence of incidental parameters for individuals. A well-known result is that the within-group estimator is biased under fixed time dimension T. On the other hand, it is known that the asymptotic theory of GMM estimators breaks down if the number of moment conditions divided by T tends to a nonzero constant. We follow an approach using a factor analytical method. We propose estimation procedures for the model in levels and in differences. In levels we estimate the sample variance of individual effects rather than the effects themselves. We control for the initial values and correlation between the individual effects and the initial values as well as the correlation between the fixed effects and the additional exogenous variables. In differences the individual effects cancel out and we only need to control for the initial values. The estimators are consistent if N tends to infinity whether or not T is fixed or tends to infinity. With the proposed estimators we get rid of the well known incidental parameter problem. Our method is easy to apply, has good small sample properties and is asymptotically efficient.

CC0196: **Panel data tests for bubbles with an application to OECD housing prices**
Presenter: **Itamar Caspi**, Bar-Ilan University and Bank of Israel, Israel
Methods are introduced for testing the null of unit root against the alternative of a mildly explosive root in dynamic panels. Rejection of the null in this case is interpreted as evidence for a bubble in some of the panel members. To this end we specify two tests: The first is for sporadic explosive periods in panel members, based on the average of individual sup ADF (SADF) statistics. The second test is for synchronized explosive periods, based on a recursive right tailed variation of a previous panel unit root test. The limiting distributions of both statistics is derived under the assumption of cross sectional independence and are shown to converge to the standard normal distribution, as the time and cross sectional dimensions go to infinity sequentially. Monte Carlo simulations show that both tests have good finite sample properties. An empirical illustration is provided by applying both tests to a panel of real house price index series as well as house price to rent ratios for 22 OECD countries for the period of 1998:Q1-2014:Q1.

CC1408: **Instrumental variable estimation of panel data models with weakly exogenous variables**
Presenter: **Kazuhiko Hayakawa**, Hiroshima University, Japan
Co-authors: Joerg Breitung, Meng Qi
We propose an instrumental variables estimator for panel data models with weakly exogenous variables. The model is allowed to include heterogeneous time trends besides the standard fixed effects. The proposed instrumental variable estimator is constructed by removing the fixed effects (and time trends) from both the model and instruments by a variant of GLS transformation. We show that the proposed estimator has the same asymptotic distribution as the bias corrected fixed effects estimator when both N and T, the dimensions of cross section and time series, are large. Monte Carlo simulation results reveal that the proposed estimator performs well in finite samples and outperforms the conventional IV/GMM estimators using instruments in levels in many cases.

CG485 Room Holden CONTRIBUTIONS ON FISCAL POLICY Chair: Peter Claeys

CC1077: **A new measure to quantify the effects of U.S. tax policy news**
Presenter: **Tom Philipp Dybowski**, University of Muenster, Germany
Co-authors: Jan Nikolaj Dybowski
Recent empirical research on tax policy proposes to expand the information set of the econometrician to solve the problems associated with foresight. Yet, variables with informational content on future tax policies are scarce and only exist for the U.S. We propose a novel approach to construct a variable that expands and remedies some drawbacks of previous measures. By using a probabilistic topic model, known as LDA ('Latent Dirichlet Allocation'), we investigate the tax policy contents of over 96.000 Presidential documents from 1945 to 2015; an insuperable task for human capacity. This allows us to obtain a probability for each document to cover tax related issues and to construct a time series, labeled as 'tax news'. The time-dynamics of this variable shows high consistency with documented U.S. tax reforms. We apply this tax news variable to expand the information set of a well-known SVAR model. Our results suggest that the tax news variable can be interpreted as an anticipated tax

shock. The attractiveness of our approach is its applicability. In contrast to the few existing measures, our tax news variable can be extracted for any country in which fiscal authorities publish a suitable corpus of tax related documents.

CC1280: **Debt maturity, monetary policy and fiscal sustainability without commitment**
Presenter: **Dmitry Matveev**, University of Mannheim, Germany
The aim is to study how the speed of optimal government debt adjustment and the monetary-fiscal policy mix that implements it depend on the maturity structure of debt when policy is chosen discretionary. Under the assumption of debt taking the form of one-period nominal bonds, for plausible levels of debt, fiscal sustainability requires prompt adjustment of debt and monetary policy bears a significant burden of implementing the adjustment. Higher average maturity reduces both the incentive of the government to alter current policy and the incentive to strategically affect future self so as to improve the price of borrowing. Accounting for a plausible average maturity makes the optimal debt adjustment much more gradual, which is in line with the existing empirical evidence on the persistence of government debt. In the case of bond portfolios with the average maturity ranging from several years and higher, it is no longer optimal for monetary policy to accommodate debt adjustment.

CC1759: **An econometric analysis of the flat tax revolution**
Presenter: **Wayne Tarrant**, Rose-Hulman Institute of Technology, United States
The flat tax goes back to at least the Biblical tithe. A progressive income tax was first espoused in a small, but famous, pamphlet in 1848 (although England had an emergency progressive tax for war costs prior to this). Within a few years many countries had adopted the progressive structure. The flat tax was only reinstated in some small countries and British protectorates until Mart Laar was elected Prime Minister of Estonia in 1992. Since Estonia's adoption of the flat tax in 1993, many formerly Communist countries have abandoned progressive income taxes. Economists had expectations of what would happen when a flat tax was enacted, but very little work has been done on actually measuring the effect. With a test bed of 21 countries in this region that currently have a flat tax, much comparison is possible. Several countries have retained progressive taxes, giving an opportunity for contrast. There are also the cases of Czech Republic and Slovakia, which have adopted and later abandoned the flat tax. Further, with over 20 years worth of economic history in some flat tax countries, we can begin to do some serious longitudinal study. We consider many economic variables to determine if there are differences due to the adoption of a flat tax. We consider unemployment rates, tax receipts, GDP growth, Gini coefficients, and market data, where such data are available. The results are mixed, but we draw statistically significant conclusions about some effects.

CC0287: **Japanese fiscal policy under the zero lower bound of nominal interest rates**
Presenter: **Hiroshi Morita**, Hitotsubashi University, Japan
The time-varying effects of fiscal policy under the zero lower bound of nominal interest rates are examined. In particular, we present a new strategy for identifying unconventional monetary policy shocks in the framework of a time-varying parameters vector autoregressive model. We characterize monetary policy shocks at the zero lower bound (ZLB) of short-term interest rates by combining the zero restrictions on the interest rate equation with the sign restrictions. Furthermore, we investigate whether the effects of fiscal policy are enhanced during ZLB periods, as predicted by theory. The main findings are as follows. First, during ZLB periods, the volatility of short-term interest rates is quite small, while that of the monetary base is large. Second, fiscal policy shocks have significant positive time-varying effects on GDP after adoption of unconventional monetary policy. On the other hand, fiscal news shocks and monetary policy have significant effects on GDP. Third, the effects of fiscal policy shocks increase during a ZLB period.

CG597 Room Chancellor's Hall CONTRIBUTIONS ON TEMPORAL AND SPATIAL ECONOMETRIC Chair: Jan PAM Jacobs

CC1520: **Spatial econometrics in Web 2.0: A method for determining the economic entropy in UE28**
Presenter: **Calin-Adrian Comes**, Petru Maior University of Tirgu-Mures, Romania
Issues of major interest from a regional and national policy perspective are addressed. The main objective is to present factual data about relevant variables that can be transformed into regional capabilities that produce sustainable and inclusive local growth. The quantitative research design aims to highlight and to explain regional capabilities that can foster the process of wealth creation by the means of diversification and specialization. Regional growth is a cumulative, spatial selective process that can amplify inequalities existent in regional markets. Initial conditions, interactions and interdependences are relevant for predicting present and future growth patterns. In the absence of counteractive policies that target market imperfections market forces will encourage regional divergence, and thus the deepening of inequalities. By providing an interactive map of regional capabilities the aim is to present a practical and applicable dimension that supports the process of decision making, for various stakeholders, thus contributing in a significant manner for maximizing regional wealth. The elements of regional economic capabilities are emphasized by appealing to interdisciplinary research using the following research methods: data mining, stored procedures, web services on application server, spatial econometrics concepts and approaches that present the regional potential through entropic indicators reflecting the links between.

CC0506: **Bayesian variable selection in spatial autoregressive models**
Presenter: **Philipp Piribauer**, Vienna University of Economics and Business, Austria
Co-authors: Jesus Crespo Cuaresma
Recently, Bayesian model averaging approaches have gained momentum in order to deal with the problem of model uncertainty by producing parameter inference unconditional on model specification issues. For spatial autoregressive models, however, standard Bayesian model averaging techniques involve the computation of Bayesian marginal likelihoods, which do not have closed form solutions. When the degree of uncertainty is large, model averaging thus leads to a severe computational burden, since the calculation of Bayesian marginal likelihoods require numerical integration techniques. Two alternative Bayesian variable selection approaches for spatial autoregressive models are put forward. Both approaches allow us to deal with the problem of model uncertainty in spatial autoregressive models in a very flexible and computationally efficient way, since the approaches can be implemented in a Gibbs sampling algorithm in a straightforward way. In a simulation study it is shown that the variable selection approaches outperform existing Bayesian model averaging techniques both in terms of computational efficiency as well as in terms of in-sample predictive performance.

CC1822: **Global high-resolution land-use change projections: A Bayesian multinomial logit approach with model uncertainty**
Presenter: **Tamas Krisztin**, IIASA, Austria
Co-authors: Petr Havlik, David Leclere, Ines Moreau
Using econometric models to estimate land-use change has a long tradition in literature. Recent contributions show the importance of including spatial information and of using a multinomial framework to take into account the inter-dependencies between the land-use classes. Few studies, however, agree on the relevant determinants of land-use change and there are no contributions so far comparing determinants on a global scale. Using multiple 5 arcminute resolution datasets of land use change between 2000 and 2010 and taking into account the transitions between forest, cropland, grassland land covers, we estimate a Bayesian multinomial logit model, using a Polya-Gamma sampling procedure. To identify and measure the determinants of land-use change and the strength of spatial separation, our model implements Bayesian model selection through SSVS priors and spatial information via Gaussian Process priors. Our results indicate that spatial proximity is of central importance in land-use change, in all regions except the pacific islands. We also show that infrastructure policy, proxied by mean time to market, seems to have a significant impact

on deforestation throughout most regions. In a second step we use aggregate, supra national land use change results from the partial equilibrium agricultural model GLOBIOM as a framework for projecting our model in ten-year intervals up to 2100 on a spatially explicit scale along multiple shared socioeconomic pathways.

CC1723: **An oblique projection approach to consumer price level prediction**
Presenter: **Christian Heinze**, University Bielefeld, Germany
Consumer price index data suitable for spatial comparison at a high resolution is scarce in many countries but of great interest to regional economists. We focus on German countries and present predictions of suitable consumer price indexes for the years 1993-2012. For this case, the available price index data comprise a small cross-section for 1993 and annual inflation rates for a subset of the German states. In a first step, we construct an auxiliary panel of pseudo price indexes in form of predictions from a low-complexity regression model. Therein, parameter estimates are based on the above mentioned data and corresponding covariate observations. Secondly, the pseudo index panel is used to estimate a VAR transition matrix - using nuclear norm regularization to enforce a low rank - alongside an innovation covariance matrix. Finally, the implied spatio-temporal covariance structure allows Kridge-type prediction - implemented as Kalman filtering - of the missing price indexes based on the available price indexes. This approach amounts to replacing the covariances of an (almost) unobserved process by those of a related and observed (pseudo-)process. This has an interpretation as an oblique projection. Therefore, the angles between certain subspaces characterize its inferiority to the corresponding orthogonal projection. We also present an expression in terms of the two processes.

CC027 Room Montague CONTRIBUTIONS IN BAYESIAN ECONOMETRICS Chair: Tom Shively

CC1599: **Exchange rate predictive densities under model uncertainty**
Presenter: **Anthony Garratt**, University of Warwick, United Kingdom
Co-authors: Emi Mise, Rodney Strachan
We re-examine the forecast performance of a well-known set of fundamentals based (uncovered interest parity, purchasing power parity, monetary fundamentals, forward premium, Taylor rules) exchange rate models but where we ask; does the random walk model dominate out-of-sample forecasts of the conditional predictive density, as opposed to point forecasts? We incorporate the fundamentals based models into a VECM framework, impose theory based long-run restrictions on the error correcting or disequilibrium term and then estimate each of these VECM's using standard Bayesian time-varying stochastic volatility techniques. A feature of this estimation is the use of precision sampling, as opposed to the kalman filter, which is computationally more efficient. Predictive densities are combined utilizing a set of weights computed using a linear mixture of experts framework (linear and logarithmic opinion pools). In our application, to monthly data for three currencies (the Euro, Pound and Yen relative to the dollar) for the period 1999m1-2014m4, we seek to examine if the combined or single model forecast densities produce well calibrated densities and how they perform relative to a random walk. Forecast evaluation will take the form of both statistical and economic evaluation.

CC1524: **Bayesian cointegration using a singular distribution on the long-run relations matrix**
Presenter: **Basile Marquier**, University of Sheffield, United Kingdom
The focus is on new methods of cointegration to establish convergence of economies, their performance, co-evolution and long-run relationships. Multivariate cointegration methods have dominated the econometrics literature over the past 20 years. They offer a framework of identifying relationships of financial assets or exchange rates or more generally of financial time series, hence they are exploited in developing long term decision making, trading and portfolio management. Over the last decade Bayesian methods have been developed for the purpose of estimating the parameters of the well established vector error correction (VEC) model and thus discovering cointegrated time series. By using Markov chain Monte Carlo methods, we propose a Gibbs sampling scheme for estimation of the parameters of the VEC model. The long- run relationships matrix is a square matrix of low rank, called the cointegration rank. This rank is the number of independent cointegration relations between the economic time series. A critical step in the analysis of the cointegration rank will be to try to avoid reliance upon Johansen tests. A singular normal prior is set over the long-run relations matrix. The proposed methodology is illustrated with a panel data of economic and financial variables and a study of the cointegrating coefficients eventually provides information about positive and negative comovements.

CC1434: **Approximate Bayesian inference with pseudo-likelihood**
Presenter: **Ray S W Chung**, The Hong Kong University of Science and Technology, China
Co-authors: Mike So
Bayesian inference can effectively deal with a wide range of complicated statistical problems like high-dimensional inference, latent variable filtering and statistical learning. In classical Bayesian analysis, we need to fully specify the likelihood of underlying models so as to carry out statistical computation for posterior inference. The requirement of likelihood limits the application of Bayesian approach in solving semi-parametric problems or problems whose full likelihood is computationally intractable. We propose an approximate Bayesian inference framework which incorporates pseudo-likelihood. It is expected that without the need of full likelihood specification, we can extend the scope of problems which Bayesian inference can solve. Two examples, the GARCH model and the spatiotemporal model, are taken to demonstrate our framework. Results in the examples show that the approximate Bayesian inference framework can provide both consistent estimates as well as good credible interval coverage.

CC0207: **Markov interacting importance samplers**
Presenter: **Marcel Scharth**, University of Sydney, Australia
Co-authors: Eduardo Mendes, Robert Kohn
A new Markov chain Monte Carlo (MCMC) sampler called the Markov Interacting Importance Sampler (MIIS) is introduced. The MIIS sampler uses conditional importance sampling (IS) approximations to jointly sample the current state of the Markov Chain and estimate conditional expectations, possibly by incorporating a full range of variance reduction techniques. We compute Rao-Blackwellized estimates based on the conditional expectations to construct control variates for estimating expectations under the target distribution. The control variates are particularly efficient when there are substantial correlations between the variables in the target distribution, a challenging setting for MCMC. An important motivating application of MIIS occurs when the exact Gibbs sampler is not available because it is infeasible to directly simulate from the conditional distributions. In this case the MIIS method can be much more efficient than a Metropolis-within-Gibbs approach. We also introduce the MIIS random walk algorithm, designed to accelerate convergence and improve upon the computational efficiency of standard random walk samplers. Simulated and empirical illustrations for Bayesian analysis show that the method significantly reduces the variance of Monte Carlo estimates compared to standard MCMC approaches, at equivalent implementation and computational effort.

CG569 Room Athlone CONTRIBUTIONS ON PRICING Chair: Lorenzo Mercuri

CC1745: **The pricing of idiosyncratic risk in option markets**
Presenter: **Christian Dorion**, HEC Montreal, Canada
Co-authors: Jean-Francois Begin, Genevieve Gauthier
The recent literature provides conflicting empirical evidence on the sign of the relationship between idiosyncratic risk and equity returns. New light is shed on this relationship by disentangling four risk factors contributing to the equity risk premium. We exploit the richness of stock option data to extract the expected risk premium associated each risk factor, thereby avoiding the exclusive use of noisy realizations of historical returns. To this end, we develop a jump-diffusion model in which a firms systematic and idiosyncratic risk have both a diffusive and a tail component. Our pricing kernel is such that each risk factor can potentially be priced. This model offers quasi-closed form solutions for the price of European options. We estimate the model on 117 firms that are or were part the S&P 500 index, using equity returns and option prices of the market index and of each individual firm. Our analysis highlights two new results. First, we find that idiosyncratic jump risk is priced to a greater extent than systematic risk. Idiosyncratic jump risk accounts for 55% of the expected equity premium, on average, compared to 45% for systematic risk (13% for the diffusive part and 32% for the tail risk). Second, we show that the diffusive part of idiosyncratic risk is not priced, once other sources of risk are accounted for. These empirical findings thus contribute to the understanding of the complex relation between idiosyncratic risk and equity risk premium.

CC0181: **A nonparametric option pricing model using higher moments**
Presenter: **Peter Julian Cayton**, The Australian National University, Australia
Co-authors: Kin-Yip Ho
A nonparametric model that includes non-Gaussian characteristics of skewness and kurtosis is proposed based on the cubic market capital asset pricing model. It is an equilibrium pricing model but risk-neutral valuation can be introduced through return data transformation. The model complies with the put-call parity principle of option pricing theory. The properties of the model are studied through simulation methods and compared with the Black-Scholes model. Simulation scenarios include cases on nonnormality in skewness and kurtosis, nonconstant variance, moneyness, contract duration, and interest rate levels. The proposed model can have negative prices in cases of out-of-money options and in simulation cases that are different from real-market situations, but the frequency of negative prices is reduced when risk-neutral valuation is implemented. The model is more adaptive and more conservative in pricing options compared to the Black-Scholes model when nonnormalities exist in the returns data.

CC1507: **Fourier transform in option pricing models**
Presenter: **Marek Kolman**, University of Economics Prague, Czech Republic
Various option pricing models are compared in within the scope of Fourier transform. It is first shown that there exists a general approach to obtain a characteristic function using a deterministic term called 'Jensen correction' ω, which makes Fourier transform methods very attractive. This Jensen correction term is a powerful tool that allows to obtain characteristic functions quickly and with almost no costs. Subsequently, we compare the Madan-Carr approach with a recent transform approach relying on Fourier-Cosine expansion where we also define a more applicable technique for bounds selection. The conclusion is that the Fourier-Cosine is very fast and easy-to-implement method which can easily outperform the standard Madan-Carr approach. With the introduced bounds selection also stable and accurate transform method for option pricing.

CC1421: **Valuing American options using fast recursive projections**
Presenter: **Paola Pederzoli**, University of Geneva Swiss Finance Institute, Switzerland
Co-authors: Antonio Cosma, Olivier Scaillet, Stefano Galluccio
A new numerical option pricing method by fast recursive projections is introduced. The projection step consists in representing the payoff and the state price density with a fast discrete transform based on a simple grid sampling. The recursive step consists in transmitting the coefficients of the representation from one date to the previous one by an explicit recursion formula. We characterise the convergence rate of the computed option price. Numerical illustrations with different American and Bermudan payoffs with discrete dividend paying stocks in the Black Scholes and Heston models show that the method is faster than comparable methods, accurate, and general. As a notable application, we characterise the early exercise boundary of an American call option on a dividend paying stock. The boundary is higher under the Merton and Heston model than under Black Scholes if the dividend is discrete, and lower in the case of a continuous dividend yield. We apply our method to study a large sample of call options on stocks with quarterly dividends during the period January 1996 through December 2012. We illustrate that the choice of the model for the underlying asset is important for both the early exercise decision, in accordance with our numerical results, and for the quantification of the dollar amount that the buyer of the option forgoes when he fails to optimally exercise to the advantage of the seller.

CG359 Room SH349 CONTRIBUTIONS ON THE ESTIMATION OF DSGE MODELS Chair: Karl Schmedders

CC1206: **Stochastic volatility for asset pricing and business cycles**
Presenter: **Oliver de Groot**, University of St Andrews, United Kingdom
A method is described to find a risk-adjusted first-order approximation for DSGE models with recursive preferences and stochastic volatility. The method can be applied to a wide class of DSGE models and does not require the primitive shocks to be log-normal. The method is tested using a simple endowment-asset pricing model with stochastic volatility for which I am able to provide an exact closed-form solution. The method captures both the first-order effects (e.g. time varying asset return premiums) of stochastic volatility for which standard perturbation methods require a third-order approximation, and the zeroth-order effects (e.g. on unconditional mean asset return premiums) of stochastic volatility for which standard perturbation methods require higher-than-third-order approximation.

CC1577: **Indeterminacy, misspecification and forecastability**
Presenter: **Marco Sorge**, University of Goettingen, Germany
Co-authors: Luca Fanelli
This paper explores the consequences of equilibrium indeterminacy in DSGE models for macroeconomic forecasting. Exploiting U.S. data on both the Great Moderation and the preceding era, we first present evidence that (i) higher (absolute) forecastability obtains in the former rather than the latter period for all models considered, and that (ii) the decline in data volatility and persistence captured across the two samples need not be associated with inferior predictive accuracy relative to univariate predictors. Then, using a small-scale New Keynesian model as laboratory, we investigate numerically whether forecastability is improved in the presence of indeterminacy. It is argued that forecasting under indeterminacy with e.g. unrestricted VAR models entails misspecification issues that are generally more severe than those one typically faces under determinacy. Irrespective of the occurrence of sunspot noise, for certain values of the arbitrary parameters governing solution multiplicity, pseudo out-of-sample VAR forecasting can outperform simple univariate predictors. For other values of these parameters, by contrast, the opposite occurs. In general, it is not possible to establish a one-to-one mapping between indeterminacy and superior forecastability, even when sunspot shocks play no role. Overall, our analysis points towards a 'good luck in bad policy' explanation of the (relative) higher forecastability of macro-econometric models prior to the Great Moderation period.

CC1763: **Estimation of a DSGE model with heterogeneous agents using an indirect inference estimator**
Presenter: **Artem Duplinskiy**, VU University Amsterdam, Netherlands
Co-authors: Alexey Gorn

Policy makers use macro models with individuals being the same in terms of their level of income and skill. Models with heterogeneous agents allow individuals to have different starting endowments (wealth) as well as different income (not everyone is employed). These models are more difficult so researchers use simulation based solution methods to solve them and calibrate the parameters. Without estimation, it is hard to tell which of the new models represents the data better. Calibration takes away the uncertainty regarding parameter values necessary for meaningful forecasts. We asked whether an indirect inference estimator could be used to estimate a simple dynamic stochastic general equilibrium model with heterogeneous agents. This estimator naturally fits to the task since it requires only that one can simulate data from the economic model for different values of its parameters. We use macro variables together with the dynamics of the income distribution to estimate the model. A Monte Carlo study reveals the importance of the income distribution data to identify the parameters. The data application highlights that the model picks up precautionary savings effect observed in the data, suggesting that II estimator can be used to fit new models to the data and advance macroeconomic research.

CC1717: **Bootstrapping DSGE models**
Presenter: **Luca Fanelli**, University of Bologna, Italy
Co-authors: Giovanni Angelini

The rational expectations hypothesis upon which dynamic stochastic general equilibrium (DSGE) models are built implies highly nonlinear restrictions on the autocorrelation structure of the data, which are usually rejected. These restrictions, hereafter denoted cross-equation restrictions (CER), represent the natural metric through which DSGE models should be evaluated empirically. The aim is: (i) to propose a nonparametric bootstrap algorithm for the LR test of the CER implied by DSGE models, which generalizes a previous approach; (ii) to propose a (mis)specification test for the null hypothesis of strong identification of the structural parameters based on the distance between the bootstrap distribution of the LR test for the CER (or of the maximum likelihood estimator of the structural parameters) and the chi-distribution (the Gaussian distribution); (iii) to investigate the finite sample size performance of the test developed in (i) and the finite sample size and power of the test developed in (ii) through a set of Monte Carlo experiments, considering the case of strong identification and the case of weak identification of the structural parameters, respectively; (iv) to provide an empirical illustration of the suggested approach, based on a well-known monetary DSGE model estimated on U.S. quarterly data.

CG553 Room Court CONTRIBUTIONS ON APPLIED ECONOMETRICS AND FINANCE Chair: Thanasis Stengos

CC1364: **On IV estimation of a dynamic linear probability model with fixed effects**
Presenter: **Andrew Pua**, Universiteit van Amsterdam, Netherlands

Research workers still estimate a dynamic linear probability model (LPM) with fixed effects when analyzing a panel of binary choices. Setting aside the possibility that the average marginal effect may not be point-identified, directly applying IV estimators to this dynamic LPM delivers inconsistent estimators for the true average marginal effect regardless of whether the cross-sectional or time series dimension grow large. We show through some examples that these inconsistent estimators are sometimes outside the nonparametric bounds proposed previously. Although there are no analytical results for GMM estimators using Arellano-Bond moment conditions, we show through an empirical example that the resulting GMM estimate of the average treatment effect of fertility on female labor participation is outside the nonparametric bounds under monotonicity.

CC1450: **The financial stability dark side of monetary policy**
Presenter: **Fabrizio Venditti**, Queen Mary University of London, United Kingdom
Co-authors: Piergiorgio Alessandri, Antonio Conti

Recent papers argue that the expansionary effect stemming from a compression of credit spreads might be smaller than the recessionary impact of their reversal. Expansionary policy shocks would then stimulate the economy in the short-run but increase macroeconomic volatility in the longer run, thus implying the existence of a so-called financial stability dark side associated to monetary interventions. Such a mechanism would create a trade-off between monetary policy and financial stability objectives. The empirical relationship between the dynamics of credit spreads and those of output in the U.S. and in the Euro area is examined, focusing on the potentially asymmetric impact of rising and falling spreads on economic activity. On the basis of predictive regressions and of structural VARs that allow for asymmetric effects of structural shocks we find that for the U.S. the effect of credit spreads on output is significantly asymmetric only for shocks of large magnitude. For the Euro area as a whole, Germany, France and Italy, evidence of asymmetric effects is generally very weak. Based on this evidence, the trade-off between monetary policy and financial stability objectives arising through this channel seems to be tenuous in normal times, but could become more relevant in the case of large shocks. In terms of policy prescriptions, our findings call for a gradual and carefully designed exit strategy from extraordinary monetary accommodation.

CC1361: **Assessing capital-based macroprudential policy using an integrated early warning GVAR model**
Presenter: **Markus Behn**, European Central Bank, Germany
Co-authors: Marco Gross, Tuomas Peltonen

We develop an integrated early warning global vector autoregressive (EW-GVAR) model to quantify the cost and benefits of capital-based macroprudential policy measures in a forward-looking manner. The model contains a logistic component to predict the probability of banking sector crises and a GVAR component to capture the dynamics of credit and other macro-financial variables in response to capital ratio shocks, thus endogenising the predictor variables used in the logistic early warning model. Our setup helps uncovering the transmission channels of capital-based measures and also allows considering cross-country spillover effects when assessing the net benefits of macroprudential measures. Our model may be used to achieve a more precise calibration of capital-based measures, thus helping to contain systemic risk while at the same time accounting for potential costs of macroprudential policy.

CC1828: **Comparative analysis of greedy and lasso-type algorithms for index tracking problems**
Presenter: **Sergei Sidorov**, Saratov State University, Russia

We compare the performance of two different portfolios obtained by means of two algorithms (greedy and LASSO-type algorithms) for index tracking error minimization in L2-norm. The index tracking problem with cardinality constraint is NP-hard problem and it usually requires the development of heuristic algorithms. Greedy algorithms also proved their effectiveness. In our empirical analysis we use publicly available data relating to five major market indices, that can be obtained from the OR-Library. The five market indices are the Hang Seng (Hong Kong, n = 31), DAX 100 (Germany, n = 85), FTSE 100 (UK, n = 89), S&P 100 (USA, n = 98) and the Nikkei 225 (Japan, n = 225) for m = 290 time periods each (weekly data). Empirical tests show that the LASSO-type approach provides better performance of tracking portfolios in terms of returns, while the greedy algorithm leads to better portfolios with respect to tracking error volatilities.

CG483 Room Senate CONTRIBUTIONS ON FORECASTING Chair: Nikolaos Kourogenis

CC1401: **Forecasting volatility with multiple horizon extreme values**
Presenter: **Ray Chou**, Academia Sinica, Taiwan
A stylized fact in empirical finance is that while volatility clusters, it reverts to its long term mean quickly, usually within months. This is especially true for models building upon intra-daily or inter-daily data of asset prices. In other words, most volatility forecasting models are of little use (if at all) in financial applications related to long (month or beyond) horizons. Extreme asset prices (high/low range) of multiple horizons are exploited. The forecasting strength of the range data comes from two facts: model free and high persistence. Unlike square returns, absolute returns, RVs or other alternatives, range is model free. Hence, it is free from model uncertainty, parameter estimation errors, microstructure bias errors and errors induced by pre-filtering of the raw data. Long horizon range also helps to preserve the persistence inherent in volatility shocks. We develop a range-based multiple horizon volatility model to characterize the dynamics of asset prices in the first two moments and the model is used for forecasting in comparison with competing models, e.g., GARCH, HAR, RV, MEM and realized GARCH. Preliminary empirical results are supportive, i.e., in contrast to the common beliefs, reliable volatility forecasts beyond a month are feasible. Potential applications are discussed.

CC1587: **Comparing density forecasts in a risk management context**
Presenter: **Hao Fang**, University of Amsterdam, Netherlands
Co-authors: Cees Diks
A testing framework is developed for comparing the accuracy of competing densities of aggregated marginal variables in the downside part of the support. Three proper scoring rules including conditional likelihood, censored likelihood and penalized weighted likelihood are used for assessing the predictive ability of (out-of-sample) densities, closely related to the Kullback-Leibler information criterion (KLIC). We consider distributions in the framework of skew-elliptical family which is analytically traceable under affine transformations. The common practice of forecast comparison in high-dimensional space is problematic because that a better forecast in multivariate evaluation does not necessarily correspond to a better portfolio return forecast, as illustrated by examples. An application to the daily returns of three US stocks suggests that the Student-t outperforms the Normal, Skew t and Skew Normal assumptions in the left tail of the portfolio return. The visualized dynamics of our test statistic provides a side proof for regime change over last thirty years.

CC1511: **Computational forecasting with FRED**
Presenter: **Charles Rahal**, University of Birmingham, United Kingdom
We develop a new dynamic 'big data' forecasting algorithm which uses 128,732 series from the Federal Reserve Economic Data (FRED) database. We evaluate it through a pseudo real-time forecasting exercise involving 50 all-transactions house price indexes at US state level, with an out-of-sample period across the recent Great Recession. We estimate approximately 482 million bivariate models drawn from a cross-section of dimension sparsely seen in the econometric literature, selecting and weighting variables across different horizons, forecast approaches, model set sizes, rolling and recursive windows and combination metrics. Our approach performs extremely favourably against each of the three commonly used benchmarks which we evaluate it against, including a comparable big data macroeconometric (FAVAR) model which is outperformed in approximately 80% of instances. We aim to provide several stylized results to guide big-data forecasting in macroeconometrics and related fields, with the algorithm showing a general preference for Bayesian Model Averaging over Smooth Akaike Information Criteria weightings, iterated multi-step over dynamic multi-step forecast generation and for recursive over rolling estimation windows. There is further strong evidence of forecast error increasing in the volatility of underlying series, and we show that our forecast errors are relatively stable, except for during the recent financial crisis.

CC1668: **A comprehensive dynamic Bayesian model combination approach to forecasting equity premia**
Presenter: **Rainer Schuessler**, Helmut Schmidt University Hamburg, Germany
Co-authors: Joscha Beckmann
We introduce a novel dynamic Bayesian model combination approach for predicting aggregate stock returns. Our method involves combining predictive densities in a data-adaptive fashion and simultaneously features (i) uncertainty about relevant predictor variables, (ii) parameter instability, (iii) time-varying volatility, (iv) time-varying model weights and (v) multivariate information. We analyze the predictability of monthly S&P 500 returns and disentangle which components of prediction models pay off in terms of statistical accuracy and economic utility. As a key feature of our approach, we formally address the (possibly) diminishing relevance of past information over time. The flexibility embedded in our approach enhances density forecasting accuracy and provides sizeable economic utility gains. We find predictability to be strongly tied to business cycle fluctuations and document disagreement between statistical and economic metrics of forecast performance.

CG395 Room Jessel CONTRIBUTIONS ON HIGH-FRECUENCY DATA Chair: Kim Christensen

CC1479: **Layering the order book: Quoting activity during the flash crash**
Presenter: **Eric Aldrich**, University of California Santa Cruz, United States
Co-authors: Joseph Grundfest, Gregory Laughlin
We present a data-driven re-examination of the Flash Crash, using millisecond-stamped quote and trade data from the Chicago Mercantile exchange (CME), augmented by nanosecond resolution data recorded at the Nasdaq Exchange. We set the Flash Crash event into context by comparing the market state on May 6, 2010 with measures of market microstructure recorded on a large number of preceding trading days in 2010. We also draw particular comparisons with August 9, 2011, during which the CME recorded its highest-ever daytime messaging traffic. Our analysis leads to a number of conclusions. (1) Volume and messaging at the CME on May 6, 2010, while high, were not unprecedented. (2) Liquidity and price correlation between the CME and Nasdaq exchanges only experienced a breakdown after the CME market stop, which ended the crash itself. (3) The structure of the CME E-mini limit order book on May 6th was highly unusual, exhibiting large imbalances between the number of resting sell orders and resting buy orders, but only at price levels deep in the book. We find little evidence, however, that these imbalances are correlated with subsequent price movements. (4) The Flash Crash itself unfolded on the CME in an orderly and disciplined manner. Our interpretation is that high-frequency traders acted as both market makers and takers, and imposed a steady decay in prices during a period in which there was effectively no significant external buy-side interest from traditional market participants.

CC1595: **Detecting infinitesimal lead-lag effects from ultra high frequency data**
Presenter: **Yuta Koike**, Institute of Statistical Mathematics, Japan
We propose a framework for modeling the lead-lag effect between two assets at ultra high frequencies, which can be accommodated to non-synchronous trading and market microstructure noise as well as the heterogeneity of lead-lag effects. We show that this framework enables us to conduct tests to detect the presence or the heterogeneity of the lead-lag effect. The methodology is illustrated by an empirical study to detect a lead-lag effect in multi-market trading.

CC1695: **Efficient multipowers**
Presenter: **Aleksey Kolokolov**, Lund University, Sweden
Co-authors: Roberto Reno
We show that the asymptotic variance of realized multipower estimators can be reduced considerably by combining different estimators and minimizing with respect to powers. After providing a general criterion for efficiency, we propose a new quarticity estimator which, thanks to the smaller variance, has the smallest mean square error among popular competitors on realistic simulations of financial prices. The application on US stocks provides further evidence of the superiority of the new estimator with respect to competitors, and in particular it shows that i) consistently with the theory, the distribution of the newly proposed quarticity estimator has smaller variance and thinner tails; ii) the new quarticity estimator does not suffer of the distortions due to the presence of jumps displayed by competitors; iii) using the new quarticity estimator instead of realized multipower variation in a standard test for the presence of jumps in asset prices (based on the difference between realized variance and bipower variation) improves the size and the power of the test.

CC1483: **On effects of jump and noise in high-frequency financial econometrics**
Presenter: **Daisuke Kurisu**, University of Tokyo, Japan
Co-authors: Naoto Kunitomo
Several new statistical procedures for high frequency financial data analysis have been developed for estimating risk quantities and testing the presence of jump in the underlying continuous-time financial processes. Although the role of micro-market noise is important in high frequency financial data, there are basic questions on the effects of presence of noise and jump in the underlying stochastic processes. When there can be jump and (micro-market) noise at the same time, it is not obvious whether the existing statistical methods are reliable or not for the applications in actual data analysis. We investigate the misspecification effects of jump and noise on some basic statistics and the testing procedures for jumps previously proposed. We have found that their first test is asymptotically robust in the small-noise asymptotic sense against possible misspecification while their second test is quite sensitive to the presence of noise.

CG626 Room Torrington CONTRIBUTIONS ON CREDIT RISK MODELLING Chair: Arvid Raknerud

CC1564: **Dynamic credit risk model of a large consumer portfolio**
Presenter: **Petr Gapko**, Academy of Sciences of the Czech Republic, Czech Republic
Co-authors: Martin Smid
We propose a dynamic model of credit risk with two factors. The model enhances the current research by introducing dynamics, incorporating the external (macroeconomic) influence and enabling to set realistic parameters of a loan portfolio. We follow the common approach that the credit risk on a loan portfolio can be decomposed into a probability of default and a loss given default and assume that both are driven by two underlying factors: one common for all borrowers in the portfolio and one individual for each single borrower. We suggest on a large mortgage portfolio econometric estimation of the model and show realistically how the factors are related and how the credit risk is influenced by macroeconomic environment, which enables a comprehensive approach to stress testing.

CC1733: **On the credit risk on bank loans**
Presenter: **Arvid Raknerud**, Statistics Norway, Norway
Co-authors: Bjorn Helge Vatne, Ida Hjelseth
We assess the risk of default on bank loans to firms by estimating logit models for the probability of bankruptcy. The primary explanatory variables in the logit models are the firms' credit rating and selected accounting variables, such as total assets, return on total assets and measures of liquidity and leverage. We have access to firm-level panel data for the period 1996-2013. The logit models are used to quantify default risk by weighting each firm's bank debt at the beginning of the year (available from accounts data) with the estimated probability that the firm will close down within the next year (and subsequently be declared as bankrupt). By aggregating over all the firms in an industry we obtain a measure of credit risk for the whole industry, which we denote risk-weighted debt. This measure is useful in order to monitor the financial stability of the banking system and evaluate aggregate default risk. We also compare the total risk-weighted debt to aggregate default rates on business loans from the bank statistics, which is available at the industry level for the period 1999-2014. The comparison reveals a striking pattern with good agreement between aggregate default rates and our measures of total risk-weighted debt calculated from firm-level data.

CC1715: **Modelling loss severity for residential mortgage loans: A three-step selection approach**
Presenter: **Hung Do**, University of Technology Sydney, Australia
Co-authors: Harald Scheule, Daniel Rosch
We develop a new framework to model the loss given default (LGD) of residential mortgage loans, one of the most important asset classes in banks. According to Finance perspective, defaulted loans with zero-LGD are intuitively more relevant to the liquidity constrains of the borrowers (such as, divorce, demotion or loss of job); whereas, the non-zero LGD loans often relates to negative equity issues. Therefore, we suggest treating zero- and non-zero LGD loans distinctively. We propose a three-step selection approach with a joint probability framework among default, cure (i.e, zero-LGD) and loss severity information. Our model better fits the Finance perspective as well as the bimodal nature of the LGD distribution, which shows a massive concentration on zero. An application of our framework on the U.S residential mortgage loans during the Global Financial Crisis shows interesting new evidence that: (i) foreclosed residential mortgage loans associated with higher credit quality borrowers are less likely to be fully recovered; (ii) regional foreclosure contagion effect exists in an essence that a rise in local foreclosure rate increases the probability of default, lowers the probability of cure and worsens the non-zero LGD; (iii) house price under distress of the foreclosed property exacerbates the non-zero LGDs locally.

CC1404: **Long term expected corporate bond spreads**
Presenter: **Erik Hennink**, Ortec Finance, Netherlands
Long term investors are interested in future risk and return of asset classes to determine efficient portfolios. Using long historical data and a relatively simple valuation model, we estimate the long term expected value of US corporate bond spreads for different ratings and maturities. From literature, we know that credit spreads are not only a compensation for the credit risk, but are also a compensation for liquidity. This finding is one of the key elements in the model and this is incorporated in the model as a constant liquidity risk premium per rating. Using long term historical data of cumulative default probabilities, the valuation model converts the historical default probabilities to risk-neutral ones such that the full credit spread curve is derived. The shapes of the constructed long term mean credit curves are consistent with the theoretical shapes of the credit spread curves. Furthermore, we find that the model-implied spread curves are in line with the historical Nelson-Siegel credit spread curves.

CC029 Room Gordon CONTRIBUTIONS ON ASYMPTOTICS IN ECONOMETRICS Chair: Tucker McElroy

CC1630: **Martingale transforms with mixed stable limits and the QMLE for conditionally heteroskedastic models**
Presenter: **Alexandros Louka**, Athens University of Economics and Business, Greece
Co-authors: Stelios Arvanitis
We derive a limit theorem for appropriately centered and scaled martingale transforms $\sum_{i=1}^{n} \xi_i V_i$ to mixed-stable limits when (ξ_i) is an iid sequence in the domain of attraction of an $\alpha-$stable distribution where $\alpha \in (0,2]$. Using the Principle of Conditioning we recover and extend known results in the literature while imposing weaker conditions. The results are useful in determining the limit theory of the QMLE in conditionally heteroskedastic models when the squared innovations are heavy-tailed. We provide the framework for the QMLE limit theory which in the ergodic case, is based on the stochastic recurrence approach and we furthermore allow for the parameter vector to lie on the boundary. Then we show that the QMLE weakly converges to an $\alpha-$stable distribution when $\alpha \in [1,2]$ and is inconsistent when $\alpha < 1$. We relax the assumption on ergodicity and provide analogous results for the non-stationary GARCH(1,1) case. We investigate the limit theory of the usual Wald statistic and provide with the asymptotic exactness and consistency of the relevant testing procedure based on subsampling. In the context of the ergodic GARCH(1,1) model we construct a testing procedure for weak stationarity and evaluate the performance of the testing procedure by means of Monte Carlo simulations.

CC1720: **Asymptotic theory for the ARMA(1,1)-EGARCH(1,1) model quasi-maximum likelihood estimator**
Presenter: **Xiaoyu Li**, University of Exeter, United Kingdom
The aim is to investigate the asymptotic properties of the quasi-maximum likelihood estimator (QMLE) of the autoregressive moving average model with an exponential generalized autoregressive conditional heteroscedasticity error (ARMA(1,1)-EGARCH(1,1)). Literature on the asymptotic theory for the ARMA-GARCH and EGARCH processes is firstly reviewed, and then it is extended to the ARMA-EGARCH process. The consistency and asymptotic normality of the QMLE have been established under mild conditions.

CC1662: **Inference and testing on the boundary in extended constant conditional correlation GARCH models**
Presenter: **Rasmus Soendergaard Pedersen**, University of Copenhagen, Denmark
We consider inference and testing in extended constant conditional correlation GARCH models in the case where the true parameter vector is a boundary point of the parameter space. This is of particular importance when testing for volatility spillovers in the model. The large-sample properties of the QMLE are derived together with the limiting distributions of the related LR, Wald, and LM statistics. Due to the boundary problem, these large-sample properties become nonstandard. The size and power properties of the tests are investigated in a simulation study. As an empirical illustration we test for (no) volatility spillovers between foreign exchange rates.

CC1641: **CAN properties of a regional estimator in the presence of national micro moments: A simulation study**
Presenter: **Yuichiro Kanazawa**, University of Tsukuba, Japan
Co-authors: Keisuke Takeshita, Kazuhiro Nakayama
We have previously analyzed the econometric properties of an estimator where the presumption is that there is one national market for cars, and the asymptotics are related to the number of products increases. Our model extended a previous one with additional style moments relating consumer demographics to the characteristics of purchased products. In many applications for nondurable consumer goods such as beer or breakfast cereal, however, the asymptotics are not in the number of products in a market, but in the number of markets. We wish to use style moments at a national level in a situation where econometricians can observe many regional markets. Via Monte Carlo simulation, we investigate asymptotic properties of the previous estimator for such nondurable consumer goods in the presence of national micro moments. As the number of market increases, the estimator exhibit a CAN property and the micro moments improve accuracy of estimates, although its contribution to the accuracy decreases as the number of markets increases as expected.

CC024 Room G21A CONTRIBUTIONS IN FINANCIAL ECONOMETRICS Chair: Florian Ielpo

CC1219: **Realized-CARE for tail risk forecasting with range and realized measures**
Presenter: **Chao Wang**, The University of Sydney, Australia
Co-authors: Richard Gerlach
A new framework named Realized Conditional Autoregressive Expectile (Realized-CARE) is proposed, through incorporating a measurement equation into the conventional CARE model, in a framework analogous to Realized-GARCH. The Range and realized measures (Realized Variance and Realized Range) are employed as the dependent variables of the measurement equation, since they have proven more efficient than return for volatility estimation. The dependence between range and realized measures and expectile can be modelled with this measurement equation. The grid search accuracy of the expectile level will be potentially improved with introducing this measurement equation. In addition, through employing a quadratic fitting target search, the speed of grid search is significantly improved. Bayesian adaptive Markov Chain Monte Carlo is used for estimation, and demonstrates its superiority compared to maximum likelihood in a simulation study. Furthermore, we propose an innovative sub-sampled realized range and also adopt an existing scaling scheme, in order to deal with the micro-structure noise of the high frequency volatility measures. Compared to the CARE, the parametric GARCH and the Realized-GARCH models, Value-at-Risk and Expected Shortfall forecasting results of 6 indices and 3 assets series favor the proposed Realized-CARE model, especially the Realized-CARE model with Realized Range and sub-sampled Realized Range.

CC1455: **Evaluating credit worthiness in a non credit culture society**
Presenter: **Mary Akinyemi**, University of Lagos, Nigeria
Co-authors: Christson Adedoyin
In the current world system where credit drives majority of transactions, it becomes increasingly necessary to evaluate credit worthiness. In a non credit culture society, that is a system where individuals have no credit history, evaluating the credit worthiness of an individual becomes a daunting task. We consider the key factors that would determine the credit worthiness of an individual and attempt to evaluate credit worthiness of customers in a Nigerian bank using statistical and artificial intelligence techniques such as Random Forests, Logistic Regression, Classification trees, Support vector machines, Neural networks and Linear discriminant analysis. We found that for all the models seem to correctly classify the credit worthy customers. However, Random Forest, Logistic Regression and Support Vector Machine better classify individuals who are not credit worthy correctly. Furthermore, Random Forest, Logistic Regression and Support Vector Machine have the highest success rates. Also, Random Forest, Logistic Regression and Support Vector Machine realizes less cost of misclassification. In addition, the Mean Decrease Gini indicates that gender and education are major factors that determine the credit worthiness of a customer in a Nigerian bank.

CC1666: **Optimal long-term allocation with pension fund liabilities**
Presenter: **Michael Rockinger**, UNIL-HEC Lausanne, Switzerland
Co-authors: Eric Jondeau
We build a macroeconomic model for Switzerland, the Euro area, and the USA that drives the dynamics of several asset classes and the liabilities of a representative Swiss (defined-contribution) pension fund. This encompassing approach allows us to generate correlations between returns on

assets and liabilities. We calibrate the economy using quarterly data between 1985:Q1 and 2013:Q2. Using a certainty equivalent approach, we demonstrate that a liabilities-hedging portfolio outperforms an assets-only strategy by between 5% and 15% per year. The main reason for such a large improvement is that the optimal assets-only portfolio is typically long in cash, whereas hedging liabilities require the pension fund to be short in cash. It follows that imposing positivity restrictions in the construction of the portfolio also results in a large cost, between 4% and 8% per year. This estimate suggests that allowing pension funds to hedge their liabilities through borrowing cash and investing in a diversified bond portfolio helps to enhance the global portfolio return.

CC1590: **An anatomy of global risk premiums**
Presenter: **Florian Ielpo**, Centre Economie de la Sorbonne, France
Co-authors: Ling-Ni Boon
Long-term investors are attuned to the thought that risk is rewarded. By making an investment with more potential variation in returns, investors demand a risk premium - the return in excess of a comparatively risk-less investment. This is particularly espoused in the long-term investments of pension funds, yet is a reductive view of financial markets. We investigate the existence of risk premiums in a global, multi-asset portfolio of a typical pension fund, and relate the variation of the risk premiums to macroeconomic fluctuations. Due to the coincident relation between risk premiums and the economic cycle, under the prevailing economic condition, the Sharpe ratios of portfolios constructed to capitalize on risk premiums are appallingly low (between -1 and 0.2). Despite the heartening corroboration of risk premiums existence for long- term investors, investors remain susceptible to their time variation.

EO306 Room MAL 541 ESTIMATION OF LARGE RANDOM MATRICES: THEORIES AND APPLICATIONS Chair: Clifford Lam

EO0193: **Outliers in the spectrum of large sample covariance matrices**
Presenter: **Jeff Yao**, The University of Hong Kong, China
Random matrix theory essentially is an attempt to understand the behaviour of the eigenvalues of a large random matrix. Because their large number, say in a few hundreds, it is useless or impossible to examine these eigenvalues individually. Rather, interesting questions concern either their joint empirical distribution or a few particular among them such as the extreme eigenvalues (the largest or the smallest ones). While the study of the joint empirical distribution of large random matrices has been long standing and well developed, the advances on the extreme eigenvalues have been much more recent. We will concentrate on a specific class of spiked covariance matrices. These matrices have attracted much attention in the last fifteen years because of their 'multidisciplinary' nature: indeed, related questions can be deeply theoretical or completely applied to real-life data analysis. We thus study the effect produced on the extreme eigenvalues by a introduction of a finite number of spikes in the population covariance matrix. Recent advances on this topic will be reviewed and some new applications to signal detection discussed.

EO0966: **Local low rank matrix completion**
Presenter: **Franz Kiraly**, University College London, United Kingdom
Co-authors: Louis Theran, Ryota Tomioka, Duncan Blythe
In recent years, the low rank matrix completion model has enjoyed quite some success for recommendation and prediction learning. Many standard algorithms in the field are designed for completing the whole matrix, which also means that achieving scalability on huge data sets is a challenging task. A question which has only been more recently considered is prediction of single entries or only a few. The novel approach based on algebra and combinatorics allows, for the first time, a systematic treatment of single-entry estimates including single-entry error bounds, and it yields, for the first time, a closed approach to the low-rank model that is intrinsically local. We will give a brief introduction to the matrix completion problem, and its algebraic combinatorial formulation. We will demonstrate how this allows us to derive simple reconstruction algorithms, and we will review a recent application in sports performance prediction.

EO1152: **Optimal estimation of a large-dimensional covariance matrix under Stein's loss**
Presenter: **Olivier Ledoit**, University of Zurich, Switzerland
Co-authors: Michael Wolf
A new method is introduced for deriving covariance matrix estimators that are decision-theoretically optimal. The key is to employ large-dimensional asymptotics: the matrix dimension and the sample size go to infinity together, with their ratio converging to a finite, nonzero limit. As the main focus, we apply this method to Stein's loss. Compared to Stein's estimator, ours has five theoretical advantages: (1) it asymptotically minimizes the loss itself, instead of an estimator of the expected loss; (2) it does not necessitate post-processing via an ad hoc algorithm (called isotonization) to restore the positivity or the ordering of the covariance matrix eigenvalues; (3) it does not ignore any terms in the function to be minimized; (4) it does not require normality; and (5) it is not limited to applications where the sample size exceeds the dimension. In addition to these theoretical advantages, our estimator also improves upon Stein's estimator in terms of finite-sample performance, as evidenced via extensive Monte Carlo simulations. To further demonstrate the effectiveness of our method, we show that some previously suggested estimators of the covariance matrix and its inverse are decision-theoretically optimal with respect to the Frobenius loss function.

EO158 Room MAL 633 RECENT ADVANCES IN QUANTILE REGRESSION Chair: Carlos Lamarche

EO0305: **Measurement errors in quantile regression models**
Presenter: **Antonio Galvao**, University of Iowa, United States
Co-authors: Sergio Firpo, Suyong Song
Estimation and inference are developed for quantile regression models with measurement errors. We propose an easily-implementable semiparametric two-step estimator when we have repeated measures for the covariates. Building on recent theory of Z-estimation with infinite-dimensional parameters, consistency and asymptotic normality of the proposed estimator are established. We also develop statistical inference procedures and show the validity of a bootstrap approach to implement the methods in practice. Monte Carlo simulations assess the finite sample performance of these proposed methods. We apply our method to a well-known example of returns to education on earnings, using data set of female monozygotic twins in U.K. We find strong heterogeneity in returns to education along the conditional distribution of earnings. In addition, the returns are relatively larger at the lower part of the distribution, providing evidence that a potential economic redistributive policy should focus on such quantiles.

EO0741: **Cluster-robust bootstrap inference in quantile regression models**
Presenter: **Andreas Hagemann**, University of Michigan, United States
A wild bootstrap procedure is developed for cluster-robust inference in linear quantile regression models. The bootstrap is shown to lead to asymptotically valid inference on the entire quantile regression process in a setting with a large number of small, heterogeneous clusters and provides consistent estimates of the asymptotic covariance function of that process. The proposed bootstrap procedure is easy to implement and performs well even when the number of clusters is much smaller than the sample size. An application to Project STAR data is provided.

EO1034: **Sparsity-based estimation of a panel quantile count data model with applications to big data**
Presenter: **Matthew Harding**, Duke University, United States
Co-authors: Carlos Lamarche
We introduce a panel quantile estimator for count data with individual heterogeneity, by constructing continuous variables whose conditional quantiles have a one-to-one relationship with the conditional count response variable. The new method is needed as a result of the increased availability of Big Data, which allows us to track event counts at the individual level for a large number of activities from webclicks and retweets to store visits and purchases. At the same time, the presence of many different subpopulations in a large dataset requires us to pay close attention to individual heterogeneity. We propose a penalized quantile regression estimator with fixed effects and investigate the conditions under which the slope parameter estimator is asymptotically Gaussian. We investigate solutions to the computational challenges resulting from the need to estimate tens of thousands of parameters in a Big Data setting and caution against penalizing in models with substantial zero inflation and endogenous covariates by using a series of Monte Carlo simulations. We present an empirical application to individual trip counts to the store based on a large panel of food purchase transactions.

EO236 Room MAL 421 NETWORK STATISTICS Chair: Richard Samworth

EO0385: **Network structural change points detection**
Presenter: **Yi Yu**, University of Cambridge, United Kingdom
Co-authors: Ivor Cribben
Motivated by the increased interest in quantifying changes in connectivity between brain regions over an experimental time course to provide a deeper insight into the fundamental properties of brain networks, a new statistical method to detect network structural change points is proposed. The novel method uses spectral clustering to study the network structure and uses a nonparametric metric to detect the changes in the structures across time course. The new method allows for situations where the number of nodes is much greater than the number of time points in the experimental time course. This method promises to offer deeper insight into the inner workings of the whole brain. We apply this new method to simulated data and to a resting-state fMRI data set. The temporal features of this novel connectivity method will provide a more accurate understanding of the large-scale characterisations of brain disorders such as Alzheimer's disease and may lead to better diagnostic and prognostic indicators.

EO1808: **Network analysis and nonparametric statistics**
Presenter: **Pierre-Andre Maugis**, University College London, United Kingdom
Co-authors: Sofia Olhede, Patrick Wolfe
Networks are ubiquitous in today's world. Any time we make observations about people, places, or things and the interactions between them, we have a network. Yet a quantitative understanding of real-world networks is in its infancy, and must be based on strong theoretical and methodological foundations. The goal is to provide some insight into these foundations from the perspective of nonparametric statistics, in particular how trade-offs between model complexity and parsimony can be balanced to yield practical algorithms with provable properties.

EO1830: **Big network data**
Presenter: **Patrick Wolfe**, University College London, United Kingdom
How do we draw sound and defensible data-analytic conclusions from networks? This question has recently risen to the forefront of mathematical statistics, and it represents a fundamental challenge for data science. I will describe new large-sample theory that helps us to view and interpret networks as statistical data objects, along with the transformation of this theory into new statistical methods to model and draw inferences from network data in the real world. The insights that result from connecting theory to practice also feed back into pure mathematics and theoretical computer science, prompting new questions at the interface of combinatorics, analysis, probability, and algorithms.

EO104 Room MAL B20 COMPLEX DATA ANALYSIS: THEORY AND METHODS Chair: Xinghao Qiao

EO0562: **Robust learning for optimal treatment decision with NP-dimensionality**
Presenter: **Rui Song**, North Carolina State University, United States
In order to identify important variables that are involved in making optimal treatment decision, a penalized least squared regression framework for a fixed number of predictors has been previously proposed which is robust against the misspecification of the conditional mean model. Two problems arise: (i) in a world of explosively big data, effective methods are needed to handle ultra-high dimensional data set, for example, with the dimension of predictors is of the non-polynomial (NP) order of the sample size; (ii) both the propensity score and conditional mean models need to be estimated from data under the NP dimensionality. We propose a two-step estimation procedure for deriving the optimal treatment regime under the NP dimensionality. In both steps, penalized regressions are employed with the non-concave penalty function, where the conditional mean model of the response given predictors may be misspecified. The asymptotic properties, such as weak oracle properties, selection consistency and oracle distributions, of the proposed estimators are investigated. In addition, we study the limiting distribution of the estimated value function for the obtained optimal treatment regime. The empirical performance of the proposed estimation method is evaluated by simulations.

EO1309: **Randomized maximum-contrast selection: Subagging for large-scale regression**
Presenter: **Jelena Bradic**, University of California San Diego, United States
We introduce a very general method for sparse and large-scale variable selection. The large-scale regression settings is such that both the number of parameters and the number of samples are extremely large. The proposed method is based on careful combination of penalized estimators, each applied to a random projection of the sample space into a low-dimensional space. In one special case that we study in detail, the random projections are divided into non-overlapping blocks; each consisting of only a small portion of the original data. Within each block we select the projection yielding the smallest out-of-sample error. Our random ensemble estimator then aggregates the results according to new maximal-contrast voting scheme to determine the final selected set. Our theoretical results illuminate the effect on performance of increasing the number of non-overlapping blocks. Moreover, we demonstrate that statistical optimality is retained along with the computational speedup. The proposed method achieves minimax rates for approximate recovery over all estimators using the full set of samples. Furthermore, our theoretical results allow the number of sub-samples to grow with the subsample size and do not require irrepresentable condition. The estimator is also compared empirically with several other popular high-dimensional estimators via an extensive simulation study, which reveals its excellent finite-sample performance.

EO1317: **High dimensional and dynamic covariance matrix estimation with application to portfolio allocation**
Presenter: **Shaojun Guo**, Chinese Academy of Sciences, China
Estimation of high dimensional covariance matrices is an interesting and important research topic. We propose a dynamic structure and develop an estimation procedure for high dimensional covariance matrices. Asymptotic properties are derived to justify the estimation procedure and simulation studies are conducted to demonstrate its performance when the sample size is finite. By exploring a financial application, an empirical study shows that portfolio allocation based on dynamic high dimensional covariance matrices can significantly outperform the market from 1995 to 2014. Our proposed method also outperforms portfolio allocation based on the sample covariance matrix and the portfolio allocation previously proposed.

EO661 Room MAL G15 COMBINATORIAL OPTIMIZATION FOR GRAPHICAL MODEL SELECTION II Chair: James Cussens

EO0754: **Learning Bayesian networks on high-dimensional genomic data**
Presenter: **Cassio Polpo de Campos**, Queens University Belfast, United Kingdom
Ideas and methods are presented for learning Bayesian networks on data sets with many thousands of variables and few data samples. The goal is to find a good balance between accuracy and efficiency of learning by employing theoretical results that can speed up computations as much as possible while losing as little as possible in accuracy. Applications arising in biomedical problems are described, where it is argued that Bayesian networks can provide meaningful and interpretable results. In particular, we discuss on the use of Bayesian networks for clustering and classification using high-dimensional genomic data sets of lymphoma patients.

EO0609: **Finding optimal Bayesian network structures with constraints learned from data**
Presenter: **Brandon Malone**, Max Planck Institute for Biology of Ageing, Germany
Co-authors: Changhe Yuan, Xiannian Fan
The Bayesian network structure learning problem (BNSL) is NP-hard. Nevertheless, several algorithms have been proposed which, in practice, can probably solve many instances quite effectively. All of these algorithms first calculate potentially optimal parent sets (POPS) for all variables, and they then use various optimization techniques to find a set of POPS, one for each variable, that constitutes an optimal network structure. We first briefly describe BNSL and give details on an admissible state space search algorithm for finding provable optimal networks. We then show how the POPS given as input to the problem constrain the parent candidates for each variable. Taken together, the parent candidates of all variables give rise to a directed cyclic graph which decomposes into a set of strongly connected components (SCCs). Each SCC corresponds to an independent subproblem. This decomposition is applicable to all current optimal BNSL algorithms. Empirical results show that solving the decomposed problem significantly improves the efficiency and scalability of admissible search-based BNSL algorithms. Further, we show that by considering only the top p POPS of each variables, we quickly find very high quality networks for large datasets.

EO1037: **Constraint reasoning and optimization for learning optimal graphical models**
Presenter: **Matti Jarvisalo**, University of Helsinki, Finland
Integration of the fields of constraint solving and machine learning has recently been identified as an important research direction with high potential. Recent work is overviewed on harnessing Boolean satisfiability and optimization procedures to learning optimal structures of graphical models such as Bayesian networks, causal graphs, and chain graphs.

EO316 Room MAL 632 SHAPE-CONSTRAINED INFERENCE AND OTHER NON-REGULAR PROBLEMS Chair: George Michailidis

EO0962: **Shape-constrained inference for the argmax of a concave function, with applications**
Presenter: **Charles Doss**, University of Minnesota, United States
Co-authors: Jon A Wellner
We consider performing inference about the location of the maximum of a function on $\mathbb{R}$ satisfying the shape restriction of concavity. In non-parametric settings, the argmax is generally not estimable at a $\sqrt{n}$-rate and does not always have a normal limiting distribution. We develop a shape-constrained likelihood ratio test (which can be inverted to form confidence intervals), compare it to other methods for performing inference, and discuss some applications of the test.

EO1277: **Inference for monotone functions under short and long range dependence**
Presenter: **Pramita Bagchi**, Ruhr University Bochum, Germany
Co-authors: Moulinath Banerjee, Stilian Stoev
We introduce new point-wise confidence interval estimates for monotone functions observed with additive, dependent noise. Our methodology applies to both short- and long-range dependence regimes for the errors. The interval estimates are obtained via the method of inversion of certain discrepancy statistics. This approach avoids the estimation of nuisance parameters such as the derivative of the unknown function, which previous methods are forced to deal with. The resulting estimates are therefore more accurate, stable, and widely applicable in practice under minimal assumptions on the trend and error structure. The dependence of the errors especially long-range dependence leads to new phenomena, where new universal limits based on convex minorant functionals of drifted fractional Brownian motion emerge. Some extensions to uniform confidence bands are also developed.

EO1593: **On smooth isotonic regression**
Presenter: **Runlong Tang**, Virginia Tech, United States
Some results on smooth isotonic regression will be discussed. We propose a method combining kernel smoothing methods and isotonic regression.

EO1846: **Some results for the single index model with applications**
Presenter: **Hanna Jankowski**, York University, Canada
Co-authors: Cecile Durot, Fadoua Balabdaoui
We consider the single index models assuming a monotone ridge function. This model is a special case of a general class of models recently considered. We discuss the interesting geometry arising from maximum likelihood estimation in this model, as well as asymptotics in the fixed design setting. The latter is motivated by a data set studying the lethal dose in a decompression sickness study.

EO190 Room MAL 540 ADVANCES IN MIXTURE MODELLING Chair: Geoffrey McLachlan

EO1142: **Model based one-class classification based on transvariation probability**
Presenter: **Angela Montanari**, Alma mater studiorum-Universita di Bologna, Italy
Co-authors: Laura Anderlucci, Francesca Fortunato
One-class classification is required when objects from one class only, the target class, are available. This is different from the traditional classification problem, which tries to distinguish between two or more classes with the training set containing objects from all of them. A method for one-class classification based on Gini's transvariation probability between a group and a constant is proposed. Transvariation probability is a measure of how typical an observation is with respect to a given group. It is assumed that the target class probability density function can be modelled by a mixture model in a dimensionally reduced space identified either by variable selection or by principal components. A formulation of multivariate transvariation probability is proposed which allows for mixture densities and its performances in one-class classification are assessed both on real and simulated data.

EO1330: **Spatial clustering of time-series via mixtures of autoregressive models and Markov random fields for image analysis**
Presenter: **Hien Nguyen**, University of Queensland, Australia
Co-authors: Geoffrey McLachlan, Jeremy Ullmann, Andrew Janke
Time-series data arise in many medical and biological imaging scenarios. In such images, a time-series is obtained at each of a large number of spatially dependent data units (e.g. electrodes, pixels, or voxels). It is often interesting to organize these data units into clusters that are modeled by an underlying probabilistic process. A two-stage procedure is presented for this task. In Stage 1, a mixture of autoregressive (MoAR) model is used to marginally cluster the time series arising at each data unit. The MoAR model is fitted using maximum marginal likelihood (MML) estimation via an MM (minorizationmaximization) algorithm. In Stage 2, a Markov random field (MRF) model is used to induce a spatial dependency structure onto the Stage 1 clustering. The MRF model is also fitted using maximum pseudolikelihood (MPL) estimation via an MM algorithm. A simulation study is presented to demonstrate the performance of the two-stage procedure. An application to the segmentation of a zebrafish brain calcium image is presented as a demonstration of the methodology.

EC1246: **Robust estimation for mixtures of skew data**
Presenter: **Francesca Greselin**, University of Milano Bicocca, Italy
Co-authors: Luis Angel Garcia-Escudero, Agustin Mayo-Iscar, Geoffrey McLachlan
Recently, observed departures from the classical Gaussian mixture model in real datasets have led to the introduction of more flexible tools for modeling heterogeneous skew data. Among the latest proposals in the literature, we consider mixtures of skew normal, to incorporate asymmetry in components, as well as mixtures of t, to down-weight the contribution of extremal observations. Clearly, mixtures of skew t have widened the application of model based clustering and classification to great many real datasets, as they can adapt to both asymmetry and leptokurtosis in the grouped data. Unfortunately, when data contamination occurs far from the bulk of the data, or even between the groups, classical inference for these models is not reliable. Our proposal is to address robust estimation of mixtures of skew normal, to resist sparse outliers and even pointwise contamination that could arise in data collection. We introduce a constructive way to obtain a robust estimator for the mixture of skew normal model, by incorporating impartial trimming and constraints in the EM algorithm. At each E-step, a low percentage of less plausible observations, under the estimated model, is tentatively trimmed; at the M-step, constraints on the scatter matrices are employed to avoid singularities and reduce spurious maximizers. Some applications on artificial and real data show the effectiveness of our proposal, and the joint role of trimming and constraints to achieve robustness.

EO304 Room MAL 532 HIGH-DIMENSIONAL DATA ANALYSIS BEYOND LINEAR MODELS Chair: Junhui Wang

EO1303: **Probability-enhanced sufficient dimension reduction for binary classification**
Presenter: **Hao Zhang**, University of Arizona, United States
Many sufficient dimension reduction (SDR) methods have been developed since the introduction of sliced inverse regression. For binary classification problems, SIR suffers the limitation of estimating at most one direction since only two slices are available. We propose a new and flexible probability-enhanced SDR method for binary classification problems using the weighted support vector machine (WSVM). The key idea is to slice the data based on conditional class probabilities of observations rather than their binary responses. We show that the central subspace based on the conditional class probability is the same as that based on the raw binary response, which justifies the proposed slicing scheme and assures no information loss. Furthermore, in order to implement the new slicing scheme, one does not need exact probability values since the only required information is the relative ordering of probability values. The new SDR bypasses the probability estimation and employs the WSVM to directly estimate the order of probability values, based on which the slicing is performed. The performance of the proposed probability-enhanced SDR scheme is evaluated by both simulated and real data examples.

EO0634: **Sparse quadratic discriminant analysis**
Presenter: **Chenlei Leng**, Warwick, United Kingdom
A novel QUadratic Discriminant Analysis procedure called QUDA is proposed for analysing high-dimensional data. The proposed method is able to identify quadratic interactions of the variables for classification. Under appropriate sparsity assumptions, we show that QUDA works even when the dimensionality is exponentially high with respect to the sample size. We develop an efficient algorithm based on the alternating direction method of multipliers method (ADMM) for finding interactions along the way, which is much faster than its competitor in the literature. The competitive performance of QUDA is illustrated via extensive simulation studies and the analysis of real datasets.

EO0217: **Sparse partially linear additive models**
Presenter: **Jacob Bien**, Cornell University, United States
Co-authors: Yin Lou, Rich Caruana, Johannes Gehrke
The generalized partially linear additive model (GPLAM) is a flexible and interpretable approach to building predictive models. It combines features in an additive manner, allowing each to have either a linear or nonlinear effect on the response. However, the choice of which features to treat as linear or nonlinear is typically assumed known. Thus, to make a GPLAM a viable approach in situations in which little is known a priori about the features, one must overcome two primary model selection challenges: deciding which features to include in the model and determining which of these features to treat nonlinearly. We introduce the sparse partially linear additive model (SPLAM), which combines model fitting and both of these model selection challenges into a single convex optimization problem. SPLAM provides a bridge between the lasso and sparse additive models. Through a statistical oracle inequality and thorough simulation, we demonstrate that SPLAM can outperform other methods across a broad spectrum of statistical regimes, including the high-dimensional setting. We develop efficient algorithms that are applied to real data sets with half a million samples and over 45,000 features with excellent predictive performance.

EG067 Room MAL 402 CONTRIBUTIONS ON COMPLEX DATA Chair: Thaddeus Tarpey

EC1635: **Rate of uniform consistency for a class of mode regression on functional stationary ergodic data**
Presenter: **Mohamed Chaouch**, United Arab Emirates University, United Arab Emirates
Co-authors: Naamane Laib, Djamal Louani
The aim is to study the asymptotic properties of a class of kernel conditional mode estimates whenever functional stationary ergodic data are considered. To be more precise, in the ergodic data setting, we consider a random element $(X;Z)$ taking values in some semi-metric abstract space $E \times F$. For a real function φ defined on the space F and $x \in E$, we consider the conditional mode of the real random variable $\varphi(Z)$ given the event $X = x$. While estimating the conditional mode function, say $\theta_\varphi(x)$, using the well-known kernel estimator, we establish the strong consistency with rate of this estimate uniformly over Vapnik-Chervonenkis classes of functions φ. Notice that the ergodic setting offers a more general framework than the usual mixing structure. Two applications to energy data are provided to illustrate some examples of the proposed approach in time series forecasting framework. The first one consists in forecasting the daily peak of electricity demand in France (measured in Giga-Watt). Whereas the second one deals with the short-term forecasting of the electrical energy (measured in Giga-Watt per Hour) that may be consumed over some time intervals that cover the peak demand.

EC1444: **Nearest neighbor ensembles for functional data**
Presenter: **Jan Gertheiss**, Georg August University of Goettingen, Germany
Co-authors: Karen Fuchs
An ensemble method for nonparametric classification with functional data is introduced that inherently provides automatic and interpretable feature selection. It is designed for single as well as multiple functional (and non-functional) covariates. The ensemble members are posterior probability estimates that are obtained using $k-$nearest-neighbors based on different semi-metrics, with each of those semi-metrics focusing on a specific curve feature. Each ensemble member, and thus each curve feature, is weighted by a specific coefficient which is estimated using a proper scoring rule with implicit lasso-type penalty, such that some coefficients can be estimated to be exactly zero. Thus, the ensemble automatically provides feature (and variable) selection. The method is illustrated in simulation studies and on real data from water quality monitoring. Besides classification, the presented approach can also be used for regression with functional covariates.

EC0330: **Bias-correction of Kalman filter predictors associated with a linear state space model with estimated parameters**
Presenter: **Magda Monteiro**, University of Aveiro, Portugal
Co-authors: Marco Costa
The aim is to discuss some practical problems on linear state space models with estimated parameters. While the existing research focuses on the prediction mean square error, some results on bias propagation in forecast and filter predictions are presented, namely, non-recursive analytical expressions. In particular, it is discussed the impact of the bias in the invariant state space models. The theoretical results provide an adaptive correction procedure based on any parameters estimation method (for instance, maximum likelihood or distribution-free estimators). This procedure is applied to two datasets: the calibration of radar precipitation estimates and the global mean land-ocean temperature index modelling.

EC037 Room MAL 539 CONTRIBUTIONS ON LINEAR MODELS Chair: Sylvain Sardy

EC1514: **Deviance analysis of age-period-cohort models**
Presenter: **Bent Nielsen**, University of Oxford, United Kingdom
The age-period-cohort model is often used to model cancer mortality data aggregated in Lexis diagrams. The model has an inbuilt identification problem that has caused much controversy. Four contributions are made. First, the identification problem is addressed by reparametrising the model in terms of freely varying parameters: a linear plane and double differences of the original time effects. Second, a deviance analysis is presented that unifies a wide range of sub-models from the literature. Third, a new plot of the deviation of the time effects from the unidentified linear trend is suggested and interpreted. Fourth, parametrisations of these deviations are suggested. The methods are applied to a classic Belgian lung cancer data.

EC1716: **Asymptotic theory for over-dispersed chain-ladder models**
Presenter: **Jonas Harnau**, University of Oxford, United Kingdom
Co-authors: Bent Nielsen
The chain-ladder technique is ubiquitous in non-life insurance claim reserving. In a Poisson model, the chain-ladder technique is maximum likelihood estimation. The equivalence of mean and variance of the Poisson is usually refuted by the data. Often, an over-dispersed Poisson structure in which mean and variance are proportional is then assumed. Then, the chain-ladder technique is maximum quasi-likelihood estimation. An asymptotic theory is provided for this situation. This leads to closed form distribution forecasts involving the t distribution. Further, an asymptotically F distributed test statistic is proposed to test for adequacy of the chain-ladder technique compared to a more general model with calendar effect. A simulation study suggests that both distribution forecasts and test statistic give reasonable approximations in finite samples. The proposed distribution forecasts are compared with the standard bootstrap approach. The results generalise to age-period-cohort models used in other fields.

EC1439: **Finite mixture regression models with concomitant variables: A simulation study**
Presenter: **Kristyna Vankatova**, Palacky University, Czech Republic
Co-authors: Eva Fiserova
Finite mixture models are a popular technique for modelling unobserved heterogeneity in the population. Within the family of mixture models, mixtures of regression models have been proposed and studied as a replacement for more traditional cluster analysis and cluster-wise regression techniques. Using mixtures of regression models, the sample representing a heterogeneous population can be clustered into groups by modelling the conditional distribution of the response given the explanatory variable as a mixture. In many applications, the parameters of a mixture of linear regression models are estimated with the expectation maximization (EM) algorithm within a maximum likelihood framework. In order to characterize the different components and improve regression parameter estimates and predictions the use of concomitant variables has been proposed. The goal is to compare the performance of standard mixture of regressions and mixture of regressions with concomitant variables. A simulation study is designed to investigate this problem and is focused on the accuracy of parameter estimations and following predictions. R package flexmix is used for modelling of mixtures using the EM algorithm.

Monday 14.12.2015 16:20 - 18:00 Parallel Session P – CFE-CMStatistics

CO460 Room MAL G15 SEASONAL ADJUSTMENT AND RECONCILIATION Chair: Gian Luigi Mazzi

CO0737: **The direct versus indirect problem in seasonal adjustment**
Presenter: **Dominique Ladiray**, INSEE, France
Co-authors: Gian Luigi Mazzi
Most of principal economic indicators are aggregates: a National Industrial Production Index (IPI) is for example computed from the NACE sub-sector indexes; the European IPI is constructed as a weighted sum of the Member states IPIs etc. In this case, the problem is to decide how the seasonally adjusted series can be computed: Should we adjust the aggregate (direct approach) or should we aggregate the adjusted series (indirect approach)? The choice between the 2 approaches is still an open question and few scientific papers exist on the subject. In practice, the choice is based on publication constraints, rule of thumb or "basic common sense" and scarcely on statistical considerations. The problem has in fact many faces and variants that make it quite complex: additive and non-additive aggregation, mixed strategies, mixed adjustment methods etc. Some descriptive statistics have already been proposed to choose between the two strategies and we use multivariate techniques, like principal component analysis, factor analysis and cluster analysis, to help the user making a sound choice.

CO1048: **Seasonal adjustment with and without revisions: A comparison of X-13ARIMA-SEATS and CAMPLET**
Presenter: **Jan PAM Jacobs**, University of Groningen, Netherlands
Co-authors: Barend Abeln
Seasonality in macroeconomic time series can obscure movements of other components in a series that are operationally more important for economic and econometric analyses. Indeed, in practice one often prefers to work with seasonally adjusted data to assess the current state of the economy and its future course. Recently, two most widely used seasonal adjustment methods, Census X-12-ARIMA and TRAMO-SEATS, merged into X-13ARIMA-SEATS to become a new industry standard. We compare and contrast X-13ARIMA-SEATS with a seasonal adjustment program called CAMPLET, an acronym of its tuning parameters. CAMPLET consists of a simple adaptive procedure which separates the seasonal component and the non-seasonal component from an observed time series. Once this process has been carried out there will be no need to revise these components at a later stage when more observations become available, in contrast with other seasonal adjustment methods. We briefly review of X-13ARIMA-SEATS and describe the main features of CAMPLET. We evaluate the outcomes of both methods in a controlled simulation framework using a variety of processes. Finally, we apply the X-13ARIMA-SEATS and CAMPLET methods to three time series: U.S. non-farm payroll employment, operational income of Ahold, and real GDP in the Netherlands.

CO0970: **Seasonal adjustment of meager time series**
Presenter: **Tucker McElroy**, Census Bureau, United States
A meager time series is one whose sampling frequency is so high relative to actual activity, such that some values are recorded as extremely small positive values, or even as zeroes. Some economic series in small economies are meager, such as agricultural variables in countries with a distinct growing season (as opposed to data that represents aggregate activity over diverse climatological zones). Multiplicative seasonal adjustment typically utilizes a logarithmic transformation, but the small values cause huge distortions when so transformed. We apply a method of extreme value adjustment based on replacing meager values by optimal projections that utilize information from the available time series dynamics. The method is illustrated on New Zealand agricultural series.

CO1019: **Model-based seasonal adjustment with JDemetra+**
Presenter: **Jean Palate**, National Bank of Belgium, Belgium
JDemetra+ is a new open source tool for time series analysis, mainly developed by the National Bank of Belgium. The software is completely written in Java. Its primary goal is to provide the European Statistical System (ESS) with modules that enables the implementation of the ESS guidelines on seasonal adjustment (SA). It is partly built around the concepts and algorithms used in the two leading SA methods, i.e. Tramo-Seats and X12-Arima/X13-Arima-Seats. It also contains a rich state space framework that enables the exploration of alternative seasonal adjustment methods. Using JDemetra+, different model-based approaches, including several forms of structural models and the canonical decomposition of seasonal ARIMA models will be compared. Some theoretical properties of the models will be highlighted and results computed on large sets of real data will be presented.

CO1022: **Alternative two step reconciliation practices**
Presenter: **Enrico Infante**, EUROSTAT, Luxembourg
Co-authors: Germana Scepi
Whenever in a system of time series both the contemporaneous and the temporal constraints are not satisfied, the procedure of adjusting the discrepancies is called reconciliation. Some authors developed two step reconciliation methods, where in the first step the temporal constraint is solved, while in the second step the contemporaneous constraint is satisfied without altering the temporal constraint. The Di Fonzo-Marini method uses the Denton benchmarking technique in the first step, while in the second step the procedure preserves the reliability of the preliminary data. The aim is to present different methods by using different well-known regression based benchmarking techniques in the first step and keeping the second step preserving the reliability. Comparisons of the results obtained by using different techniques in the first step will be presented. Comparisons with simultaneous approaches are also made. A real example using European quarterly sector accounts data is also presented.

CG053 Room MAL 532 CONTRIBUTIONS ON BUSINESS CYCLE Chair: Antonello DAgostino

CC0419: **Revisiting the transitional dynamics of business-cycle phases with mixed frequency data**
Presenter: **Marie Bessec**, University Paris Dauphine, France
A Markov-Switching model is introduced where transition probabilities depend on higher frequency indicators and their lags, through polynomial weighting schemes. The MSV-MIDAS model is estimated via maximum likelihood methods. The estimation relies on a slightly modified version of Hamilton's recursive filter. We use Monte Carlo simulations to assess the robustness of the estimation procedure and related test-statistics. The results show that ML provides accurate estimates, but they suggest some caution in the tests on the parameters involved in the transition probabilities. We apply this new model to the detection and forecast of business cycle turning points. We properly detect recessions in the United States and the United Kingdom by exploiting the link between GDP growth and higher frequency variables from financial and energy markets. Spread term is a particularly useful indicator to predict recessions in the United States, while stock returns have the strongest explanatory power around British turning points.

CC1624: **News, slope of the term structure and uncertainty shocks: An empirical evaluation of its interconnections**
Presenter: **Danilo Cesar Cascaldi Garcia**, University of Warwick, United Kingdom
There are two main theories in the literature on business cycles driven by agents beliefs: the news shock - the shock on the future total factor productivity (TFP) that is foreseen by the economic agents -, and the uncertainty shock - the shock on general uncertainty that has the ability to

impact real economic outcome. We provide novel empirical evidence that these two shocks are linked adopting a BVAR model in levels and a maximum forecast error variance identification. We also revisit the well-known correlation between news shocks and shocks to the slope of the term structure, which vanishes when employing a new utilization-adjusted TFP series. Finally, we propose a novel identification scheme for news shocks, orthogonal to uncertainty shocks, that confirms the positive comovement among macroeconomic variables at impact. The findings are that the positive effects of news shocks are boosted in an environment of lower uncertainty.

CC1556: **Real-time nowcasting the US Output gap: Singular spectrum analyis at work**
Presenter: **Miguel de Carvalho**, Pontificia Universidad Catolica de Chile, Chile
Co-authors: Antonio Rua
We explore a new approach for nowcasting the output gap based on singular spectrum analysis. Resorting to real-time vintages, a recursive exercise is conducted so to assess the real-time reliability of our approach for nowcasting the US output gap, in comparison with some well-known benchmark models. For our applied setting of interest, the preferred version of our approach consists of a multivariate singular spectrum analysis, where we use a Fisher g test to infer which components, within the standard business cycle range, should be included in the grouping step. We find that singular spectrum analysis provides a reliable assessment of the cyclical position of the economy in real-time, with the multivariate approach outperforming substantially the univariate counterpart.

CC1548: **Constructing fluctuation data from emerging national account statistics**
Presenter: **Jerome Trinh**, THEMA - CREST, France
Assessing the economic fluctuations of emerging markets is an arduous exercise due to the scarcity of reliable infra-annual data, especially the time series of national accounts. Techniques such as interpolation, distribution and extrapolation of annual data by using related quarterly or monthly series can be adopted to construct infra-annual series of national accounts. However for these countries, the period of coincident account and indicator series is short and the regression model linking the two likely to be subject to structural changes. A residual-based ADF-type test designed to test the null of the presence of a unit root in the residuals of the model allowing for a change in the level, the trend and/or the slope coefficients is in definition similar to a test of cointegration with regime shift, for which the asymptotic properties have been exposed in the literature. Finite sample critical values for each different type of structural change model are computed considering both endogenous and exogenous break dates. The procedure is performed on Chinese annual and quarterly data, thus allowing an analysis of the relative amplitudes and the correlations between the fluctuations of the components of the national accounts.

CC1545: **Regional business cycles accross Europe**
Presenter: **Eduardo Bandres**, Universidad de Zaragoza, Spain
Co-authors: Ana Gomez-Loscos, Lola Gadea
Much effort has been devoted in the existing literature to studying the business cycle in Europe. Numerous studies have analysed business cycles and the synchronisation between countries that make up the Monetary Union, while little attention has been given to the regional framework. Against this background, our aim is to draw a map of the regional business cycles in the Euro area countries to help to identify the role of the economic geography and macro factors. As previously shown for the Spanish case, the national cycle may hide different regional rhythms of economic activity, which may have implications for the implementation of economic policies. To perform the analysis, we use different dating techniques, namely non-parametric, parametric and finite mixtures. In a preliminary study with Eurostat data we find evidence in favour of a unique cluster amongst the European countries, and 4 different groups of European regions. In a further analysis, this framework has been implemented using a more extensive database, which comprises a larger number of regions and a longer sample size. Using this dataset, we enrich our findings by calculating an index of regional synchronisation by country and relating the business cycle path to the structural characteristics of each economy. The results have important implications for policy-makers, both at the European level, in terms of convergence policies, and at the country level, in terms of domestic policies.

CG357 Room MAL 632 CONTRIBUTIONS ON EVALUATION OF FORECASTING Chair: Adam Clements

CC0747: **Comparing predictive accuracy under long memory**
Presenter: **Michael Will**, Leibniz University Hannover, Germany
Co-authors: Christian Leschinski, Robinson Kruse
The popular Diebold-Mariano test is extended to situations where the forecast error loss differential series has long memory. It is shown that this circumstance can arise frequently, because long memory is transmitted from the forecasts and the forecast objective to the forecast error loss differentials. The nature of this transmission depends on the (un)biasedness of forecasts. The case of fractional cointegration is hereby analyzed as well. Further results show that the conventional Diebold-Mariano test is invalidated under these circumstances and two robust statistics based on the memory and autocorrelation consistent (MAC) estimator and an extended fixed-bandwidth (EFB) approach are considered. The subsequent Monte Carlo study allows general conclusions about the relative performance of these methods to estimate the long-run variance. Furthermore we use the adaptive local polynomial whittle (ALPW) estimator to mitigate the effect of short-run dynamics. As an empirical application, we consider SPF forecasts for the three-month US Treasury Bills.

CC1410: **Comparing predictive accuracy in small samples**
Presenter: **Laura Coroneo**, University of York, United Kingdom
Co-authors: Fabrizio Iacone
We consider fixed-b and fixed-m asymptotics for a test of predictive accuracy. We show that this approach allows to obtain predictive accuracy tests that are correctly sized even in small samples. We apply the alternative asympotics for the test to evaluate the predictive accuracy of the Survey of Professional Forecasters (SPF) against a simple random walk. Our results show that the predictive ability of the SPF was partially spurious, especially in the last decade.

CC1468: **A fast model confidence set implementation for large and growing collections of models**
Presenter: **Sylvain Barde**, University of Kent, United Kingdom
A new algorithm is proposed for finding the confidence set of a collection of forecasts or prediction models. Existing numerical implementations for finding the confidence set use an elimination approach where one starts with the full collection of models and successively eliminates the worst performing until the null of equal predictive ability is no longer rejected at a given confidence level. The intuition behind the proposed implementation lies in reversing the process: one starts with a collection of two models and as models are successively added to the collection both the model rankings and $p-$values are updated. The first benefit of this updating approach is a reduction of one polynomial order in both the time complexity and memory cost of finding the confidence set of a collection of M models, falling respectively from $O\left(M^3\right)$ to $O\left(M^2\right)$ and from $O\left(M^2\right)$ to $O(M)$. This theoretical prediction is confirmed by a Monte Carlo benchmarking analysis of the algorithms. The second key benefit of the updating approach is that it intuitively allows for further models to be added at a later point in time, thus enabling collaborative efforts using the model confidence set procedure.

CC1572: **Micro information dynamics: Decomposing the forecasting power of aggregate indicators**
Presenter: **Klaus Wohlrabe**, ifo Institute for Economic Research, Germany

Many popular business indicators, such as the Ifo Business Climate, are based on surveys. Surveys collect information from often very heterogeneous respondents. In aggregate indicators, however, this rich panel information remains mostly unused. Panel information is used, in particular, differences in assessment and predictive ability between groups, in two ways: First, to examine which subset of respondents contributes most to the forecasting performance of an aggregate indicator and whether focusing on certain subsets can improve it. Second, to explore how panel information can improve the forecasting performance of an aggregate indicator based on the full sample. We classify respondents into subgroups based on size, sector, and characteristics of their past survey responses. Furthermore, we introduce dispersion and disagreement measures based on micro survey data. Our focus is forecasting industrial production in Germany with the large micro data set underlying the Ifo Business Survey. We demonstrate that well-defined subgroups of firms convey a large share of the information embedded in the aggregate indicator. Most importantly, exploiting the heterogeneity in responses between subgroups improves the forecasting accuracy.

CC1744: **On probability forecasts from betting odds: The CaSco normalization**
Presenter: **Vincenzo Candila**, University of Salerno, Italy
Co-authors: Antonio Scognamillo

Nowadays, the betting odds of professional bookmakers are widespread considered the most accurate source of sport forecasts. However, (the inverse of) the betting odds do not represent the true probability forecasts of the bookmaker. This is because they incorporate the bookmakers margin and the long-shot bias. In literature there have been proposed at least three normalization methods that normalize the inverse of the odds to obtain the probability forecasts for each outcome of a match: the basic, the regression and the Shin normalization. A new method is proposed, the CaSco normalization, for the fixed odds offered on betting markets. This procedure is free from any assumptions, differently from the Shin's method, and simply to apply as the basic normalization. On the other hand, it does not need any historical set of betting odds and match outcomes as the regression normalization. Moreover, it takes into account the long-shot bias phenomenon as the Shin method. The performance of the new procedure is evaluated using the betting odds provided by different professional bookmakers on more than 20.000 male tennis matches. The results show that the CaSco normalization has a better forecasting ability than that of the other approaches.

CC028 Room MAL 633 CONTRIBUTIONS IN APPLIED ECONOMETRICS Chair: Geoff Kenny

CC1380: **US monetary policy in a globalized world**
Presenter: **Martin Feldkircher**, Oesterreichische Nationalbank, Austria
Co-authors: Florian Huber

We analyze the Fed's response to global supply, demand and monetary policy shocks identified by conventional sign restrictions. Since financial globalization has increased over the recent years, monetary policymakers are nowadays less independent from external factors in controlling and setting domestic interest rates. Hence we use a time-varying framework by augmenting the global vector autoregressive model to allow for changes in parameters and residual variances. The proposed time-varying parameter stochastic volatility global vector autoregressive model (TVP-SV-GVAR) is then estimated for no less than 36 countries. To cope with such a data rich environment technically, we draw on a recent contribution on Cholesky stochastic volatility models. Preliminary results indicate that monetary policy in the USA does respond to global macroeconomic shocks. In particular, US short-term interest rates respond in a hump-shaped manner to global shocks to GDP growth with a peak response at around 2-3 quarters. By contrast, the response of US interest rates to global inflation shocks appears stronger in the 1980s and 1990s, compared to more recent periods.

CC1704: **Regional growth and business cycles in Japan**
Presenter: **Kazuhiko Kakamu**, Kobe University, Japan
Co-authors: Yoshihiro Ohtsuka

The consistency and gaps in national and regional business cycles in Japan are examined from a Bayesian point of view. The Tokyo monopolar system started in the mid-1970s, and recent descriptive statistics, such as migration and per capita income, show that the system continues, despite severe crises such as the bursting of bubbles and the Lehman shock. We explore the relationship between national and regional business cycles in the system using a spatio-temporal Markov-switching model with the Markov chain Monte Carlo method. Our empirical results show that, overall, the regional business cycle in the Kanto region, including Tokyo, is identical to the national business cycle. Moreover, we find that switches in the degree of spatial dependency occur around the turning points of business cycles, and that the degree of spatial dependency tends to be higher during a recession.

CC1269: **Untangling hotel industries inefficiency: An SFA approach applied to a renowned Portuguese hotel chain**
Presenter: **Nuno Ferreira**, ISCTE-IUL, Portugal
Co-authors: Manuela Maria Oliveira

The aim is to explore the technical efficiency of five hotels from Teixeira Duarte Group - a renowned Portuguese hotel chain. An efficiency ranking is established from these five hotel units located in Portugal using Stochastic Frontier Analysis. This methodology allows us to discriminate between measurement error and systematic inefficiencies in the estimation process enabling to investigate the main inefficiency causes. Several suggestions concerning efficiency improvement are undertaken for each hotel studied.

CC1657: **On trend inflation: An empirical analysis with a regime-switching model**
Presenter: **Sohei Kaihatsu**, Bank of Japan, Japan
Co-authors: Jouchi Nakajima

We propose a new econometric framework for estimating trend inflation and the slope of the Phillips curve with a regime-switching model. As a unique aspect of our approach, we assume regimes for the trend inflation at one-percent intervals, and estimate the probability of the trend inflation being in each regime. The trend inflation described in the discrete manner provides for an easily interpretable explanation of estimation results as well as a robust estimate. An empirical result indicates that Japan's trend inflation stayed at zero percent for about 15 years after the late 1990s, and then shifted away from zero percent after the introduction of the price stability target and the quantitative and qualitative monetary easing. The U.S. result shows considerably stable trend inflation at two percent since the late 1990s.

CC1186: **Depicting the dynamics of the trading process with the ACM-ACD model**
Presenter: **Katarzyna Bien-Barkowska**, Warsaw School of Economics, Poland

Modelling ultra-high-frequency (UHF) data stays at the forefront of financial econometrics. Recent access to intraday tick-by-tick financial data that might contain not only trade characteristics but also detailed information about buy/sell, market/limit orders allows researchers, practitioners and regulators to learn about trading strategies and their impact on the process of price formation at the micro level. We provide the empirical market microstructure study of interdependencies among different events (submissions of buy and sell market and limit orders and cancellations of limit orders) in the interbank FX market (Reuters Dealing 3000 Spot Matching System). The study documents striking regularities in the manner currency dealers submit and cancel their orders. From the econometric viewpoint, the sequence of different micro-events (orders) can be reflected

as the multidimensional marked point process. We apply the Autoregressive Conditional Multinomial Autoregressive Conditional Duration Model (ACM-ACD) model which allows us to capture jointly (1) clustering of order submissions and cancellations in time and (2) lead-lag dependencies between different order types (for example the significant positive cross-correlation between market buy (sell) orders and limit sell (buy) orders). We derive the news impact functions for the individual order categories and provide insight about high-frequency trading strategies used in FX markets.

CG351 Room MAL 421 CONTRIBUTIONS ON MCMC AND BAYESIAN ECONOMETRICS Chair: Catherine Forbes

CC1513: **Data driven particle filters for particle Markov chain Monte Carlo**
Presenter: **Patrick Leung**, Monash University, Australia
Co-authors: Catherine Forbes, Gael Martin

New automated proposal distributions are proposed for sequential Monte Carlo (SMC) algorithms, including particle filtering and related sequential importance sampling methods. The weights for these proposal distributions are easily established, as is the unbiasedness property of the resultant likelihood estimators, so that the methods may be used within a particle Markov chain Monte Carlo (PMCMC) inferential setting. Simulation exercises, based on a range of important financial models, are used to demonstrate the linkage between the signal to noise ratio of the system and the performance of the new particle filters, in comparison with existing filters. In particular, we demonstrate that one of our proposed filters performs well in a high signal-to-noise ratio setting, that is, when the observation is informative in identifying the location of the unobserved state. A second filter, deliberately designed to draw proposals that are informed by both the current observation and past states, is shown to work well across a range of signal-to-noise ratios. We then extend the study to a PMCMC setting in which we document the performance of the PMCMC algorithm using the new filters to estimate the likelihood function, once again in comparison with existing alternatives. The comparison is based on the optimal computing time required to estimate the posterior distribution of the parameter of interest.

CC1392: **The stock market and real economy: A Bayesian nonparametric approach**
Presenter: **Qiao Yang**, University of Toronto, Canada

The study of the joint dynamics between the stock market and real economy has a long history. A bivariate infinite hidden Markov model (IHMM) is proposed to investigate the joint dynamics of monthly SP500 returns and industrial production growth rates from 1926 to 2014. The bivariate IHMM is a nonparametric Bayesian approach which extends the vector autoregression (VAR) model to include Markov switching of infinite dimension. Comparing to conventional approaches, the bivariate IHMM is able to accommodate multivariate modelling, regime switching as well as structural changes in unified framework. The bivariate IHMM shows significant improvements in out-of-sample density forecasts as well as evidence of capturing structural change after 1984. In addition, we find past stock returns only have significant relation to future economic growth rates when regime- dependence is allowed. We show both jointly modelling and regime-dependence are two essential building blocks for fully capturing the joint dynamics of stock returns and real growth rates.

CC1685: **Electricity spot prices: A model based on stable CARMA processes and its Bayesian estimation**
Presenter: **Gernot Mueller**, Augsburg University, Germany
Co-authors: Armin Seibert

In recent years, electricity markets throughout the world have undergone massive changes due to deregulations. Extreme price volatility has forced producers and wholesale consumers to hedge not only against volume risk but also against price movements. Consequently, statistical modelling and estimation of electricity prices are an important issue for the risk management of electricity markets. We consider a model for the electricity spot price dynamics, which is able to capture seasonality, low-frequency dynamics and the extreme spikes in the market. In particular, we model the large fluctuations by a non-Gaussian stable CARMA process. After looking at a first stepwise estimation procedure we develop a Bayesian approach to fit the model to data. Furthermore, we apply the procedures to base load and peak load data from the German electricity exchange EEX.

CC1671: **Affine term structure models for the UK yield curve**
Presenter: **Konstantinos Bisiotis**, Athens University of Economics and Business, Greece

We propose an affine term structure model incorporating structural breaks for yield curves. In accordance with the recent literature we use macroeconomic variables as well as hidden unobserved variables for the modeling of stylized facts of the term structure of interest rates. Our starting point is a simple multifactor affine model which is extended to allow for regime switching and structural breaks using the Dirichlet process. We focus on the appropriate restrictions on the parameters of the model needed to guarantee well posedness for the affine models and considered potential issues arising in the identification and estimation procedures, which are based on minimum chi square estimation, and Bayesian methods. The model is tested on data related to the UK yield curve.

CC1836: **Bayesian analysis in non-linear non-Gaussian state-space models using particle Gibbs**
Presenter: **Oliver Grothe**, Karlsruhe Institute of Technology, Germany
Co-authors: Roman Liesenfeld, Tore Selland Kleppe

We consider the Particle Gibbs (PG) procedure as a tool to perform a Bayesian analysis of non-linear, non-Gaussian state space model, which combines MCMC techniques with Sequential Monte Carlo (SMC). A potential advantage of the PG approach is that sampling from the states given parameters and observations can be partitioned into a sequence of smaller sampling problems. However, PG usually suffers from very poor mixing. We propose to improve its efficiency by relying upon a SMC procedure known as Particle Efficient Importance Sampling (PEIS) recently developed. PEIS combines SMC with an Efficient IS (EIS) procedure and exploits that EIS produces by a sequence of auxiliary regressions close global density approximations to potentially high-dimensional target densities. We show that, within PG, PEIS substantially improves the particle approximation relative to existing approaches, e.g., Ancestor Sampling.

CG021 Room MAL 541 CONTRIBUTIONS IN DYNAMIC FACTOR ANALYSIS Chair: Michele Lenza

CC1517: **A shadow policy rate to calibrate US monetary policy at the zero lower bound**
Presenter: **Marco Lombardi**, Bank for International Settlements, Switzerland
Co-authors: Feng Zhu

The recent global financial crisis, the Great Recession and the subsequent implementation of a variety of unconventional policy measures have raised the issue of how to correctly measure the stance of monetary policy when policy interest rates reach the zero lower bound (ZLB). We propose a new "shadow policy rate" for the US economy, using a large set of data representing the various facets of the US Federal Reserve's policy stance. Changes in term premia at various maturities and asset purchases by the Fed are key drivers of this shadow rate. We document that our shadow policy rate tracks the effective federal funds rate very closely before the recent crisis. More importantly, it provides a reasonable gauge of US monetary policy stance when the ZLB becomes binding. This facilitates the assessment of the policy stance against familiar Taylor rule benchmarks. Finally, we show that in structural vector autoregressive (VAR) models, the shadow policy rate helps identify monetary policy shocks that better reflect the Federal Reserve's unconventional policy measures.

CC1734: **Maximum likelihood estimation of time-varying factor loadings in high-dimensional factor models**
Presenter: **Jakob Mikkelsen**, Aarhus University, Denmark
Co-authors: Eric Hillebrand, Giovanni Urga
We consider the maximum likelihood estimation of a parametric model for loadings in factor models of high dimension. We specify the loadings to evolve as stationary vector autoregressions and show that consistent estimates can be obtained by a two-step estimation procedure. In the first step principal components are extracted from the data to form factor estimates. In the second step we estimate a set of univariate regression models with time-varying coefficients for the observable data and unobservable factors. Replacing the unobservable factors with principal components gives a feasible likelihood function that is asymptotically equivalent to the infeasible likelihood function with unobservable factors and therefore gives consistent estimates of the loading parameters.

CC1770: **Bootstrap inference for structural analysis in factor models**
Presenter: **Victor Bystrov**, University of Lodz, Poland
Alternative bootstrap procedures of constructing confidence bands for responses to structural shocks in latent factors are compared. Factors are estimated by the method of principal components and identified using restrictions which facilitate economic interpretation of structural shocks. The coverage ratios of the confidence bands, which are computed using alternative procedures, are compared in a Monte Carlo exercise. The relative role of the bias, caused by a small time-series dimension, and the bias, caused by a small cross-sectional dimension, is discussed. An empirical example is supplied.

CC1477: **The effects of monetary policy on industry-level stock returns in a changing world**
Presenter: **Pierre Guerin**, Bank of Canada, Canada
Co-authors: Danilo Leiva-Leon
The effects of monetary policy shocks on asset prices is investigated. A key aspect of our analysis is that we study the response of stock returns to monetary policy shocks at the industry-level (rather than considering the aggregate stock market response). Our analysis shows that there is a substantial degree of heterogeneity in the responses to monetary policy shocks across industries. In doing so, we first focus on high-frequency (daily) regressions, using an event-study approach to identify monetary policy shocks. Second, when investigating lower frequency (quarterly) responses based on a time-varying factor-augmented VAR model, we uncover evidence in favor of variation both over time and across industries in the responses to a monetary policy shock. Moreover, we find that responses to monetary policy shocks are best explained by the cyclicality, volatility and the interconnectedness of the returns of a given industry. Another contribution is methodological in that we show how to estimate a time-varying factor-augmented VAR model with stochastic volatility and time-varying factor loadings, which is helpful to assess changes in the cyclicality of industry-specific returns.

CC0275: **Country shocks, monetary policy expectations and ECB decisions: A dynamic nonlinear approach**
Presenter: **Danilo Leiva-Leon**, Central Bank of Chile, Chile
Previous studies have shown that the effectiveness of monetary policy depends, to a large extent, on the market expectations of its future actions. An econometric framework is proposed to address the effect of the current state of the economy on monetary policy expectations. Specifically, we study the effect of contractionary (or expansionary) demand (or supply) shocks hitting the Euro area countries on the expectations of the ECB's monetary policy in two stages. In the first stage, we construct indexes of real activity and inflation dynamics for each country, based on soft and hard indicators. In the second stage, we use those indexes to provide assessments on the type of aggregate shock hitting each country and assess its effect on monetary policy expectations at different horizons. Our results indicate that expectations are responsive to aggregate contractionary shocks, but not to expansionary shocks. Particularly, contractionary demand shocks have a negative effect on short term monetary policy expectations, while contractionary supply shocks have negative effect on medium and long term expectations. Moreover, shocks to different economies do not have significantly different effects on expectations, although some differences across countries arise.

CG457 Room MAL 540 CONTRIBUTIONS IN INFLATION ANALYSIS AND FORECASTING Chair: Maritta Paloviita

CC1663: **Inflation forecasts: On market-based and survey-based measures**
Presenter: **Magdalena Grothe**, European Central Bank, Germany
The aim is to analyse the predictive power of market-based and survey-based inflation expectations for actual inflation. We use the data on inflation swaps and the forecasts from the Survey of Professional Forecasters for the Euro area and United States. The results show that both, market-based and survey-based measures have a non-negligible predictive power for inflation developments, as compared to statistical benchmark models. Therefore, for horizons of one and two years ahead, market-based and survey-based inflation expectations actually convey information on future inflation developments.

CC1374: **On neural networks in forecasting inflation**
Presenter: **Karol Szafranek**, Narodowy Bank Polski, Poland
Forecasting inflation is a crucial ingredient in monetary policy conduct. Forecasting of the Polish headline inflation is investigated by using a large number of randomly generated artificial neural networks. Using principal component analysis we extract the relevant information from an extensive macroeconomic database and use bootstrap averaging to generate 10000 forecasts from three kinds of artificial neural networks: a single feedforward, an Elman recurrent and a multilayer perceptron with the number of neurons in the hidden layer following the zero-truncated Poisson distribution. CPI point forecasts are calculated by adopting different combination techniques from projected densities. Results advocate that in majority of horizons forecasts generated from a neural network forest significantly outperform a panel of benchmark projections (judgment forecasts, dynamic factor model, Bayesian autoregressive model, leading indicator model, new-Keynesian hybrid Philips curve model, simple autoregressive process and combinations of aforementioned). Considerable reduction of forecast error statistics are confirmed by the HLN-DM test statistics. Moreover, simple average forecasts generally yield lowest errors among distinguished combination forecasts, although the discrepancies between different kinds of averaging are negligible. Robustness check suggests also that moderately parsimonious architecture of neural networks hidden layer is beneficial for forecasting accuracy.

CC1607: **Short term inflation forecasting: The M.E.T.A. approach**
Presenter: **Andrea Silvestrini**, Bank of Italy, Italy
Co-authors: Giacomo Sbrana, Fabrizio Venditti
Forecasting inflation is an important and challenging task. We assume that the core inflation components evolve as a multivariate local level process. This model, which is theoretically attractive for modelling inflation dynamics, has been used only to a limited extent to date owing to computational complications with the conventional multivariate maximum likelihood estimator, especially when the system is large. We propose the use of a method called Moments Estimation Through Aggregation (M.E.T.A.), which reduces computational costs significantly and delivers fast and accurate parameter estimates, as we show in a Monte Carlo exercise. In an application to Euro area inflation, we find that our forecasts compare well with those generated by alternative univariate and multivariate models as well as with those elicited by professional forecasters.

CC0322: **Assessing the economic value of a probabilistic forecast for inflation in the presence of an inflation target**
Presenter: **Shaun Vahey**, Warwick University, United Kingdom
Co-authors: Liz Wakerly, Craig Thamotheram, Christopher McDonald
Central banks devote considerable resources to modelling and assessing the probabilities of inflation events. We consider the fundamental issue of what makes a "good" probability forecast for a central bank operating within an inflation targeting framework. It is well-known that the utility of a forecast depends on the decision making problem of the user. We provide two examples in which conventional measures of forecast performance give misleading information about the usefulness of a forecast from the perspective of an inflation targeting central bank wishing to communicate ex ante inflation risks to the public. Evaluations based on root mean squared prediction error and the average logarithmic score neglect the relative costs of actions available to the policymaker in our decision-theoretic framework. Two examples, one using actual forecasts for inflation the United Kingdom published by the Bank of England, the other using forecasts for inflation in New Zealand generated from vector autoregressions, illustrate the prevalence and importance of the issue for monetary policymakers concerned with inflation targeting.

CC026 Room MAL 402 CONTRIBUTIONS IN TIME SERIES ECONOMETRICS Chair: Anders Rahbek

CC1484: **Mean reversion and stationarity: A new perspective from the asymptotics of diffusion models**
Presenter: **Jihyun Kim**, Toulouse School of Economics, France
Co-authors: Joon Park
The mean reversion and unit root properties of general diffusion models and their discrete samples are analyzed. In particular, we find that the Dickey-Fuller unit root test applied to discrete samples from a diffusion model becomes a test of no mean reversion rather than a unit root, or more generally, nonstationarity in the underlying diffusion. The unit root test has a well defined limit distribution if and only if the underlying diffusion has no mean reversion, and diverges to minus infinity in probability if and only if the underlying diffusion has mean reversion. It is shown, on the other hand, that diffusions are mean-reverting as long as their drift terms play the dominant role, and therefore, nonstationary diffusions may well have mean reversion.

CC1792: **Detrended fluctuation analysis as a regression framework: Estimating dependence at different scales**
Presenter: **Ladislav Kristoufek**, Czech Academy of Sciences, Czech Republic
Detrended fluctuation analysis (DFA) was introduced in early 1990s as a method for analyzing fractal properties of underlying data. The method was later popularized in the long-range correlations and multifractal analyses. Recently, DFA has been generalized for long-range cross-correlation analysis as well as an examination of correlations between nonstationary series. We propose a framework based on the detrended fluctuation analysis which allows for a regression analysis of possible nonstationary and long-range dependent data at different scales. The methodology is based on the least squares framework, which is briefly recalled and translated into the language of variances and covariances. The detrended fluctuation analysis together with its bivariate generalization of the detrended cross- correlation analysis (DCCA) are described in some detail as a connecting bridge to the DFA-based regression. The DFA framework is introduced for the bivariate setting with procedures to estimate parameters, standard errors of the estimates, and the coefficient of determination (R^2), all characteristic for a specified scale. The theoretical concepts are further supported by Monte Carlo simulations. The framework is then applied to several phenomena from various disciplines: relationship between temperature and humidity, stock market betas, elasticity between corn and ethanol, and transmission between influenza outbursts and the Google Flu Trends indicator.

CC0787: **Towards multivariate seasonal adjustment and joint outlier treatment in a system of hierarchical time series**
Presenter: **Suad Elezovic**, Statistics Sweden, Sweden
A practical approach for joint seasonal adjustment in a system of hierarchical time series is illustrated. The proposed procedure is performed in two steps; a preparation step where the nature of relationships within groups of time series is investigated and the estimation step, involving actual simultaneous seasonal adjustment through a multivariate procedure. Then, the results from multivariate seasonal adjustment are compared with a standard univariate method through a simulation study and through a real data example from the Swedish National Accounts. This approach is based on an assumption that a standard univariate seasonal adjustment might be improved by utilizing the dependence structure among related components within a given set of time series. In particular, the trend-cycle components are studied in more detail in order to identify joint breaks in several time series. Joint treatment of level-shift outliers is one of the major issues in the preparatory step of the proposed procedure. Furthermore, a simulation procedure used for identification of optimal critical values for outlier detection is illustrated and the results are discussed.

CC1738: **Causality tests and delay measures in a frequency band**
Presenter: **Sven Schreiber**, Macroeconomic Policy Institute IMK and Free U Berlin, Germany
Co-authors: Joerg Breitung
We first extend the frequency-specific Granger causality test from the literature to a more general null hypothesis that allows non-causality at unknown frequencies within an interval, instead of having to prespecify a single frequency. This setup corresponds better to most hypotheses that are typically analyzed in applied research and is easy to implement within a linear VAR framework, while still being statistically rigorous. The resulting test is shown to be somewhat conservative (hence not exceeding its nominal size) but with generally good power. In an empirical application dealing with the dynamics of US temperatures and CO2 emissions we find that emissions cause temperature changes only at very low frequencies with more than 30 years of oscillation. Secondly we analyze the frequency-specific time shift of a target variable with respect to the Granger-causal input variable. In order to assess the corresponding estimation uncertainty of this lead/delay measure we make use of the fact that the time shift in the frequency domain is a function of the impulse response coefficients via the phase shift, after applying appropriate adjustments to the phase shift for this particular case of a causal (one-sided) filter. Hence standard errors are provided by applying the analytical delta approximation and through simulation-based techniques.

CC1365: **Multivariate spurious long memory and a robust local Whittle estimator**
Presenter: **Christian Leschinski**, Leibniz University Hannover, Germany
Co-authors: Philipp Sibbertsen
The estimation of memory parameters is considered in multivariate processes that are subject to low frequency contamination such as mean shifts or trends. For univariate series, it is well established that they low frequency contamination can generate spurious long memory. We show that this also applies to the multivariate case. To do so we consider a multivariate generalization of the random level shift process and derive the expectation of its periodogram. We then derive a multivariate version of the local Whittle estimator, that is modified so that it is robust to spurious long memory. The estimator is consistent and asymptotically normal distributed. Due to the form of the modification, the estimator is not simply a multivariate extension of the available univariate methods. In fact, our Monte Carlo study shows that in the univariate special case our estimator outperforms its univariate competitors in most situations. In the multivariate case, the Monte Carlo results demonstrate that in contrast to standard multivariate local Whittle methods (or GSE estimators), our estimator successfully mitigates the impact of level shifts. An application to log-absolute returns of the SP500 shows that part of the memory in this volatility series can be attributed to low frequency contamination.

CG023 Room MAL 539 CONTRIBUTIONS ON STOCHASTIC VOLATILITY Chair: Drew Creal

CC1729: **Using high-frequency returns in the Bayesian estimation of stochastic-volatility jump-diffusion models**
Presenter: **Milan Ficura**, University of Economics in Prague, Czech Republic
Co-authors: Jiri Witzany
A methodology is presented of how to utilize high-frequency returns in the Bayesian estimation of Stochastic-Volatility Jump-Diffusion (SVJD) models with self-exciting jumps. Two different approaches are developed and compared. In the first approach the SVJD models are applied directly to the intraday price returns which requires a modification of the models in order to handle the intraday seasonality of volatility and jump intensity in the analyzed time series. In the second approach, non-parametric power-variation estimators are used as a source of additional information in the Bayesian estimation of the SVJD models. The previous methods, using realized variance, are extended in order to incorporate also the non-parametric jump estimators into the models. The performance of the newly developed models is assessed based on their in-sample fit to the data as well as out-of-sample predictive accuracy. The results indicate that the use of high-frequency data in the estimation of SVJD models presents a highly promising area for future research and applications.

CC0457: **Baysian non-linear jump-GARCH models with self-excitation in high frequency finance via SMC samplers**
Presenter: **Alan Hawkes**, Swansea University, United Kingdom
Co-authors: Jing Chen, Gareth Peters, Mike Buckle
The aim is to incorporate the one-dimension Hawkes process self-excitation into Non-linear Jump GARCH diffusion structures which mixes both daily and intra-daily information into assessment of the returns series and the volatility structures. We consider two specifications of the likelihood, one in which a pre-processing of the data is performed which involves a jump detection analysis and then the Hawkes Jump-NGARCH model is calibrated conditionally on the pre-processing. The second approach involves specification of a stochastic model for the number of intra-daily jumps and the jump sizes. In addition, we consider a Bayesian formulation and calibrate this model via a specialised SMC Sampler formulation which exploits an unbiased Monte Carlo estimator of the likelihood in the estimation.

CC1442: **Bayesian semiparametric modeling of multivariate stochastic volatility**
Presenter: **Martina Danielova Zaharieva**, University of Muenster, Germany
The proposed model is a multivariate stochastic volatility model (MSV), in which the errors are modeled as an infinite scale mixture of multivariate Gaussian distributions. A Bayesian non-parametric approach, in particular a Dirichlet process mixture (DPM), is adopted. This allows for highly flexible modeling of the return distribution with respect to the kurtosis. The structure of the MSV model is based on Cholesky decomposition, which simplifies the estimation of the latent volatility states and allows the use of standard filtering methods. For the DPM part of the model an efficient sampling algorithm related to the slice sampler is proposed. Finally, an empirical application regarding volatility shock transmissions between different markets is applicable.

CC1633: **A GMM-RM estimation of the GARCH jump diffusion model**
Presenter: **Isao Ishida**, Konan University, Japan
Co-authors: Shuichi Nagata
We investigate GMM estimation of the GARCH jump diffusion model for asset price processes based on moments of daily returns and realized measures (RMs). For this purpose, we augment the Corradi-Distaso GMM-RM estimator with additional moments for estimating the leverage parameter, and also propose an alternative version based on conditional moments. The GARCH diffusion has enough analytical tractability to render GMM-RM estimators easy to implement even without the help of a spot volatility proxy, not guaranteed to be available for all assets, while the recently proposed approximate closed-form maximum likelihood estimators rely on either an options-implied volatility as a proxy for the unobserved spot volatility or computationally intensive techniques for integrating the spot volatility out of the likelihood. We report the empirical results of applying our GMM-RM estimators to high-frequency S&P 500 data. We carefully treat microstructure noise in high-frequency overvations and the existence of overnight non-trading hours.

CC1553: **The forecast of financial volatility using volume-scaled returns through stochastic volatility models and intraday data**
Presenter: **Antonio Santos**, University of Coimbra, Portugal
The volatility is one of the most studied subjects in the financial literature. Model-free and model-dependent measures have been proposed to measure the evolution of the volatility through time. With the availability of intraday data, measures like the realized volatility have gained considerable attention. The environment associated with the transactions in the financial markets constitutes a challenge to the realized volatility measure, where there is a clear departure from the theoretical setting associated with the definition of the measure. Using intraday data, the aim is to characterize the evolution of the intraday volatility through models commonly used to characterize the volatility evolution at lower frequencies. We use stochastic volatility models to characterize and forecast the evolution of intraday volatility, where the main novelty is the utilization of volume-scaled returns. All characteristics amenable of being characterized by the stochastic volatility model are found in this kind of returns, and volatility forecasts at higher frequencies are made possible of being obtained.

EO262 Room Beveridge Hall NEW TRENDS IN SURVIVAL ANALYSIS Chair: Anouar El Ghouch

EO0500: **P-spline estimation in varying coefficient models with right censored data**
Presenter: **Anneleen Verhasselt**, Hasselt Univeristy, Belgium
Co-authors: Paul Janssen, Kim Hendrickx
Modeling the relationship between the response and one or more covariates is an important activity in statistics. Several regression models have been proposed in the literature among which the classical linear regression model is the simplest model. Varying coefficient models provide an extension of the classical linear model; they are still linear in the covariates, but the regression coefficients are smooth functions in one or more other variables. The attraction of these models is the ability of capturing complex relationships in the data. In a variety of scientific fields responses are subject to random right censoring, among others time-to-event in medicine (survival data). We estimate the varying coefficients - for random right censored data - using P-splines. To apply the P-spline method in case responses are subject to right censoring we first transform the observed response (which can be a time-to-event or a censored observation) into a 'synthetic' response, where the conditional means (conditioning is on the covariates) of the time-to-event and the synthetic response are matched. The P-spline method is then applied to the transformed data. We study the asymptotic properties of our estimators and discuss how to select the smoothing parameters and the transformation parameter in a data-driven way. In addition, we demonstrate with simulated and real data the finite sample behavior of the estimators.

EO0810: **Estimating the error distribution function in heteroskedastic nonparametric regression with cure model data**
Presenter: **Justin Chown**, Universite catholique de Louvain, Belgium
Conceptual aspects of nonparametric estimation in the context of cure model data is first introduced: model identifiability, estimation efficiency and estimator construction. This type of data presents many interesting theoretical challenges due to censoring and inescapable dependence on covariates. Then a suitable estimator of the distribution function of the errors from a heteroskedastic nonparametric regression of cure model

data is discussed. The approach specifically avoids the popular technique of constructing residuals, which is counter-intuitive. Rather, we use the underlying relationship between the unconditional distribution function of the errors and the conditional distribution function of the responses given the covariates. This approach is shown to be consistent at the so-called parametric rate of convergence.

EO0945: **A two-stage estimation procedure to analyze bivariate left-censored zero-inflated semi-continuous lifetimes**
Presenter: **Roel Braekers**, Hasselt University, Belgium
Co-authors: Yves Grouwels

In some clinical or environmental studies, researchers are interested in a bivariate set of semi-continuous lifetimes. Hereby the outcome variables take a value zero with a discrete probability or a non-zero value from a continuous distribution. Due to the measuring mechanism observing an outcome variable is not fully possible and each of the outcomes are left-censored. We introduce a bivariate mixture regression model to analyze this kind of data. We consider parametric regression models for the marginal zero-response probabilities and their association. For the non-zero continuous distribution, we introduce both a parametric regression model and a semi-parametric Cox's regression model. We model the association between the different continuous distributions through a copula function. Since maximizing a full likelihood function is computational difficult, we consider a two-stage estimation procedure. As results, we obtained the asymptotic normality of the association parameters. Next we compare the finite sample behaviour of the different estimators through a simulation study. Furthermore we illustrate the estimators on a practical example on ethanol-induced sleeping time in mice.

EO1241: **Nonparametric estimation of the conditional survival function for ordered multivariate failure time data**
Presenter: **Luis Machado**, University of Minho, Portugal

In longitudinal studies of disease, patients may experience several events through a follow-up period. In these studies, the sequentially ordered events are often of interest and lead to problems that have received much attention recently. Issues of interest include the estimation of bivariate survival, marginal distributions and the conditional distribution of gap times. We consider the estimation of the survival function conditional to a previous event. Different nonparametric approaches will be considered for estimating these quantities, all based on the Kaplan-Meier estimator of the survival function. We explore the finite sample behavior of the estimators through simulations. Real data illustration based on a German breasts cancer study is included.

EO651 Room Woburn NON- AND SEMI-PARAMETRIC FUNCTIONAL STATISTICS II Chair: Aldo Goia

EO0710: **Recursive estimation of the geometric median and of the median covariation matrix in Hilbert spaces**
Presenter: **Antoine Godichon**, Universite de Bourgogne, France
Co-authors: Herve Cardot

With the development of automatic sensors, it is more and more usual to study large samples of observations taking values in high dimensional spaces (such as functional spaces) that also may be observed sequentially. Moreover, these large samples are often contaminated. In this context, the Geometric Median (or L^1-median), which is a generalization of the real median for metric spaces, and the Median Covariation Matrix are interesting robust location and dispersion parameters. Efficient and fast recursive estimators of the median will be considered: a stochastic gradient algorithm and its averaged version. Note that some results on their convergence exist in the literature, but, new ones, as L^p rates of convergence, will be given. Then, an averaged stochastic gradient algorithm will be given for the estimation of the Median Covaration Matrix as well as its convergence in quadratic mean.

EO0587: **Hyperspectral image segmentation based on functional kernel density estimation**
Presenter: **Laurent Delsol**, University of Orleans, France
Co-authors: Cecile Louchet

Splitting a picture into a set of homogenous regions is a common problem, called segmentation, in image analysis. The detection of such regions is usually a relevant way to identify specific parts of the scene. Various methods have been proposed to segment gray-level or multispectral images. The maximum a posteriori approach, based on Potts random field as prior and density estimation on each region, is an interesting use of Bayesian statistics in that domain. On the other hand, a great variety of functional statistical methods are nowadays available to deal with data sets of curves. The kernel density estimator has been adapted to such data. We focus on hyperspectral images for which each pixel is described through a curve (discretized on a thin grid) and discuss the way functional kernel density estimation and maximum a posteriori approach may be combined. The choice of granularity and smoothing parameters will be discussed. And we will present some simulations and applications on real world images.

EO1430: **Clustering multivariate functional data with phase variation**
Presenter: **Juhyun Park**, Lancaster University, United Kingdom
Co-authors: Jeongyoun Ahn

We consider the problem of clustering multivariate functional data when phase variation prevails and amplitude variation is considered as a nuisance. The issues with amplitude and phase variations are relatively well understood in one dimensional curves as vertical and horizontal variabilities, but its extension to multivariate functional data is not well established. In addition, most approaches developed in this context have taken the view that phase variation is a nuisance and thus needs to be adjusted for in explaining amplitude variation. Since these functions are not comparable, the same argument does not apply to the complementary situation as ours. We reformulate the phase variation in multivariate functional data with task-appropriate parametrization of the warping functions and propose a simple solution to clustering multi-functional objects based on phase variation only. We illustrate our method using simulation studies as well as a real data application with multivariate growth curves.

EO0899: **An inferential theory of clustering for functional data**
Presenter: **Mattia Ciollaro**, Carnegie Mellon University, United States
Co-authors: Christopher Genovese, Daren Wang

Recently, it has been shown that Morse theory can be exploited to develop a sound inferential background for clustering: one can rigorously define both population and empirical clusters by means of the gradient flows associated to the population density p and the estimated density $\hat{p}$. In this framework, clusters are well-defined entities corresponding to the basins of attraction of the density's critical points. The population parameter of interest corresponds to the collection of population clusters associated to p, and a natural estimator is given by the collection of empirical clusters associated to $\hat{p}$. While this framework is already well-developed for finite-dimensional random vectors, very little is known about the extent to which this theory generalizes to infinite-dimensional random variables (functional data). We discuss the challenges that functional data pose for the development of a clustering inferential paradigm based on Morse theory, and we describe how these challenges can be addressed. We focus on smooth random curves, but the theory that we develop can be further extended to more complex functional data types. We will also discuss how population clusters of functional data can be estimated in practice and how to assess the statistical significance of the estimated clusters.

EC032 Room Bloomsbury CONTRIBUTIONS IN DIMENSION REDUCTION Chair: Bing Li

EC1350: **Independent component analysis for matrix-valued data**
Presenter: **Joni Virta**, University of Turku, Finland
Co-authors: Bing Li, Klaus Nordhausen, Hannu Oja
In preprocessing high-dimensional matrix-valued data, e.g. images, a common procedure is to vectorize the observed matrices and subject the resulting vectors to one of the many methods used for independent component analysis (ICA). However, the structure of the original matrix is lost in the vectorization along with any meaningful interpretations of the rows and columns. To provide a more suitable alternative, we propose the Folded fourth order blind identification (FOLD), a matrix-valued analogy of the classic Fourth order blind identification (FOBI). In FOLD, instead of vectorizing, we stay in the matrix form and in a sense perform FOBI simultaneously on both the rows and the columns of the observed matrices. Furthermore, being an extension of FOBI, FOLD shares with it its computational simplicity. A simulated example is used to showcase the method's usefulness in discriminant analysis.

EC1416: **Sliced inverse regression for time series**
Presenter: **Markus Matilainen**, University of Turku, Finland
Co-authors: Christophe Croux, Klaus Nordhausen, Hannu Oja
When analysing data with a response variable y and explanatory variables $\mathbf{x}$, modelling may become infeasible when number of variables gets higher. It can also cause computational problems and visualization of data becomes harder. To avoid this kind of problems we can use Sliced Inverse Regression (SIR), which is a supervised dimension reduction method and it is used to study a relationship between the response y and the variables $\mathbf{x}$. However, in case of time series SIR algorithm does not use any information on lagged values directly. One way to deal with this is to treat explanatory variables and their past values and the past values of response variable as explanatory variables. We suggest a new method, which is based on SIR algorithm, but instead of using regular supervised covariance matrices, we use lagged supervised covariance matrices. We show how our new algorithm works with a few illustrative examples.

EC1384: **One-stage estimation of principal component regression with sparse regularization**
Presenter: **Shuichi Kawano**, The University of Electro-Communications, Japan
Co-authors: Hironori Fujisawa, Toyoyuki Takada, Toshihiko Shiroishi
Principal component regression (PCR) is a two-stage procedure that selects some principal components and then constructs a regression model regarding them as new explanatory variables. Note that the principal components are obtained from only explanatory variables and not considered with the response variable. To address this problem, we propose the sparse principal component regression (SPCR) that is a one-stage procedure for PCR. SPCR enables us to adaptively obtain sparse principal component loadings that are related to the response variable and select the number of principal components simultaneously. SPCR can be obtained by the convex optimization problem for each parameter with the coordinate descent algorithm. Numerical results are presented to illustrate the effectiveness of SPCR.

EC1711: **Three-way principal component analysis for obtaining perfectly simple loadings**
Presenter: **Kohei Uno**, Osaka University, Japan
Three-way principal component analysis (PCA) is a dimension reduction method for analyzing a three-way data array of $I \times J \times K$. In a famous three-way PCA procedure Tucker3, the data array is decomposed into three component loading matrices, A ($I \times P$), B ($J \times Q$) and C ($K \times R$) and core array G ($P \times Q \times R$). A strategy for obtaining interpretable sparse loadings may be to incorporate an L_1 penalty into Tucker3. However, it leads to a problem that the sparse loading matrices resulting in the L_1-penalized Tucker3 are not column-orthonormal. It is not convenient for interpretation. In order to overcome the problem, we propose a new Tucker3 procedure without a penalty function. In the proposed method, component matrices are constrained to be column-orthonormal and have interpretable sparse structure called perfect simple structure. Under those constraints, the Tucker3 loss function is minimized with an alternating least square (ALS) algorithm. A key point in the algorithm is reparameterizing each component loading matrix as the product of a column-orthonormal one and a binary matrix standing for the membership of variables to components.

EC1537: **Inverse moments and machine learning for sufficient dimension reduction**
Presenter: **Andreas Artemiou**, Cardiff University, United Kingdom
Recent advances in dimension reduction promote the use of SVM and its variants for dimension reduction in regression problems with high dimensional predictors. The use of classifiers which aim to improve the accuracy of classification, in a dimension reduction framework where we aim to align the estimated surface with the actual regression surface, can be questioned; mainly due to the fact that some artificially selected inliers play more role than the rest of the observations. We discuss a different approach which reweights all the points and creates more robust dimension reduction methods.

EG630 Room Holden CONTRIBUTIONS ON VARIABLE SELECTION Chair: Yichao Wu

EC1354: **Statistical analysis of data from restricted randomised experiments via shrinkage methods**
Presenter: **Sadiah Aljeddani**, University of Southampton, United Kingdom
Co-authors: Kalliopi Mylona, Steven Gilmour
For model selection purposes in experimental contexts, researchers often use stepwise regression or subset selection. In situations involving restricted randomization, such as block experiments and split-plot experiments, this has to be done manually and, especially for experiments involving multiple responses, it involves numerous model estimations. We propose a modification of penalized least squares, for the analysis of data from experiments conducted under restricted randomization, which performs model selection and model estimation simultaneously. Under the penalized generalized least squares, non-significant variables are more likely to be dropped out the models. We demonstrate the usefulness of the approach using various practical examples, and study its properties in a simulation study.

EC1617: **Variable selection for model-based clustering using the integrated-complete-data likelihood**
Presenter: **Mohammed Sedki**, Paris-Sud university, France
Co-authors: Matthieu Marbac
Variable selection in cluster analysis is important yet challenging. It can be achieved by regularization methods, which realize a trade-off between the clustering accuracy and the number of selected variables by using a lasso-type penalty. However, the calibration of the penalty term can suffer from criticisms. Model selection methods are an efficient alternative, yet they require a difficult optimization of an information criterion which involves combinatorial problems. First, most of these optimization algorithms are based on a suboptimal procedure (e.g. stepwise method). Second, the algorithms are often greedy because they need multiple calls of EM algorithms. We propose to use a new information criterion based on the integrated complete-data likelihood. It does not require the maximum likelihood estimate and its maximization appears to be simple and computationally efficient. The original contribution of our approach is to perform the model selection without requiring any parameter estimation. Then, parameter inference is needed only for the unique selected model. This approach is used for the variable selection of a Gaussian mixture model with conditional independence assumption. The numerical experiments on simulated and benchmark datasets show that the proposed method

often outperforms two classical approaches for variable selection. The proposed approach is implemented in the R package VarSelLCM available on CRAN.

EC1567: **Positive unlabelled feature selection using constrained affinity graph embedding**
Presenter: **Yufei Han**, Symantec Research Lab, France
Co-authors: Yun Shen

In real-world binary classification scenarios, such as network intrusion detection and web page categorisation, samples from negative class usually requires prohibitive overheads to label. Only a small proportion of positive data can be labelled explicitly by trustable oracles. Selecting relevant and non-redundant features given limited positively labelled data and unlabelled data is therefore highly desirable for accurate classification. So far as we know, no previously similar study attacks this problem. The proposed positive-unlabelled feature selection method learns a L_1-norm regularised robust linear regression of a constrained spectral graph embedding of training data on the feature representation. A set of noisy but informative must-link and cannot-link constraints are extracted using the given positively labelled samples and affinity graph of training data. These constraints are used to generate a constrained spectral graph embedding of training data, injecting partial supervision information into the feature selection procedure. The robust L_1-norm regularised regression model originates from correntropy theory. It is designed to suppress the impacts of noise in the pairwise constraints and identify the most informative features corresponding to the non-zero regression coefficients simultaneously. Experiments on two public benchmark data sets and one real-world network intrusion data set verify the method.

EC1756: **Variable selection through sparse principal component analysis for a multinomial logit model**
Presenter: **Antonio Lucadamo**, University of Sannio, Italy
Co-authors: Luca Greco

In chemometrics, soils are often classified, through laboratory chemical analysis, according to their texture. This procedure is time and cost expensive and, for this reason, techniques as, for example, Linear Discriminant Analysis or MultiNomial Logit analysis are often preferable. In these methods the values of the reflectance at different wave lengths are used as explicative variables, and the textural classes are the modalities of the dependent variable. In this way, the obtained estimates can be used to predict new soils. Anyway, the use of many explicative variables, often highly correlated, can cause some problems in parameter estimation and in classification. It is then necessary to choose a criterion that can help in reducing the dimensionality and in selecting variables useful for the classification aim. Our idea is then to use Sparse Principal Component Analysis as the variable selection tool for Multinomial Logit Model.

EC1463: **The 100-fold cross validation for small sample method**
Presenter: **Shuichi Shinmura**, Seikei, Japan

We propose a new model selection procedure of the discriminant analysis. Fisher formulated a linear discriminant function (LDF). He never formulated the standard error of error rate and discriminant coefficient. Therefore, there was no sophisticated model selection procedure the same as the regression analysis. Until now, we could use the leave-one-out method. Now, we can use a powerful computer power and software. We had already proposed the k-fold cross-validation for small sample method. We tried to compare eight LDFs by two means of error rates such M1s and M2s in the training and validation samples. Moreover, we got the precise results of the comparison of eight LDFs. We proposed to select the model with the minimum M2 among all possible combination of variables as the best model. Moreover, we compare a revised Optimal LDF by Integer Programming (Revised IP-OLDF) based on the minimum number of misclassification (MNM) criterion with other seven LDFs by the best model. We confirm the new model selection procedure by the several data and have significant results as follows:1) M2s of Fisher LDF are worse than Revised IP-OLDF.2) A soft-margin SVM for penalty c=1 is worse than another mathematical programming based LDFs and logistic regression. 3) We obtain the remarkable results about the 95% CI of the coefficients.

EG652 Room G21A CONTRIBUTIONS ON STATISTICAL MODELLING FOR ECONOMICS AND FINANCE Chair: Johan Lyhagen

EC1433: **COGARCH(p,q) models in the yuima package: Application to financial time series**
Presenter: **Lorenzo Mercuri**, University of Milan, Italy
Co-authors: Stefano Iacus, Edit Rroji

We present the simulation and the estimation of a COGARCH(p,q) model in the R package yuima. Several routines are introduced. Indeed for the generation of a COGARCH(p,q) trajectory, the user can choose between two alternative schemes. The first is based on the Euler discretization of the stochastic differential equations that identifies a COGARCH(p,q) model while the second one considers the explicit solution of the equations defining the variance process. Estimation is based on the matching of the empirical with the theoretical autocorrelation function. In this case three different approaches are implemented: minimization of the mean square error, minimization of the absolute mean error and the generalized method of moments where the weighting matrix is continuously updated. In the empirical analysis, we show how estimate the model using daily data and then generate the sample paths of the process with intraday frequency. We compute different risk measures (value at risk and expected shortfall) and we provide a comparison with GARCH(p,q) models.

EC1066: **A novel statistical and machine learning hybrid approach to predicting S&P500 using sentiment analysis**
Presenter: **Daniel Stamate**, Goldsmiths University of London, United Kingdom
Co-authors: Fionn Murtagh, Rapheal Olaniyan

The frequent ups and downs are characteristic to the stock market. The conventional standard models that assume that investors act rationally have not been able to capture the irregularities in the stock market patterns for years. As a result, behavioural finance is embraced to attempt to correct these model shortcomings by adding some factors to capture sentimental contagion which may be at play in determining the stockmarket. We address the predictive influence of online expressed sentiment on the stock market returns and volatility by using a non-parametric nonlinear approach that corrects specific limitations encountered in previous approaches. A novel approach to developing sentiment analysis and stock market predictive models based on GARCH, EGARCH and recurrent neural network frameworks is presented, and it is compared to previous statistical and/or machine learning approaches addressing this problem, proving its advantages and superiority over the latter. The sentiment information extracted via text mining from online blogs includes variants of indexes expressing relevant sentiment, in particular anxiety, whose predictive value on the dynamic of S&P500 is rigorously analysed using linear and nonlinear Granger causality and Monte Carlo simulations. Future extensions envisage incorporating the necessary apparatus and efficient mechanism to handle also stream data.

EC1673: **Crime forecasting with spatial panel models for counts**
Presenter: **Stephanie Glaser**, University of Hohenheim, Germany
Co-authors: Robert Jung

Crime forecasting has gained increasing interest in the last years but only seldom spatial models are used to exploit the underlying geographic processes in the data for estimation and forecasting. Also, the count nature of the data is often ignored. We apply methods of spatial econometrics to overcome these shortcomings. In particular we propose an additive model with serial and spatial autoregressive terms as well as fixed effects. The model is estimated using system GMM. For forecasting purposes we obtain point forecasts as well as data coherent density forecasts using a Poisson distribution. We apply the model to monthly crime counts at census tract level from Pittsburgh.

EC1432: **Modelling bivariate moving average time series of Poisson counts**
Presenter: **Naushad Mamode Khan**, University of Mauritius, Mauritius

The bivariate integer-valued moving average models based under the Poisson assumption have not gained enough attention in time series literature. The only work undertaken in this direction is the fitting of a bivariate moving average model to two time series of stock transaction series subject to some time-varying covariates. However, the dependence structure between the two series has not been clearly addressed particularly under a dynamic non-stationary covariance set-up. The methods used to estimate the mean and dependence parameters were Generalized Least Squares and Generalized Method of Moments but under discrete assumption, the least squares technique is not appropriate while method of moments do not yield efficient parameter estimates as compared to likelihood-based approaches. We overcome these shortcomings through the use of the copulas techniques to model the non-stationary covariances and the application of the generalized quasi-likelihood approach to estimate the regression, serial correlation coefficients and the dependence parameters. These techniques are assessed through simulation studies and a real-life application based on the Mauritian stock exchange market is provided.

EC1508: **Estimation of dynamic approximate factor models with incomplete data**
Presenter: **Franz Ramsauer**, Technische Universitaet Muenchen, Germany
Co-authors: Aleksey Min, Rudi Zagst, Francesco Sandrini, Lorenzo Portelli, Monica Defend

We consider the estimation of dynamic approximate factor models with homoscedastic, cross-sectionally correlated errors using incomplete data. Besides missing observations, data incompleteness can systematically arise from the inclusion of mixed-frequency information. The dynamics of the latent factors is described by a vector autoregressive process of order p. In contrast to estimation approaches using the Kalman filter or smoother, the presented estimation method comprises two expectation-maximization algorithms which are alternately used. To determine the unknown factor dimension and the unknown autoregressive order p we propose a new two-step model selection criterion. The performance of the estimation procedure and the model selection criterion is investigated in an intensive Monte Carlo study. Finally, we apply the dynamic approximate factor model to real-economy data supporting asset allocation decisions and risk management. For this purpose, an autoregressive model with the estimated factor span of the mixed-frequency data as exogenous variables maps the behavior of the monthly log-returns of the S&P 500 price index. While analyzing the characteristics of the S&P 500 returns, we construct prediction intervals covering the range of the subsequent returns, detect the main drivers for this development and define a dynamic trading strategy to benefit from the gained information.

EG145 Room Gordon CONTRIBUTIONS ON SPATIAL AND SPATIO-TEMPORAL PROCESSES Chair: Brian Reich

EC1476: **From time to space-time: Tempo-spatial Ornstein-Uhlenbeck processes**
Presenter: **Michele Nguyen**, Imperial College London, United Kingdom
Co-authors: Almut Veraart

Ornstein-Uhlenbeck (OU) processes have been frequently used to model stochastic volatility in finance. This describes alternating periods of high and low activity. Such features are also observed in wind velocities. To model spatially varying intermittency, the $\mathrm{OU}_\wedge$ process was introduced as a tempo-spatial extension. It is defined as a random field in space-time and obtained by integrating a Levy basis weighted by an exponentially decaying kernel. The integration region is restricted to be translation invariant and non-anticipative. The $\mathrm{OU}_\wedge$ process shares many properties with the OU process and boosts flexibility in a tractable way. For example, while both have exponential temporal autocorrelation, the spatial autocorrelation of the $mathrmOU_\wedge$ process varies according to its integration set. Simulation and inference for $\mathrm{OU}_\wedge$ processes is novel. Two algorithms have been designed for the canonical process. While the first generates data on the usual rectangular grid, the second uses diamond grids to mimic the edges of the process's triangular integration sets. Detailed studies involving moments-based estimation shows that the latter algorithm generates data with more accurate tempo-spatial dependency structures. The practical relevance of $\mathrm{OU}_\wedge$ processes is further demonstrated using radiation anomaly data from the International Research Institute for Climate and Society.

EC1639: **Spatial autocorrelation: Robustness of measures and tests**
Presenter: **Marie Ernst**, University of Liege, Belgium
Co-authors: Gentiane Haesbroeck

Distinguishing the analysis of spatial data from classical analysis is only meaningful if the spatial components bring information. Therefore, testing if the spatial autocorrelation is significant may confirm or deny the need to consider spatial analysis over the classical one. Spatial autocorrelation expresses the dependence between values at neighbouring locations. Several measures of spatial autocorrelation are defined in the literature. Moran's index, Geary's ratio and Getis-Ord statistic are the most used statistics. Tests based on these measures have been developed in the literature using asymptotic and permutation results. They are used in practice in many fields, for instance in geography, economics, biogeosciences, medicine, ... However, these tests should be cautiously applied because they are not robust. A single contaminated observation can significantly modify their results. The aim is twofold. Firstly, the already available tools for measuring spatial autocorrelation will be reviewed with an emphasis on the study and comparison of their robustness. Secondly, alternative methods will be proposed to robustly estimate the spatial autocorrelation.

EC1773: **Conditional risk assessment in spatial and spatio-temporal settings: Effect of covariates.**
Presenter: **Jose Luis Romero**, University of Granada, Spain
Co-authors: Ana Esther Madrid, Jose Miguel Angulo

Authorities and public or private organizations devoted to studying and analyzing a wide range of real phenomena related to Geosciences, Finance, Climatology, among other fields, have the primary focus of interest on risk assessment of hazard situations, with the aim to provide prevention protocols that minimize damage to the population and material costs. A risk assessment methodology in spatial and spatio-temporal settings, based on a conditional approach, is addressed. Effect of covariates on risk evaluation is analyzed. Relations between widely used risk measures such as value-at-risk and expected shortfall, and structural characteristics of excursion sets of the underlying random field such as exceedance area or excess volume, for varying thresholds, are established. Applications of this methodology in spatial and spatio-temporal frameworks, involving the effect of covariates, are illustrated in fields such as Hydrology and Environmental Health.

EC1779: **Mutual information related measures of multifractal dimensional dependence: Application to seismicity**
Presenter: **Francisco Javier Esquivel**, University of Granada, Spain
Co-authors: Jose Miguel Angulo

Systems with complex behavior, characterized by properties such as scaling and long-range interactions, are present in many fields of science. Diverse statistical tools are applied to their study. In particular, dimensional dependence assessment helps understanding the structural dynamics of such systems. In this context, Tsallis-entropy-based approaches have been successfully used to explain the degree of association between components in systems in disequilibrium. Normalized dependence measures are formulated in terms of different versions of Tsallis mutual information. Extensions to the multifractal domain are derived based on generalized Tsallis dimensions. The introduced measures are used to study the spatio-temporal dynamics in a real seismic series.

EC0187: **Generalized spatial autoregressive conditional heteroscedasticity**
Presenter: **Philipp Otto**, European University Viadrina, Germany
Co-authors: Robert Garthoff
A univariate spatial stochastic process in a $q-$dimensional space is considered. This space could be a subset of the $q-$dimensional real numbers or the $q-$dimensional integers. Regarding the first case a continuous space is present, whereas the process is a spatial lattice process considering the second case. The commonly applied spatial autoregressive model assumes that the variance is constant over all locations in the multidimensional space. This approach is extended by assuming that the variance can vary over space building clusters of high and low variance. Analogue to the ARCH and GARCH model in time series, a spatial model is developed such that the observations of the process are uncorrelated over space, but the squared observations are correlated. The autocorrelation of the second moments can be modeled using matrices of spatial weights. In contrast to the common temporal model, where the distribution at a certain point of time knowing the full information set of the previous period is well defined, the information set of all spatially lagged locations of some location i depends on the observation in location i. Hence, the distribution of the process is not straightforward. Moreover, one could define the residual process of a spatial autoregressive process as spatial ARCH process. Eventually, this model can be applied to explain the variation of the population density.

EC033 Room Senate CONTRIBUTIONS IN ROBUST STATISTICS Chair: Christian Hennig

EC1535: **Robust time series analysis: A comparative simulation study**
Presenter: **Alexander Duerre**, TU Dortmund, Germany
Co-authors: Roland Fried, Tobias Liboschik
Classical time series methods get unreliable in case of outlying observations and inefficient under heavy tailed distributions. Therefore many robust procedures have been developed, which overcome these problems. Unfortunately theoretical properties and also comparative studies are missing in the literature, so that it remains unsure, which procedure one should apply in which situation. We want to fill this gap by comparing the most promising approaches complemented by some new estimators, which have not been used before in the time series context, in a thorough simulation study. We concentrate on the estimation of the autocorrelation structure as well as on fitting linear time series models.

EC1755: **Robust joint modelling: Revealing the impact of outliers**
Presenter: **Lisa McCrink**, Queens University Belfast, United Kingdom
The use of joint modelling techniques has grown in popularity as rapidly as the collection of the data that such models analyse. The appeal of simultaneously analysing an individual's changing repeated measurements over time and the effect such changes have on their time-to-event process is evident, with literature confirming the improvements in estimation compared to independent models. Despite such growth in this research field, only a limited amount of study has investigated the impact of longitudinal outliers on the commonly used normality assumptions of the random terms. The need to investigate this further is highlighted both through a simulation study and an illustrative example exploring the factors that affect the survival of Northern Irish renal patients who are undergoing haemodialysis. The detrimental impact of outliers on both the accuracy and efficiency of estimates is demonstrated, whilst introducing novel methodology for accommodating outliers in one of the most common types of joint models, that which links a linear mixed effects model with a Cox proportional hazards model. Motivated by outlying patients shown to have significantly worse survival than typical renal patients, the focus is on the identification of outliers presented using graphical techniques which are shown to accurately identify outliers. This has the potential to aid clinicians to change the treatment plan of outlying renal patients to improve their survival rates.

EC1696: **Robust and fast add one for linear models**
Presenter: **Mark Hannay**, University of Geneva, Switzerland
We are interested in a fast robust test for add one, while testing the composite hypothesis in the linear model. The Wald, likelihood ratio and score tests of the composite hypothesis have robust equivalents, with already known asymptotic properties. While testing multiple different directions in the add one, both the robust Wald and robust likelihood ratio equivalent require the fit of the corresponding full model, which can be computationally intensive. The robust score test on the other hand is fast, yet if one uses redescenders, the test has undesirable properties under the alternative hypothesis. We therefore propose a new test, the robust residual regression test, which is fast to compute (not requiring a full fit), asymptotically equivalent to the robust Wald test for small deviances and doesn't suffer from the problems the robust score test does. We then proceed to compare all the tests in a simulation study.

EC1694: **Robust heritability and predictive accuracy estimation in plant breeding**
Presenter: **Vanda Lourenco**, Faculty of Sciences and Technology - New University of Lisbon, Portugal
Co-authors: Piepho Hans-Peter, Joseph O. Ogutu
Genomic prediction (GP) is widely used in plant breeding to help find the best genotypes for selection. Two of the most important measures of GP, whose accurate estimation is essential for genomic selection (GS), are predictive accuracy and heritability. Still, as in many other applications, the models of choice for analyzing field data are regression models which are known to underperform when their underlying assumptions are violated. Among other drawbacks, the violation of the normality assumption may lead to biased parameter estimates. These and other biases often translate into inaccurate heritability and predictive accuracy estimates which in turn may negatively impact GS. Since phenotypic data are prone to contamination and almost never perfectly conform to the normality premise, one way of improving GS involves refining its estimation accuracy. Robust statistical methods provide a natural framework, since they are known to overcome some of the handicaps of the likelihood-based classical methodology like departure from normality. Therefore, a robust approach to a recently proposed two-stage method from the literature is presented. The robust technique is compared with its classical counterpart through simulation considering several plausible contamination scenarios. Finally, an example application involving a maize dataset is given to illustrate the adequacy and usefulness of the robust approach.

EC1341: **Robust regression using unnormalized model under heterogeneous contamination**
Presenter: **Hironori Fujisawa**, Institute of Statistical Mathematics, Japan
Co-authors: Takafumi Kanamori
The estimator can be defined as the minimizer of the divergence between the estimated density (e.g. empirically estimated density) and parametric density. It is known that the gamma-divergence is very useful for robust estimation against heavy contamination and can show a sufficiently small latent bias even if the contamination rate is not small. Recently, a new idea was proposed; the estimated density remains unchanged, but the parametric density is replaced by an unnormalized model. We can simultaneously estimate the parameter and the contamination rate. The idea of using the unnormalized model enables us to incorporate the usefulness of the gamma-divergence into other divergences. The idea can be easily applied to the regression model when the contamination is homogeneous. In addition, it can also be applied to the regression model even if the contamination is heterogeneous when the underlying model belongs to a location-scale family.

EG347 Room Jessel CONTRIBUTIONS ON CLUSTERING AND CLASSIFICATION Chair: Hao Zhang

EC1573: **A new approach to the computation of dissimilarity for uncertain data**
Presenter: **Maha Bakoben**, Imperial College London, United Kingdom
Co-authors: Tony Bellotti, Niall Adams
The inspection of uncertainty associated with model parameters is fundamental in statistical inference. When such estimates of uncertainties are available, it is also critical to incorporate them in the dissimilarity computations of the model parameters estimated for a set of data objects. Existing approaches typically construct distance metrics using divergence measures that capture the probabilistic overlap of the data objects. Strong assumptions are typically required to provide tractable computations. In the study, data objects are represented by parameters estimated from some statistical models. We develop a new dissimilarity from a measure of geometric overlap of the joint confidence intervals implied by the covariance matrices associated with parameter estimates. The mathematical computation of overlap in this context is difficult, and we use Monte Carlo procedures to address this issue. The proposed dissimilarity measure was motivated by a clustering problem, and performance results from clustering simulation studies are presented. These demonstrate that this new approach can outperform conventional dissimilarity measures such as Euclidean, Kullback-Leibler and Mahalanobis when measurement error is present.

EC1526: **Group structures of prion diseases via multilevel mixed type data**
Presenter: **Giorgia Rocco**, Sapienza University of Rome, Italy
Co-authors: Luca Tardella, Romolo Nonno, Michele Di Bari, Maria Puopolo
We report on a hybrid approach to analyze a dataset derived from an experimental study on prion diseases. The data comes from inoculating different strains (inocula) of the diseases to bank voles. The aim is to understand at what extent some phenotypic outcomes such as survival times and profiles of brain lesions are able to detect the underlying heterogeneous origin of the data. From a statistical point of view the challenge is the approach through clustering and classification of multilevel mixed type data.

EC1706: **Classification and regression methods for predicting disease risk based on genetic and epidemiological data**
Presenter: **Lamiae Azizi**, University of Sydney, Australia
It is today generally accepted that susceptibility to many diseases is not simply the consequence of a single marker, but occurs as a result of an assembly of both genetic and non-genetic risk factors. Thus, in integrative genomics, a large number of these so-called predictors are often explored for their association with disease risk. In this context, the determination of which combination of predictors are associated with disease risk, and the prediction of the disease status based on this combination, represent major challenges. We study the predictive performance of the classical Logistic Regression and its version combined with both a stepwise selection step and the LASSO, and compare it to other several tree-based decision methods. These tree-based models for regression and classification are geared towards prediction and are commonly used to address the statistical and computational challenges arising from employing a large number of predictors with a high degree of collinearity between them. Using real cancer data-set, we compare the predictive performance of the respective models using various statistical validation criterion and discuss how they can be applied cautiously to predict the risk of a disease in the context of high dimensional genomic and phenotypic data.

EC1652: **Sparse linear discriminant analysis with common principal components**
Presenter: **Tsegay Gebrehiwot Gebru**, The Open University, United Kingdom
Co-authors: Nickolay Trendafilov
Linear discriminant analysis (LDA) is a commonly used method for classifying a new observation into one of $g-$populations. However, in high-dimensional classification problems the classical LDA has poor performance. When the number of variables is much larger than the number of observations, the within-group covariance matrix is singular which leads to unstable results. In addition, the large number of input variables needs considerable reduction which nowadays is addressed by producing sparse discriminant functions. We propose a method to tackle the (low-sample) high-dimensional discrimination problem by using common principal components (CPC). LDA based on CPC is a general approach to the problem because it does not need the assumption of equal covariance matrix in each groups. We find sparse CPCs by modifying a stepwise estimation method. Our aim is to find few important spare discriminant vectors which are easily interpretable. For numerical illustrations, the method is applied on some known real data sets and compared to other methods of sparse LDA.

EC1501: **Mixture model with multiple allocations for clustering spatially correlated observations for the analysis of NGS data**
Presenter: **Saverio Ranciati**, Universita di Bologna, Italy
Co-authors: Cinzia Viroli, Ernst Wit
Model-based clustering is a technique widely used to group a collection of units into mutually exclusive groups. There are, however, situations in which an observation could in principle belong to more than one cluster. In the context of *next-generation sequencing* experiments, for example, the signal observed in the data might be produced by two (or more) different biological processes operating together and a gene could participate in both (or all) of them. We propose a novel approach to cluster NGS discrete data with a mixture model, allowing each unit to belong potentially to more than one group: these multiple allocation clusters can be flexibly defined via a function combining the features of the original groups without introducing new parameters. The formulation naturally produces a 'background group' in which values close to zero can be allocated, acting as a correction for the zero-inflation that arises in this type of data. We take into account the spatial dependency between observations describing it through a latent *conditional auto-regressive* (CAR) model that can reflect different dependency patterns. We assess the performance of our model within a simulation study and then we apply it to a RNA-seq real dataset.

EG169 Room SH349 CONTRIBUTIONS IN STATISTICAL MODELING AND COMPUTATION Chair: Cristian Gatu

EC1578: **EM algorithm for estimating the parameters of stochastic epidemic models with random environment**
Presenter: **Tewfik Kernane**, University of Sciences and Technology Houari Boumediene, Algeria
Co-authors: Sidali Becheket, Hamid Elmaroufy
Epidemic models are used to understand the evolution of infectious diseases such as predicting the effects of public health programs. The SIS (Susceptible-Infected-Susceptible) model is often used to model diseases for which there is no immunity, including for example gonorrhea and pneumococcus. Meanwhile, the SIR (Susceptible-Infected-Removed) model is used for epidemics where after a random time of infection the infected individual is removed. We obtain diffusion approximation of the stochastic SIS and SIR epidemic models in random environment. Simulation is often used to predict the spreading of an epidemic. We provide exact Gillespie algorithms for simulating these two models as well as the Euler-Maruyama scheme for simulating from the diffusion approximations. The simulation depend on the estimation of the parameters governing the dynamics of such models. We propose to use the Expectation-Maximisation (EM) algorithm for the estimation of the parameters and the transition probabilities of the unobserved random environment. Simulation examples are performed to illustrate the effectiveness of the EM algorithm.

EC1728: **Additive modelling for scaled t data with residual autocorrelation**
Presenter: **Natalya Pya**, Nazarbayev University, Kazakhstan

Generalized additive models (GAM) are widely used in environmental science and industry for modelling flexible nonlinear relationship among multiple variables. The extensive use of GAMs is explained by their flexibility and the existence of reliable and efficient computational methods for their estimation. The standard GAM assumes that the response variables are independent random variables from the exponential family of distributions. Many generalizations of GAMs have been proposed, such as modelling outside the exponential family setting, modelling multivariate response, modelling beyond the mean. Despite the availability of the generalizations, the current methodological framework for these models could be improved on adequate modelling of practical data sets. Additive modelling of data sets that are heavier tailed than normal distribution and also reveal short range temporal autocorrelation is proposed. The suggested approach is merged with the general GAM framework which is implemented in R.

EC1267: **prop.comb.RR: R package for linear combination of proportions**
Presenter: **Maria Alvarez Hernandez**, University of Vigo, Spain
Co-authors: Javier Roca Pardinas

Inferences about a linear combination of several independent binomial proportions have aroused a great amount of interest, especially in applied research such as multicenter clinical trials, public-health surveys or epidemiology studies. The prop.comb.RR package is a software application for R to make inferences about a general linear combination of proportions and for classic cases as a proportion, the difference of proportions or the relative risk (particular instances of a linear combination). Inferences are solved not only from the perspective of the confidence interval, but also from the perspective of hypothesis tests. The implemented methods are optimal procedures selected in published literature and the well-known score method.

EC1538: **CatPredi: An R package to categorise continuous variables in a logistic regression model**
Presenter: **Irantzu Barrio**, Universidad del Pais Vasco, Spain
Co-authors: Inmaculada Arostegui, Maria Xose Rodriguez-Alvarez

In the medical field, prediction models are gaining importance as a support for decision-making. Although it is not recommended from a statistical point of view due to loss of information and power, the categorisation of continuous variables is a common practice in the development of clinical prediction models. However, there are no unified criteria for the selection of the cut points in the categorisation process. In order to provide valid cut points whenever a categorisation is going to be performed, we have developed a methodology for a logistic regression model based on the maximisation of the AUC. This methodology has been implemented in an R package called CatPredi. This is a package of R functions that allows the user to categorise a continuous predictor variable in a univariate or multiple logistic regression model. It provides the optimal location of cut points for a chosen number of cut points, fits the prediction model with the categorised predictor variable and returns the estimated and bias-corrected discriminative ability index for this model. Additionally, it allows a comparison of two categorisation proposals for different number of cut points and the selection of the optimal number of cut points.

EC0276: **Robust inference in regression models**
Presenter: **Gay Marie Saavedra**, University of the Philippines - Diliman, Philippines

The aim is to propose a hybrid method of forward-search and bootstrap approaches to estimate a model with contaminated data or a structural change. The method is compared to ordinary least squares, simple bootstrap, and simple forward-search. A simulation study illustrates that the new algorithm yields narrower confidence intervals, even when there is model misspecification and contamination. Moreover, it has a comparable predictive ability to simple forward search and yields comparable estimates of the regression coefficients with lower percent bias. Thus, the hybrid of forward-search and bootstrap method can produce better estimates regression coefficients in the presence of structural changes and/or data contamination.

EG097 Room Chancellor's Hall CONTRIBUTIONS ON DEPENDENCE MODELS AND COPULAS Chair: Ostap Okhrin

EC1592: **Structure learning of sparse R-vines using directed acyclic graphs**
Presenter: **Dominik Mueller**, Technische Universitaet Muenchen, Germany
Co-authors: Claudia Czado

Modeling dependencies in high dimensional systems has become an increasingly important topic in the recent research. While most approaches rely on restrictive distributional assumptions, vine copulas allow to choose from a wide array of parametric copula families. This freedom comes however with a quadratically increasing modeling complexity, and exponential freedom in choosing a structure describing the corresponding (conditional) bivariate dependencies in the dataset. Graphical models using directed acyclic graphs (DAGs) were used frequently to identify conditional independencies in datasets. These models may now be exploited to yield sparse representations of joint dependence models. We develop heuristics which might outperform the current structure estimation for vine copulas using directed acyclic graphs and yield more parsimonious representations of the joint dependencies in some datasets.

EC1606: **Simplified vine copula models: Approximations based on the simplifying assumption**
Presenter: **Fabian Spanhel**, Ludwig-Maximilians-University Munich, Germany
Co-authors: Malte Kurz

In the last decade, simplified vine copula models (SVCMs), or pair-copula constructions, have become an important tool in high-dimensional dependence modeling. So far, specification and estimation of SVCMs has been conducted under the simplifying assumption, i.e., all bivariate conditional copulas of the data generating vine are assumed to be bivariate unconditional copulas. For the first time, we consider the case if the simplifying assumption does not hold for the data generating process and an SVCM acts as an approximation of a multivariate copula. Several results concerning optimal simplified vine copula approximations (SVCAs) and their properties are established. We show that step-by-step estimators of pair-copula constructions converge to the partial vine copula approximation (PVCA) if the simplifying assumption does not hold. The PVCA can be regarded as a generalization of the partial correlation matrix where partial correlations are replaced by $j-$th order partial copulas. We prove that the PVCA does not globally minimize the KL divergence from the true copula and show that the best approximation satisfying the simplifying assumption is given by a vine pseudo-copula. Moreover, we demonstrate how spurious conditional (in)dependencies may arise in SVCAs. Finally, the practical implications of the theoretical results for vine copula modeling are investigated.

EC1632: **The empirical beta copula**
Presenter: **Hideatsu Tsukahara**, Seijo University, Japan
Co-authors: Johan Segers, Masaaki Sibuya, Nathan Uyttendaele

Given a sample from a multivariate distribution F, the uniform random variates generated independently and rearranged in the order specified by the vector of ranks look like a sample from the copula of F. This idea can be regarded as Baker's construction of copulas based on order statistics with the ranks being coefficients, and led us to define the empirical beta copula. It is then fairly easy to show that the empirical beta copula is a particular case of the empirical Bernstein copula. The advantage is that we do not need any smoothing parameter. Also it is extremely simple to

simulate a sample from the empirical beta copula. We establish the assumptions under which the standard asymptotic results hold for the empirical Bernstein copula. They are significantly weaker than those given in the literature. Our Monte Carlo simulation study shows that in all cases, the empirical beta copula outperforms the empirical copula in terms of the bias and the integrated mean squared error. Compared with the empirical Bernstein copula with optimal smoothing rate, its performance is still significantly better in several cases, especially in terms of bias.

EC1724: **Closure property of randomly weighted sums with long-tailed increments**
Presenter: **Lina Dindiene**, Vilnius University, Lithuania
Co-authors: Remigijus Leipus
We deal with the tail behavior of randomly weighted sums. We assume that initial random variables $X_1,\dots,X_n$ are dependent with heavy-tailed distribution functions $F_1,\dots,F_n$, respectively. Random weights $\Theta_1,\dots,\Theta_n$ are independent of $X_1,\dots,X_n$. Under some dependence structure between $X_1,\dots,X_n$ we show that the closure property of the sum $S_n^{\Theta} := \Theta_1 X_1 + \cdots + \Theta_n X_n$ holds, i.e., given that distributions $F_1,\dots,F_n$ are long-tailed, the distribution function of sum S_n^{Θ} belongs to the same class. Some copula-based examples illustrate the results.

EC1659: **Estimating copula density using ranks and sub-sampling**
Presenter: **Jerome Collet**, Electricite de France RD Division, France
The goal is to estimate the copula density of a d-dimensional random variable, without parametric assumptions. Given a sample of size n, for a given sub-sample size $m < n$, one draws many sub-samples. For each sub-sample and each observation, one obtains a vector of ranks (in set $\{1,\cdots,m\}^d$). For each point $\mathbf{r}$ of $\{1,\cdots,m\}^d$, we propose to count the sub-samples from which one obtains $\mathbf{r}$ (where there is an observation which d ranks are the point $\mathbf{r}$). This counting is the estimator we propose, it converges in some sense to the copula density. The sub-sample size is equivalent to a bandwidth, so its choice is a trade-off between bias and variance. We study the convergence speed in two cases: the independent one and a case encompassing additive models. We provide theoretical results, we confirm it using simulation studies and comparisons with usual methods.

EG009 Room Montague CONTRIBUTIONS ON STATISTICS FOR FUNCTIONAL DATA Chair: Jane-Ling Wang

EC1614: **Notes on Schutte and Takashi functional inequalities and their applications in functional data analysis**
Presenter: **Ewa Szlachtowska**, AGH University of Science and Technology, Poland
Co-authors: Daniel Kosiorowski, Dominik Mielczarek, Jerzy Rydlewski
Prediction errors are frequently used to select and to assess the models under study. The magnitude of estimation error shows the quality of prediction. Examination of prediction errors has the advantage to confirm the correctness of the finally chosen model. We will present the basic concepts of functional data analysis. We extrapolate classical notions of probability, such as the expected value of a random variable, the distribution of the random variable or variance of functional data. Then we extrapolate these concepts to $p-$power integrable random variables over a bounded and Lebesgue measurable set. Next we prove Chebyshev type inequalities for random variables function. To prove these inequalities we will use the fact that L_p space can readily be expressed as a space with semi-inner product. In addition, we will use Schutte and Takashi functional inequalities. In the last part we assume that a random variable X is $p-$power integrable. Then we show that the estimation error of functional data does not exceed the value $o(1)/nlog(n)$, where n is the number of estimated parameters. This inequality reflects the difference between the random variable X and its estimator. As a result the inequality will be used to estimate the convergence of functional prediction in terms of the L_p norm. To prove the convergence of prediction we use Schutte distribution for Banach spaces and a version of Takashi inequality in Banach spaces.

EC1690: **Inference for sparse to non-sparse functional data with covariate-adjustments**
Presenter: **Dominik Liebl**, University Bonn, Germany
The motivation arises from the problem of analyzing the nonlinear price effects of Germany's abrupt nuclear phaseout after the nuclear disaster in Fukushima Daiichi, Japan, in mid-March 2011. The technical side deals with the nonparametric local linear estimation of the mean and covariance function from a latent functional time series of random functions with covariate-adjustments. By contrast to the classical case of sparse functional data, the amount of discretization points per function is allowed to diverge with the number of functions in our asymptotic analysis. This broader asymptotic scenario includes all cases from (very) sparsely to (very) densely sampled discretization points per function and therefore allows to take into account the somewhat intermediate case in our application. The derived bias, variance, and bandwidth results are used to test for differences in the electricity prices before and after Germany's nuclear phaseout for given values of the most important exogenous factors: electricity demand and air temperature.

EC1783: **Consistency of $h-$mode depth**
Presenter: **Stanislav Nagy**, KU Leuven, Belgium
Consistency results for the sample $h-$mode depth in the general case of Banach-valued data are established. The rate of convergence is provided, which is linked to the rate at which the sample sequence of bandwidths converges to its population version. The robustness of the $h-$mode depth, as well as the convergence of the associated modes of the distribution, are also studied.

EC1628: **Hilbertian ARMA model for forecasting functional time series**
Presenter: **Jose Portela**, Universidad Pontificia Comillas, Spain
Co-authors: Antonio Munoz, Estrella Alonso
A new forecasting method for functional time series is proposed. This model attempts to generalize the standard scalar ARMA time series model to the L^2 Hilbert space in order to forecast functional time series. A functional time series is the realization of a stochastic process where each observation is a continuous function defined in a finite interval $[a,b]$. Forecasting these time series require a model that can operate with continuous functions. The structure of the proposed model is a regression where functional parameters operate on functional variables. The variables can be lagged values of the series (autoregressive terms), past observed errors (moving average terms) or exogenous variables. The functional parameters used are integral operators in the L^2 space. In our approach, the kernels of the operators are given as a linear combination of sigmoid functions. The parameters of each sigmoid are estimated using a Quasi-Newton algorithm minimizing the sum of squared errors. This is a novel approach because the iterative algorithm allows estimating the moving average terms. The new model is tested with functional time series obtained from real data of the Spanish electricity market and compared with other functional reference models.

EC1766: **Model-based functional clustering to detect climate changes**
Presenter: **Per Arnqvist**, Umea University, Sweden
Co-authors: Sara Sjostedt de Luna
A model-based method to cluster functions into k latent homogeneous groups is proposed. The functions are represented by linear combinations of known basis functions, e.g. B-splines. The (spline) coefficients are assumed to be normally distributed allowing for different mean and covariance structures for the underlying k groups. Extending previous results, we derive/propose an EM algorithm to estimate the parameters of the model and the posterior probabilities used to assign group labels to each function. The model can handle unequally spaced and also different number of observations of the sampled functions. The method also opens up for incorporating covariates in the model. Information criteria, such as AIC, is suggested to determine the number of latent groups.The method is illustrated by analyzing a varved lake sediment core from the lake Kassjon

(N. Sweden). The sediment record consists of around 6400 varves, each varve having a functional seasonal pattern. Image analysis was used to generate the observed data of yearly profiles (in terms of grey-scale variation) and the varying thickness was measured as the number of pixels within a varve (year).

EG325 Room Bedford CONTRIBUTIONS IN TIME SERIES ANALYSIS II Chair: Sara Taskinen

EC1682: **Predictive model choice in time series analysis under stationarity and non-stationarity**
Presenter: **Tobias Kley**, London School of Economics and Political Science, United Kingdom
Co-authors: Piotr Fryzlewicz, Philip Preuss
In statistical research there usually exists a choice between structurally simpler or more complex models. We argue that, even if a more complex model were true, a simple one may be advantageous to work with under parameter uncertainty. We present an alternative model choice methodology for time series analysis, where one of two competing approaches is chosen based on its empirical finite-sample performance with respect to a certain task. A rigorous, theoretical analysis of the procedure is provided in the framework of choice between stationarity and local stationarity when the task is to forecast. We state conditions that imply when it is preferable to base the forecasts on the more volatile time-varying estimates and when it is advantageous to forecast as if the data were from a stationary process, even though it is not. We also consider different frameworks, as for example choosing between linear and non-linear time series models, and provide the results of an extensive simulation study as well as an empirical example.

EC1405: **Application of the generalized resampling scheme to spectral density estimation of APC time series**
Presenter: **Lukasz Lenart**, Cracow University of Economics, Poland
Using Generalized Resampling Scheme the consistent estimator (in the mean square sense) of the spectral density function for Almost Periodically Correlated (APC) time series was introduced. The non-zero mean case was considered. The class of APC time series is a generalization of PC (Periodically Correlated) time series. The problem of spectral density estimation for PC time series with known period T was considered in the literature. In general APC case we assume that the mean function and the autocovariance functions are almost periodic functions with unknown set of frequencies in Fourier representation. Therefore, the same technique as in PC case cannot be applied. To obtain the consistent estimator the Generalized Resampling Scheme and the asymptotic distribution for normalized estimator of Fourier coefficient in mean representation were used. The applications for business cycle analysis were considered.

EC1575: **Hyvärinen scoring rule estimators for a single linear time series**
Presenter: **Silvia Columbu**, University of Cagliari, Italy
Co-authors: Valentina Mameli, Monica Musio
Composite likelihood methods appears to be appealing alternative to the full likelihood when the latter is to complex to evaluate explicitly. Even if they have traditionally been considered as an approximation to the full likelihood, they can be justified in its own right, as leading to an unbiased estimating equation. These methods are particularly attractive since generally inherit many of the good properties of inference based on the full likelihood function. However, recently it has been proved that in the first order moving average model, the pairwise likelihood method, which is a special case of composite likelihood, has very poor asymptotic efficiency. We propose the use of proper scoring rules, in particular of the Hyvärinen score as an alternative to the full or composite likelihood in situations in which the latter shows poor performances. We consider a traditional time series setting where only a single realization of a time series of increasing length is available. Under this framework, asymptotic properties of the so-called minimum Hyvärinen score estimator will be explored and investigated.

EC1502: **An evolution strategy for ARMA model selection and estimation**
Presenter: **Basad Al-sarray**, Franche comte unieversity, France
An evolution strategy is a soft computing algorithm branch of artificial intelligent field, consists of three principal steps, Mutation, Recombination and Selection. The idea is to apply variants of the ES algorithm to ARMA model selection and estimation based on minimum prediction error, by applying the VC dimension lower bound of convergence rate on level sets of the fitness function. Theoretical, simulation studies and applications on finance data will be presented.

EG159 Room Torrington CONTRIBUTIONS ON QUANTILE REGRESSION Chair: James Taylor

EC0336: **Joint modelling of quantile regression and survival time of lung function decline in cystic fibrosis patients**
Presenter: **Elisabeth Waldmann**, Friedrich-Alexander-Universitaet Erlangen-Nuernberg, Germany
Co-authors: David Taylor-Robinson
When modelling repeated measurement and time to event data simultaneously, analysis is often based on combining mixed models with survival analysis. Some data sets however have a more complex structure than the one underlying normality assumption. In the case of modelling lung function of cystic fibrosis (CF) patients over a long period of time the impact of the covariates differs between quantiles of the dependent variable. An important question in CF research is to understand the impact of the lung function trajectory on the onset of a variety of different infections. The aim is to explain the setup of a Bayesian joint quantile model based on replacing the Gaussian distribution in the longitudinal model with an asymmetric Laplace distribution and illustrates it using data from the United Kingdom CF registry.

EC0863: **Robust quantile regression estimation and variable selection**
Presenter: **Pavel Cizek**, Tilburg University, Netherlands
Penalized quantile regression (QR) that incorporates variable selection in high-dimensional models has been extensively studied in the past decade. Given its sensitivity to heavy-tailed distributions and to outliers and leverage points, the re- or down-weighting of penalization term and distant observations were proposed. Concentrating on the latter issue, the sensitivity of penalized QR to outliers and leverage points, it can be shown that the existing approaches to robust penalized QR are not globally robust and the main existing method proved to withstand a large fraction of outliers and leverage points is specific to penalized least squares regression. Moreover in many cases of outlier-robust penalized regression, the asymptotic behavior is not established. Hence to facilitate outlier-robust penalized QR estimation, a locally weighted penalized QR is proposed. The local weighting, which relies on an initial not-necessarily-consistent estimator of the model parameters, guarantees the robustness of the method. At the same time, the local weighting is designed in such a way that the asymptotic distribution of the estimator is well defined and conditionally equivalent to standard penalized quantile regression. We derive the asymptotic distribution and robust properties of the proposed method and study its finite-sample behavior by means of Monte Carlo simulations.

EC0688: **Predicting the expected shortfall corresponding to value at risk forecasts produced by quantile models**
Presenter: **James Taylor**, University of Oxford, United Kingdom
Value at risk (VaR) is widely used as a measure of financial market risk. One approach to VaR forecasting is to model directly the conditional quantile for a chosen probability level. This is the approach taken with conditional autoregressive VaR models, which are estimated using quantile regression. Directly modelling a quantile avoids the need for a distributional assumption, and allows the dynamics of a quantile to differ for

each probability level. However, by focussing on a particular quantile, such approaches provide no information regarding the magnitude of the exceedances beyond the quantile. The conditional expectation of these exceedances is the expected shortfall (ES), which is a measure of market risk that is receiving increasing attention. We introduce a method for predicting the ES corresponding to the VaR forecasts produced by quantile models. The method enables the joint modelling of the VaR and ES, with parameters estimated using maximum likelihood based on an asymmetric Laplace density. This estimation approach has been shown to be equivalent to quantile regression. We present an empirical study using daily stock indices.

EC1551: **Model selection in expectile regression**
Presenter: **Elmar Spiegel**, University of Goettingen, Germany
Co-authors: Thomas Kneib, Fabian Sobotka

Expectile regression can be seen as a mixture of quantile regression and normal linear regression, since an expectile e_τ is the solution of the least asymmetric weighted squared error function $\sum_i w_\tau(y_i)(y_i - e_\tau)^2$, with data y_i, asymmetry parameter $\tau \in (0,1)$ and weighting function $w_\tau(y_i) = (1-\tau)I(y_i < e_\tau) + \tau I(y_i \geq e_\tau)$. So expectile regression estimates the influence of covariates on the whole distribution of the response, but does not assume a specific distribution function. Since expectile regression depends on the L_2 norm, it is a generalization of the normal linear regression and inherits its advantages of easily including splines and spatial data. In general, model selection gains greater influence with more complex data and increasing complexity of the models. For expectile regression, model selection is especially interesting, since there are two different perspectives to be considered. On the one hand the best model for the current asymmetry parameter identifies which covariate has relevant influence on this special part of the distribution. On the other side the optimal model to describe the whole distribution is advantageous when comparing the effects of different asymmetries. For both approaches criteria based and shrinkage methods are possible. Finally a simulation study and an application on undernourished children illustrate the approaches.

EC1420: **D-vine copula based quantile regression**
Presenter: **Daniel Kraus**, Technische Universitaet Muenchen, Germany
Co-authors: Claudia Czado

Quantile regression, that is, the prediction of a random variable's quantiles conditioned on other random variables taking on certain values, has perpetually gained importance in statistical modeling and financial applications. We introduce a new quantile regression method based on sequentially fitting a likelihood optimal D-vine copula to given data resulting in highly flexible models with easily extractable conditional quantiles. As a subclass of regular vine copulas, D-vines enable the modeling of multivariate copulas in terms of bivariate building blocks, a so-called pair-copula construction (PCC). The proposed algorithm works fast and accurate even in high dimensions and incorporates an automatic variable selection. In a simulation study the improved accuracy and saved computational time of the approach in comparison with established quantile regression methods is highlighted. An extensive financial application to international credit default swap (CDS) data including stress testing and Value at Risk (VaR) prediction demonstrates the usefulness of the proposed method.

EG233 Room Athlone CONTRIBUTIONS ON CONTROL AND RELIABILITY Chair: Sophie Mercier

EC1777: **Hazard rate and reliability function predictions for future observations**
Presenter: **Baris Surucu**, Orta Dogu Teknik Universitesi, Turkey

Estimating reliability and hazard rate functions for various types of distributions based on sample information is an interesting problem in the statistical literature. The common and the simplest approach to this problem is to estimate the unknown parameters for the underlying distribution of concern and replace these unknown parameters in reliability and hazard rate functions by their estimators. However, this may lead to inconsistent coverage probabilities for the estimated functions. This classical approach may even perform much worse if one is interested in hazard rate and reliability function predictions for future observations. We will firstly introduce an efficient approximation algorithm for hazard rate and reliability functions, which achieves highly accurate coverage probabilities for their confidence intervals. We will also show that the convergence to asymptotic distributions for the functional estimators is quite fast. Secondly, we will discuss how predicted future order statistics of a specified distribution are obtained and utilized in this context. Results of a simulation study will also be shown for various types of location-scale distributions to demonstrate the efficiency of the proposed method for future observations.

EC1674: **Weighted likelihood-based inference for the stress strength reliability parameter**
Presenter: **Luca Greco**, University of Sannio - Benevento, Italy

The problem of evaluating the stress-strength reliability parameter is considered when both stress and strength data are prone to contamination. It is well-known that standard likelihood-based inference can be badly affected by mild departures from model assumptions, that often occur in the form of several outliers. The proposed robust method relies on the weighted likelihood methodology. This approach has the advantage to lead to robust estimators, tests and confidence intervals that share the main asymptotic properties of their classical counterparts based on the likelihood function.

EC1758: **Reliability calculus of a Markov renewal process on a crack propagation problem**
Presenter: **Chrysanthi Papamichail**, University of Technology of Compiegne, France
Co-authors: Salim Bouzebda, Nikolaos Limnios

The aim is to study a stochastic differential system that describes the evolution of a degradation mechanism, the fatigue crack propagation. A Markov process is the perturbing process of the system that models the crack evolution. With the help of Markov renewal theory, we study the reliability of a structure and propose for it a new analytical solution. The method we propose reduces the complexity of the reliability calculus compared with the previous resolution method. As numerical applications, we tested our method on a numerical example and on an experimental data set, which gave results in good agreement with a Monte Carlo estimation.

EC031 Room Court CONTRIBUTIONS ON BAYESIAN METHODS Chair: Richard Gerlach

EC1604: **On a systematic fanning out of income profiles**
Presenter: **Sarah Meyer**, University of Muenster, Germany

The nature of labour income risk is important for various economic decisions. Two different approaches to modelling labour income profiles have been established in the literature: the heterogeneous income profile model (HIP) and the restricted income profile model (RIP). Both models mainly differ in their assumptions about whether differences in labour income profiles are driven deterministically or stochastically. Our aim is to empirically investigate which of the two models is more suitable to describe real income data. To this end, a dynamic linear model is proposed. It allows for both individual-specific and time-varying coefficients. Estimation of the latent state vector and of the unknown variance parameters is carried out using Gibbs sampling, a Markov chain Monte Carlo (MCMC) algorithm. In our application, the Forward Filtering Backward Sampling (FFBS) algorithm is used as a building block, while further draws result from the conjugate prior distributions. Applying our framework to the German SOEP (GSOEP) data, we find evidence that the earnings data disprove the RIP approach. For validation we employ the Bayes factor, which confirms our finding.

EC1714: **Sequential MCMC with multiple adaptive proposals**
Presenter: **Leopoldo Catania**, Tor Vergata University of Rome, Italy
Co-authors: Mauro Bernardi
The increased amount of large dimensional data calls for new algorithms for sequentially extracting signals from data and estimating model parameters. We propose a new Sequential MCMC algorithm that continuously updates the proposal parameters as the new information arrives. The key innovation of the proposed methodology relies on the mixture of transition kernels of the MCMC algorithm. The validity of the method is tested on large dimensional finite mixture and hidden Markov models as well as on dynamic models with time varying parameters.

EC1757: **Bayesian hierarchical models for financial networks**
Presenter: **Axel Gandy**, Imperial College London, United Kingdom
Co-authors: Luitgard Veraart
We discuss how Bayesian hierarchical models can be used to describe financial networks and how these models can be fitted. We particularly consider the situation when the network is not fully observed.

EC1397: **Some new approaches for performing efficient and reliable Bayesian inference on large datasets**
Presenter: **Clare McGrory**, University of Queensland, Australia
Some new computational approaches for performing time efficient, practical and reliable Bayesian inference on large datasets will be presented. Massive datasets are becoming commonplace in modern statistical applications. While the ability to capture more information than we could before is exciting and enhances the potential to unlock even more interesting features of the data, the downside is that data storage as well as analysis can be challenging. We consider some important applications where these difficulties arise. A time-efficient approach for satellite image analysis will be described. The approach centres on the idea of carefully choosing a representative weighted subsample of the complete dataset in order to model the images in a fraction of the time. Hybrid algorithms combining more than one Bayesian inferential technique to create an approach which targets the posterior distribution more efficiently thereby saving on time while aiming to improve accuracy will also be explored. We will outline hybrid schemes for mixture model estimation and hidden Markov modelling, considering applications to genetics problems and time series data arising from ocean regime shift modelling.

EC0455: **Bayesian detection of abnormal segments in multiple time series**
Presenter: **Lawrence Bardwell**, Lancaster, United Kingdom
Co-authors: Paul Fearnhead
A novel Bayesian approach is presented to analysing multiple time-series with the aim of detecting abnormal regions. These are regions where the properties of the data change from some normal or baseline behaviour. We allow for the possibility that such changes will only be present in a, potentially small, subset of the time series. We develop a general model for this problem, and show how it is possible to accurately and efficiently perform Bayesian inference, based upon recursions that enable independent sampling from the posterior distribution. A motivating application for this problem comes from detecting copy number variation (CNVs), using data from multiple individuals. Pooling information across individuals can increase the power of detecting CNVs, but often a specific CNV will only be present in a small subset of the individuals. We evaluate the Bayesian method on both simulated and real CNV data, and give evidence that this approach is more accurate than a recently proposed method for analysing such data. We also prove some properties of our method.

Authors Index