PROGRAMME AND ABSTRACTS

7th CSDA International Conference on

Computational and Financial Econometrics (CFE 2013)

http://www.cfenetwork.org/CFE2013

and

6th International Conference of the
ERCIM (European Research Consortium for Informatics and Mathematics) Working Group on

Computational and Methodological Statistics (ERCIM 2013)

http://www.cmstatistics.org/ERCIM2013

Senate House, University of London, UK
14-16 December 2013

ERCIM
European Research Consortium
for Informatics and Mathematics

Queen Mary University of London
http://www.qmul.ac.uk

Birkbeck UNIVERSITY OF LONDON
http://www.bbk.ac.uk

LSE THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE
http://www.lse.ac.uk

Dear Friends and Colleagues,

We warmly welcome you to London, for the Seventh International Conference on *Computational and Financial Econometrics* (CFE 2013) and the Sixth International Conference of the ERCIM Working Group on *Computational and Methodological Statistics* (ERCIM 2013). As many of you know, this annual conference has been established as a leading joint international meeting for the interface of computing, empirical finance, econometrics and statistics.

The conference aims at bringing together researchers and practitioners to discuss recent developments in computational methods for economics, finance, and statistics. The CFE-ERCIM 2013 programme consists of 270 sessions, 5 plenary talks and over 1100 presentations. There are over 1200 participants. The founder editor of the journal Computational Statistics & Data Analysis (CSDA) Stanley Azen and co-editor Jae Chang Lee are being honoured during the conference dinner for their dedication and editorial work.

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 the 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, Birkbeck University of London, London School of Economics, Liberbank, Elsevier and ERCIM.

Looking forward, the CFE-ERCIM 2014 will be held at the University of Pisa, Pisa (Italy) on 6-8 December 2014. You are invited and encouraged to actively participate in these events.

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

The CFE-ERCIM 2013 co-chairs and the International Organizing Committee.

ERCIM Working Group on
COMPUTATIONAL AND METHODOLOGICAL STATISTICS

http://www.cmstatistics.org

AIMS AND SCOPE

The working group (WG) CMStatistics focuses on all computational and methodological aspects of statistics. Of particular interest is research in important statistical applications areas where both computational and/or methodological aspects have a major impact. The aim is threefold: first, to consolidate the research in computational and methodological statistics that is scattered throughout Europe; second to provide researches with a network from which they can obtain an unrivalled sources of information about the most recent developments in computational and methodological statistics as well as its applications; third to edit quality publications of high impact and significance in the broad interface of computing, methodological statistics and its applications.

The scope of the WG is broad enough to include members in all areas of methododological statistics and those of computing that have an impact on statistical techniques. Applications of statistics in diverse disciplines are strongly represented. These areas include economics, medicine, epidemiology, biology, finance, physics, chemistry, climatology and communication. The range of topics addressed and the depth of coverage establish the WG as an essential research network in the interdisciplinary area of advanced computational and methodological statistics.

The WG 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 (currently for the Journal of Computational Statistics & Data Analysis).

Specialized teams

Currently the ERCIM WG has over 800 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 Metholological Statistics WG CMStatistics

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 web site), or by email at info@cmstatistics.org.

CFEnetwork
COMPUTATIONAL AND FINANCIAL ECONOMETRICS

http://www.CFEnetwork.org

AIMS AND SCOPE

Computational and Financial Econometrics (CFEnetwork) is an autonomous subgroup linked to CMStatistics. This network will enhance the interface of theoretical and applied econometrics, financial econometrics and computation in order to advance a powerful interdisciplinary research field with immediate applications. The aim is first, to promote research in computational and financial econometrics; second to provide researches with a network from which they can obtain an unrivalled sources of information about the most recent developments in computational and financial econometrics as well as its applications; third to edit quality publications of high impact and significance.

Computational and financial econometrics comprise a broad field that has clearly interested a wide variety of researchers in economics, finance, statistics, mathematics and computing. Examples include estimation and inference on econometric models, model selection, panel data, measurement error, Bayesian methods, time series analyses, portfolio allocation, option pricing, quantitative risk management, systemic risk and market microstructure, to name but a few. While such studies are often theoretical, they can also have a strong empirical element and often have a significant computational aspect dealing with issues like high-dimensionality and large number of observations. Algorithmic developments are also of interest since existing algorithms often do not utilize the best computational techniques for efficiency, stability, or conditioning. So also are developments of environments for conducting econometrics, which are inherently computer based. Integrated econometrics packages have grown well over the years, but still have much room for development.

The CFEnetwork comprises a number of specialized teams in various research areas of computational and financial econometrics. The teams contribute to the activities of the network by organizing sessions, tracks and tutorials during the annual CFEnetwork meetings, editing special issues (currently publish under the CSDA Annals of CFE) and submitting research proposals.

Specialized teams

Currently the CFEnetwork has approximately 500 members and the following specialized teams

AE: Applied Econometrics

BE: Bayesian Econometrics

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 contact the Chairs of the specialized groups (see the web site), or by email at info@cfenetwork.org.

SCHEDULE

CFE 2013

Saturday, 14th December 2013

08:25 - 08:35	Opening CFE
08:35 - 09:25	Plenary Session A
09:35 - 10:50	Parallel Sessions B
10:50 - 11:20	Coffee Break
11:20 - 13:00	Parallel Sessions D
13:00 - 14:30	Lunch Break
14:30 - 16:10	Parallel Sessions E
16:10 - 16:40	Coffee Break
16:40 - 18:45	Parallel Sessions F
20:00 - 21:30	Reception

Sunday, 15th December 2013

09:35 - 10:25	Plenary Session H
10:25 - 10:55	Coffee Break
10:55 - 12:35	Parallel Sessions I
12:35 - 14:40	Lunch Break
14:40 - 16:20	Parallel Sessions L
16:20 - 16:50	Coffee Break
16:50 - 18:30	Parallel Sessions M
20:30 - 24:00	Christmas Conference Dinner

Monday, 16th December 2013

08:45 - 10:25	Parallel Sessions O
10:25 - 10:55	Coffee Break
10:55 - 13:00	Parallel Sessions P
13:00 - 14:30	Lunch Break
14:30 - 15:50	Parallel Sessions Q
15:50 - 16:20	Coffee Break
16:20 - 17:40	Parallel Sessions R
17:50 - 18:40	Plenary Session S
18:40 - 18:45	Closing

ERCIM 2013

Saturday, 14th December 2013

09:50 - 10:00	Opening ERCIM
10:00 - 10:50	Plenary Session C
10:50 - 11:20	Coffee Break
11:20 - 13:00	Parallel Sessions D
13:00 - 14:30	Lunch Break
14:30 - 16:10	Parallel Sessions E
16:10 - 16:40	Coffee Break
16:40 - 18:45	Parallel Sessions F
20:00 - 21:30	Reception

Sunday, 15th December 2013

08:45 - 10:25	Parallel Sessions G
10:25 - 10:55	Coffee Break
10:55 - 12:10	Parallel Sessions J
12:20 - 13:10	Plenary Session K
13:10 - 14:40	Lunch Break
14:40 - 16:20	Parallel Sessions L
16:20 - 16:50	Coffee Break
16:50 - 18:55	Parallel Sessions N
20:30 - 24:00	Christmas Conference Dinner

Monday, 16th December 2013

08:45 - 10:25	Parallel Sessions O
10:25 - 10:55	Coffee Break
10:55 - 13:00	Parallel Sessions P
13:00 - 14:30	Lunch Break
14:30 - 15:50	Parallel Sessions Q
15:50 - 16:20	Coffee Break
16:20 - 17:40	Parallel Sessions R
17:50 - 18:40	Plenary Session S
18:40 - 18:45	Closing

TUTORIALS, MEETINGS AND SOCIAL EVENTS

TUTORIALS

The tutorials will take place on Friday the 13th of December 2013 at Clore Management Center, Birkbeck University of London. The first one is given by Alastair Young (Achieving accuracy and correctness in parametric frequentist inference) at 9:00-13:30. The second one is given by Herman K. van Dijk (Simulation Based Bayesian Econometric Inference for Forecasting and Decision Analysis: Introduction and Recent Developments) at 15:00 - 19:30.

SPECIAL MEETINGS by invitation to group members

- CSDA Editorial Board meeting, *Birkbeck*, Friday 13th of December 2013, 18:00-19:30.
- CFE meeting, *Senate Room, Senate House*, Saturday 14th of December 2013, 18:40-19:00.
- ERCIM co-chairs meeting, *Senate Room, Senate House*, Sunday 15th of December 2013, 18:50-19:05 (by invitation).

SOCIAL EVENTS

- *The coffee breaks* will take place at the Macmillan Hall and Grand Lobby and top of ceremonial stairs of the Senate House.
- *Welcome Reception, Saturday 14th of December, from 20:00 to 21:30.* The Welcome Reception is open to all registrants and accompanying persons who have purchased a reception ticket. It will take place at the Hotel Russell (1-8 Russell Square, see map at page VIII). 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.
- *Christmas Conference Dinner, Sunday 15th of December, 20:30.* 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 VIII). The Dinner is in Honour of the CSDA Founding Editor Stan Azen and Co-Editor Jae C. Lee. Conference registrants and accompanying persons should bring their conference dinner tickets in order to attend the conference dinner.

Addresses of venues:

- University of London, Senate House, Malet Street, London WC1E 7HU.
- Birkbeck, University of London, Malet Street, London WC1E 7HX.

Registration, exhibitors and networking activities

The registration and exhibitors will be located in the MacMillan Hall of the Senate House. The MacMillan Hall will also be used for networking and internet access. The Room B04 at Birkbeck is available for meetings upon request.

Lecture rooms

The paper presentations will take place at the Senate House and at the main building of Birkbeck (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. The opening ceremony will take place at the Beveridge Hall of the Senate House. 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 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	120	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	Russell	40	Ground	Senate House
Athlone	30	Ground	Senate House	Holden	30	First	Senate House
Montague	30	Ground	Senate House	Deller	100	Lower Ground	Senate House
Senate	80	First	Senate House	Macmillan Hall	220	Ground	Senate House
349	70	Third	Senate House	G21A	30	Ground	Senate House
B02	40	Basement	Birkbeck Malet St	B18	65	Basement	Birkbeck Malet St
B20	99	Basement	Birkbeck Malet St	B33	165	Basement	Birkbeck Malet St
B34	222	Basement	Birkbeck Malet St	B35	125	Basement	Birkbeck Malet St
B36	123	Basement	Birkbeck Malet St	G16	60	Ground	Birkbeck Malet St
B30	40	Basement	Birkbeck Malet St	B29	30	Basement	Birkbeck Malet St
G15	48	Ground	Birkbeck Malet St				

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 to 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. IT technicians will be available during the conference and should be contacted in case of problems.

Posters

The poster sessions will take place at the Macmillan Hall. 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.

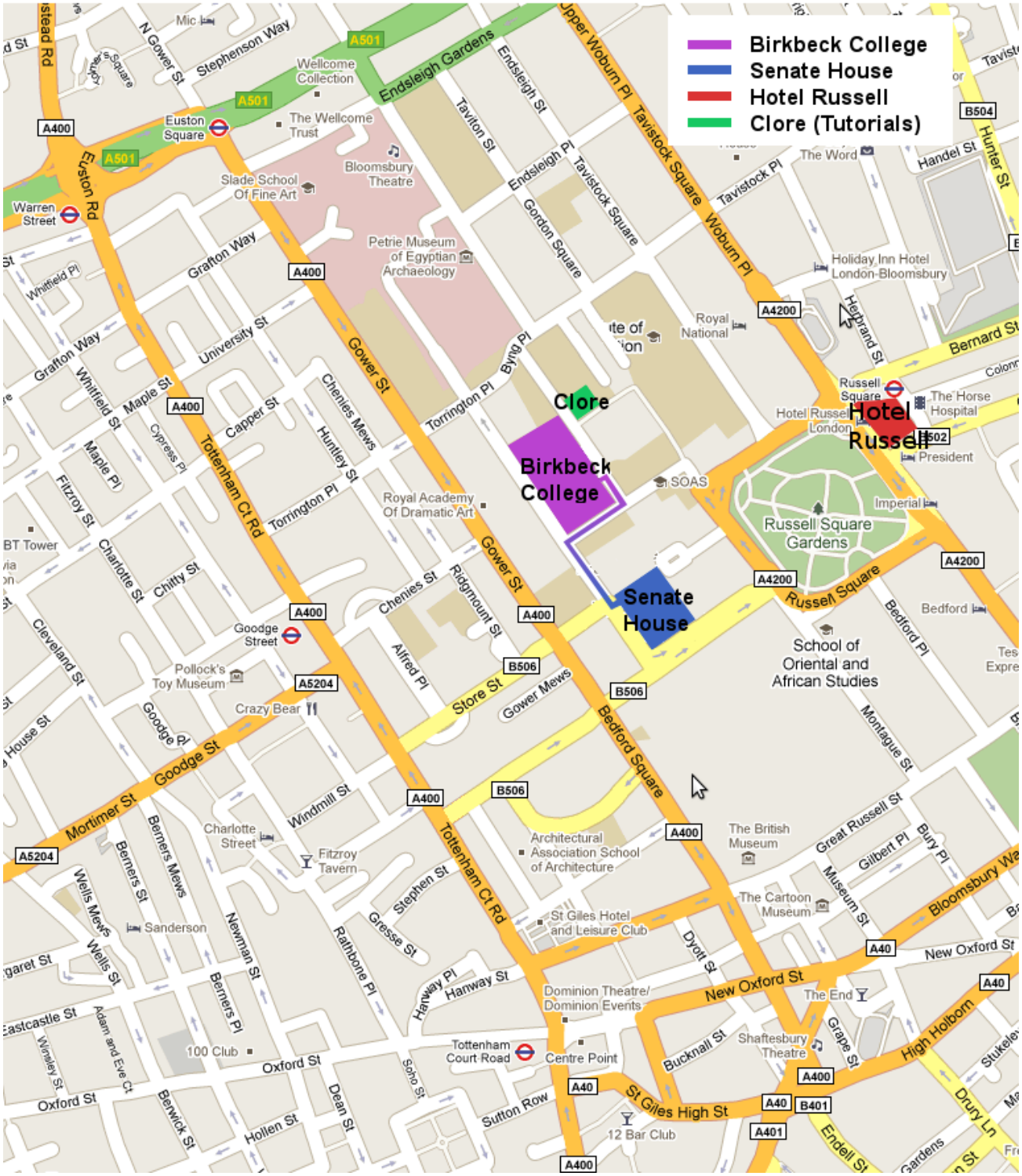
Internet

There will be 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 in the announcement board by the registration desk of the MacMillan Hall at Senate House.

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.

Map of the venue and nearby area



PUBLICATION OUTLETS

Journal of Computational Statistics & Data Analysis (CSDA)

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

Selected peer-reviewed papers will be published in the Journal of Computational Statistics & Data Analysis. Submissions for the CSDA should contain a strong computational statistics, or data analytic component.

Selected papers, which will be subject to peer review, will be considered for publication in a special issue, or in a regular issue of the journal Computational Statistics & Data Analysis. Theoretical papers or papers with simulation as the main contribution are not suitable for the special issues. Authors who are uncertain about the suitability of their papers should contact the special issue editors.

Papers will go through the usual review procedures and will be accepted or rejected based on the recommendations of the editors and referees. However, the review process will be streamlined to facilitate the timely publication of the papers. Papers that are considered for publication must contain original unpublished work and they must not be submitted concurrently to any other journal. Papers should be submitted using the Elsevier Electronic Submission tool EES: http://ees.elsevier.com/csda (in the EES please choose the appropriate special issue). All manuscripts should be double spaced or they will be returned immediately for revision.

Any questions may be directed via email to: csda@cfe-csda.org.

Annals of Computational and Financial Econometrics

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

Selected peer-reviewed papers will be published in the CSDA Annals of Computational and Financial Econometrics (as a supplement of the journal of Computational Statistics & Data Analysis). Submissions for the CSDA Annals of CFE should contain both a computational and an econometric or financial-econometric component.

Special Issues

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

- CSDA is planning for the 2013-2014 the following special issues:
  - Advances in Mixture Models.
  - Advances in Survival Analysis.
  - Advances in Data Mining and Robust Statistics.
  - Bayesian Econometrics.
  - Time Series Econometrics.

SPONSORS

http://www.qmul.ac.uk

http://www.bbk.ac.uk

http://www.lse.ac.uk

http://www.elsevier.com

http://www.ercim.eu

http://www.cajastur.es

EXHIBITORS

Elsevier (URL http://www.elsevier.com)

Numerical Algorithms Group (NAG) (URL http://www.nag.co.uk/)

Springer (URL http://www.springer.com)

Atlantis Press (URL http://www.atlantis-press.com)

Cambridge University Press (URL http://www.cambridge.org/)

Timberlake Consultants (URL http://www.timberlake.co.uk)

ENDORSED SOCIETIES & GROUPS

ERCIM Working Group on *Computational and Methodological Statistics*

Computational and Financial Econometrics *CFEnetwork*

The Society for Computational Economics

International Statistical Institute

International Association for Statistical Computing

Contents

Saturday 14.12.2013 08:35-09:25 Room: Beveridge Chair: Siem Jan Koopman Keynote talk CFE 1

Bootstrap methods for moment condition models

Speaker: **Richard J. Smith, University of Cambridge, UK**

Many if not most time series models in econometrics can be expressed in moment condition form. The application of bootstrap methods to such models is considered. As is widely appreciated bootstrap methods can deliver improved coverage for confidence regions, more accurate size for test statistics and better bias properties for estimators. Some examples of moment condition time series econometric models are initially presented together with brief descriptions of the generalised method of moments and generalised empirical likelihood estimation methods for dependent data. Various bootstrap methods for dependent data are next reviewed. A novel asymptotically valid bootstrap procedure appropriate for stationary times series is then introduced which initially transforms the original data using kernel reweighting together with an application to the transformed data of the standard m out of n bootstrap for i.i.d. observations. The kernel-based bootstrap with truncated (flat) kernel reweighting bears some similarities to the moving block bootstrap and to the transformation-based bootstrap which also uses a data transformation to obtain asymptotically independent observations in contradistinction to the kernel-based transformed data which remain asymptotically dependent. The efficacy of the kernel-based bootstrap is illustrated in a number of simulation studies.

Saturday 14.12.2013 10:00 - 10:50 Room: Beveridge Chair: Jae C. Lee Keynote talk ERCIM 1

The use of survey weights in regression modelling

Speaker: **Chris Skinner, London School of Economics, UK**

Survey weighting provides a unifying approach to point estimation in surveys. Weight construction often starts from inverse probability (Horvitz-Thompson) weighting and then incorporates auxiliary information through calibration methods. Although the practice of using weights in descriptive surveys is very established, their use in regression analysis and other modelling methods has a more contested history. The main focus will be on the uses of weights in regression modelling to correct for biases from informative sampling. The use of weights to protect against model misspecification will also be discussed. One disadvantage of weighting is that it can lead to a loss of efficiency. Some more recent work on weight modification to reduce this loss of efficiency will be presented. Gains from such modification will be shown in one application to the analysis of cross-national survey data.

Sunday 15.12.2013 09:35 - 10:25 Room: Beveridge Chair: John Galbraith Keynote talk CFE 2

Analyzing and forecasting US inflation

Speaker: **Herman K. van Dijk, Erasmus University Rotterdam, VU Amsterdam and Tinbergen Institute, The Netherlands**

Changing time series properties of US inflation and economic activity, measured as marginal costs, are modeled within a set of extended Phillips Curve (PC) models. It is shown that mechanical removal or modeling of simple low frequency movements in the data may yield poor predictive results which depend on the model specification used. Basic PC models are extended to include structural time series models that describe typical time varying patterns in levels and volatilities. Forward and backward looking expectation components for inflation are incorporated and their relative importance is evaluated. Survey data on expected inflation are introduced to strengthen the information in the likelihood. Use is made of simulation based Bayesian techniques for the empirical analysis. No credible evidence is found on endogeneity and long run stability between inflation and marginal costs. Backward-looking inflation appears stronger than forward-looking one. Levels and volatilities of inflation are estimated more precisely using rich PC models. The extended PC structures compare favorably with existing basic Bayesian vector autoregressive and stochastic volatility models in terms of fit and prediction. Tails of the complete predictive distributions indicate an increase in the probability of deflation in recent years.

Sunday 15.12.2013 12:20 - 13:10 Room: Beveridge Chair: Peter Mueller Keynote talk ERCIM 2

Bayesian inference in graphical models

Speaker: **Peter Green, Universityof Bristol, UK and UTS, Australia**

The structure in a multivariate distribution is largely captured by the conditional independence relationships that hold among the variables, often represented graphically, and inferring these from data is an important step in understanding a complex stochastic system. Simultaneous inference about the conditional independence graph and parameters of the model is known as joint structural and quantitative learning in the machine learning literature: it is appealing to conduct this in Bayesian paradigm, but this can pose computational challenges, because of the huge size of the model space that is involved, unless there are very few variables. After a general introduction to the area and the problems it poses, I will present some recent joint work with Alun Thomas (Utah), that exploits new results on perturbations to graphs that maintain decomposability and on enumeration of junction trees to construct a Markov chain sampler on junction trees that can be used to compute joint inference about structure and parameters in graphical models on quite a large scale. I will discuss some implications for inference about genetic networks.

Monday 16.12.2013 17:50 - 18:40 Room: Beveridge Chair: Alastair Young Keynote talk CFE-ERCIM

Signal detection in high dimension: Testing sphericity against spiked alternatives

Speaker: **Marc Hallin, Universite Libre de Bruxelle, Belgium**

The problem of testing the null hypothesis of sphericity is considered for a high-dimensional covariance matrix against the alternative of a finite (unspecified) number of symmetry-breaking directions (multispiked alternatives) from the point of view of the asymptotic theory of statistical experiments. The region lying below the so-called *phase transition* or *impossibility threshold* is shown to be a *contiguity* region. Simple analytical expressions are derived for the asymptotic power envelope and the asymptotic powers of existing tests. These asymptotic powers are shown to lie very substantially below the power envelope; some of them even trivially coincide with the size of the test. In contrast, the asymptotic power of the likelihood ratio test is shown to be uniformly close to the same.

Saturday 14.12.2013 09:35 - 10:50 Parallel Session B – CFE

CS08 Room Torrington MODELLING CREDIT RISK IN FINANCIAL MARKETS Chair: Alessandra Canepa

C943: **An empirical investigation about what causes banking crises**
Presenter: **David Meenagh**, Cardiff University, United Kingdom
Co-authors: Vo Phuong Mai Le, Patrick Minford

The Bernanke-Gertler-Gilchrist model is added to a modified version of the Smets-Wouters model of the US in order to explore the causes of the banking crisis. We then extract the model's implied residuals on US unfiltered data since 1984 to replicate how the model predicts the crisis. The main banking shock tracks the unfolding 'sub-prime' shock. This shock worsens the banking crisis but 'traditional' shocks explain the bulk of the crisis; the non-stationarity of the productivity shock plays a key role. Crises occur when there is a 'run' of bad shocks; based on this sample they occur on average once every 40 years and when they occur around half are accompanied by financial crisis. Financial shocks on their own, even when extreme, do not cause crises - provided the government acts swiftly to counteract such a shock as happened in this sample.

C987: **Reforms, incentives and banking sector productivity**
Presenter: **Kul Luintel**, Cardiff Business School, United Kingdom
Co-authors: Pushkar Bajracharya, Sheikh Selim

The aim is to study whether profit maximizing behaviour of banks, following financial liberalization, leads to distinct improvements in banking sector productivity. We model banks as profit cum welfare maximizing firms and derive their optimal level of efforts. Banking sector productivity is a function of banker's level of efforts, relative input-output prices and some technical parameters. Our micro-founded analytical model is put to test on a panel of Nepalese commercial banks. Nepal embarked into deep and far reaching banking and financial reforms during 1986-1994, which has brought about profound structural changes in the country's financial system. We employ up-to-date econometric methods in estimating the model. We find that following financial reforms, bankers' efforts and bank profitability have improved considerably but banking sector productivity has hardly improved. It appears that increased bankers effort is directed to attain higher level of profits through increased volumes of deposits and credits rather than the delivering productivity improvements.

C1100: **Housing market and credit risk: Evidence from the United States**
Presenter: **Alessandra Canepa**, Brunel University, United Kingdom
Co-authors: Mauro Costantini, Mohammad Tajik

The relation between real estate prices and credit risk is considered. Using a large panel of US banks the determinants of problem loans are investigated with special emphasis given to the role of real estate price fluctuations. This study also examines how real estate prices affect credit risk across different financial institution organisational structures and loan categories. It is found that house prices are able to convey signals about the evolution of riskiness of banks over the business cycle. Overall, there is substantial evidence that bad debts are an important transmission channel of macroeconomic shocks to balance sheets of banks.

CS12 Room Russell BAYESIAN NONLINEAR ECONOMETRICS Chair: Roberto Casarin

C517: **A Bayesian dynamic multi-factor model of instability in prices and quantities of risk**
Presenter: **Francesco Ravazzolo**, Norges Bank, Norway
Co-authors: Daniele Bianchi, Massimo Guidolin

The aim is to propose a Bayesian estimation framework for a typical multi-factor model with time-varying risk exposures and premia to price U.S. publicly traded assets. The model assumes that risk exposures and idiosyncratic volatility follow a break-point latent process, allowing for changes at any point on time but not restricting them to change at all points. The empirical application to 40 years of U.S. data and 23 portfolios shows that the approach yields sensible results compared to previous two-step methods based on naive recursive estimation schemes, as well as a set of alternative model restrictions. In particular, by considering model instability (together with parameter uncertainty), most portfolios do not appear to have been misspriced and a few risk premia are precisely estimated. A variance decomposition test shows that most of the predictable variation comes from the market risk premium. Yet, a Bayes factor investigation goes in favour of the model proposed. These results suggest that a more realistic Bayesian dynamic model with random structural breaks might sensibly improve the asset pricing performances of otherwise standard multi-factor linear models.

C527: **Particle learning for Bayesian non-parametric Markov switching stochastic volatility models with financial applications**
Presenter: **Audrone Virbickaite**, Universidad Carlos III de Madrid, Spain
Co-authors: Hedibert Lopes, Pedro Galeano, Concepcion Ausin

A Markov Switching Stochastic Volatility Model with DPM errors (MSSV-DPM) is estimated. The MSSV model can capture the structural changes in the volatility process, while the use of Bayesian non-parametric methods allows us to model the return distribution as a flexible mixture of infinite Gaussian distributions. The stick-breaking mixture weights are assumed to be dependent and evolve in time via a Beta auto-regressive process (BAR). Inference and prediction are carried out using a class of Sequential Monte Carlo Methods - Particle Learning. Finally, some real data illustrations and financial applications are presented.

C657: **Real-time learning, macroeconomic uncertainty, and the variance risk premium**
Presenter: **Daniele Bianchi**, Bocconi University, Italy

Macroeconomic uncertainty and periods of extreme market turmoil have recently plagued financial markets. This environment has prompted an increasing interest in option-like trading strategies aimed to capture a return premium over time, as a compensation for the risk of losses during sudden upward movements in market volatility (variance risk premium); these tend to coincide with general turbulence, financial/economic collapses, investors stress and high uncertainty as a whole. It is shown that the variance risk premium is tightly related to the level of uncertainty about the economic outlook. Unlike the reference literature we amend the conventional wisdom that views the agent observe the structure of the economy, and consider real time learning about both states and parameters. It is found that the variance premium peaks when the uncertainty about the economic fundamentals is higher. The underlying framework is a representative agent endowment economy with recursive preferences. Sequential learning is solved by a particle filtering and learning algorithm.

CS13 Room Gordon PROBABILISTIC FORECASTING: STATISTICAL ISSUES Chair: Wojtek Charemza

C295: **On the usefulness of product spreads for forecasting: An empirical evaluation of the Verleger hypothesis**
Presenter: **Christiane Baumeister**, Bank of Canada, Canada
Co-authors: Lutz Kilian, Xiaoqing Zhou

Notwithstanding a resurgence in research on out-of-sample forecasts of the price of oil in recent years, there is one important approach to forecasting the real price of oil which has not been studied systematically to date. This approach is based on the premise that demand for crude oil derives from the demand for refined products such as gasoline or heating oil. Oil industry analysts such as Philip Verleger and financial analysts widely believe that there is predictive power in the product spread, defined as the difference between suitably weighted refined product market prices and refiners' purchase price of crude oil. Our objective is to evaluate this proposition. We derive from first principles a number of alternative forecasting model specifications involving product spreads and compare these models to the no-change forecast of the real price of oil. We show that not all product spread models are useful for out-of-sample forecasting, but some models are, especially at horizons between one and two years. The most accurate model is a time-varying parameter model of gasoline and heating oil spot spreads that allows the marginal product market to change over time. We document MSPE reductions as high as 20% and directional accuracy as high as 63% at the two-year horizon, making product spread models a good complement to forecasting models based on economic fundamentals, which work best at short horizons.

C690: **Parsimonious models for probabilistic macroeconomic forecasting**
Presenter: **Patrick McSharry**, University of Oxford, United Kingdom

Forecasting the future is difficult and history has shown just how inadequate efforts have been in numerous disciplines. Quantifying the risk of extreme events is even more challenging given the need to understand the dynamics of the tail of the distribution. The need to seek parsimonious models that are capable of describing complex macroeconomic systems and yet simple enough to avoid over-fitting historical data is emphasised. Nonlinearity and regime-switching behaviour will be used to highlight the importance of using appropriate techniques for specifying, estimating and evaluating models. It will be argued that honest forecasts, involving multiple future scenarios and probabilistic information, and appropriate communication tools are required to foster collaboration between quantitative modellers and policy-makers. Whether the goal is to improve economic forecasts or develop innovative financial products for adapting to economic shocks, transparent and independently validated mathematical models will be necessary for managing these risks. A parsimonious nonparametric model for forecasting US GNP is used for demonstration purposes.

C481: **Forecasting VARs, model selection, and shrinkage**
Presenter: **Christian Kascha**, University of Zurich, Switzerland
Co-authors: Carsten Trenkler

Various shrinkage estimation and/or selection methods such as the LASSO have become popular. However, there is limited evidence on the forecasting performance of these methods for vector autoregressive (VAR) models, apart from evidence on the performance of Bayesian shrinkage. In particular, one does not know when there is a difference between these methods, which strategy is likely to work best under which circumstances and how these methods work in interaction with other specification choices such as the size of the VAR and its lag length. The aim is to try to fill part of this gap by comparing the forecast performance of (i) traditional selection methods (ii) regularization methods and (iii) empirical Bayesian methods for a quarterly US data set. The main results are as follows. It is found that an expanding estimation window combined with cross-validation for determining the tuning parameters of the models works best in the proposed context, even though other choices do not yield very different results. There was no method that worked best under all circumstances, however, traditional selection methods such as top-down selection or sequential elimination of regressors were dominated by the other methods. The choice of the lag length (estimation method) was often more decisive for forecasting performance than the choice of the shrinkage method. In agreement with the literature, it is found that augmenting the system size combined with shrinkage yields considerably better forecasting performance even though it is found that these improvements essentially disappear after horizon one.

CS16 Room Athlone MULTIVARIATE TIME SERIES METHODS FOR MACROECONOMICS Chair: Gianluca Cubadda

C266: **Realized volatility and business cycle fluctuations: A mixed-frequency VAR approach**
Presenter: **Alain Hecq**, Maastricht University, Netherlands
Co-authors: Marcelle Chauvet, Thomas Goetz

The link between uncertainty in financial markets and economic fluctuations is analyzed. In particular, we test for Granger non-causality between the monthly growth rate of Industrial Production Index and the logarithm of daily bipower variation, an observed measure of volatility, in both directions. Due to the mismatch of sampling frequencies, we analyze Granger non-causality testing in a mixed-frequency vector autoregressive model (VAR). Due to the constituents of the regressand-vector, we define the term nowcasting causality as the mixed-frequency version of instantaneous causality. We show that, starting from a mixed-frequency VAR, the presence or absence of nowcasting causality has a crucial impact on testing for Granger causality in standard mixed- frequency regression models. Furthermore, given a realistic sample size, the number of high-frequency observations per low-frequency period, 20 for our specific example, leads to parameter proliferation problems in case we attempt to estimate the VAR unrestrictedly. Hence, we propose several parameter reduction techniques, among which are reduced rank conditions and the imposition of an ARX(1) structure on the high-frequency variables. The performance of these various approaches in terms Granger non-causality testing behavior is investigated in a Monte Carlo experiment. Subsequently, the techniques are illustrated for the initial empirical research question.

C426: **Multivariate Markov-switching models and tail risk interdependence measures**
Presenter: **Mauro Bernardi**, Sapienza University of Rome, Italy
Co-authors: Antonello Maruotti, Lea Petrella

Markov switching models are often used to analyse financial returns because of their ability to capture frequently observed stylized facts. A multivariate Student-t version of the model is considered as a viable alternative to the usual multivariate Gaussian distribution, providing a natural robust extension that accounts for heavy-tails and time varying correlations. Moreover, these modelling assumptions allow us to capture extreme tail co-movements which are of fundamental importance to assess the underlying dependence structure of asset returns during extreme events such as financial crisis. For the considered model new risk interdependence measures are provided which generalise the existing ones, like the Conditional Value-at-Risk (CoVaR), capturing interconnections among multiple market participants. Application on a set of U.S. banks is considered to show that the right specification of the conditional distribution of the Markov process along with a multiple risk interdependence measure can help better understand how overall risk shares among institutions. This is particularly relevant during periods of market turbulence when several institutions may contemporaneously experience distress instances.

C618: **Nowcasting regional economic growth in the UK**
Presenter: **James Mitchell**, University of Warwick, United Kingdom

The aim is to set out a mixed-frequency multivariate unobserved components model suitable for the nowcasting and forecasting of regional output data in the UK. At present 'official' regional output data, from the UK's Office for National Statistics, are available at the annual frequency only

and are published with a considerable lag. This hinders decision making by both government and business. The proposed model uses quarterly employment data, published at the regional level, along with annual regional labour productivity data to infer regional output movements. Temporal and cross-sectional aggregation constraints are simultaneously imposed so that the quarterly regional data are both consistent with the annual totals and official quarterly data for UK GDP. Gains to waiting until these UK totals are published are found, 27 days after the end of the reference quarter, before estimating the quarterly regional output data. The out-of-sample performance of the methodology is then considered exploiting quarterly output data recently published for Scotland. The uncertainty associated with the regional nowcasts and forecasts is quantified and evaluated using density forecast methods. Comparisons are made with qualitative survey-based indicators of regional economic growth and it is seen that these can present a misleading impression of regional growth patterns. Finally, the model is used to shed light on the productivity puzzle, namely the low labour productivity observed in the aftermath of the 2007-8 financial crisis.

CS20 Room Bloomsbury NON-LINEAR DYNAMIC MODELS AND APPLICATIONS Chair: Laurent Ferrara

C288: **Cyclicality and term structure of Value-at-Risk within a threshold autoregression setup**
Presenter: **Frederique Bec**, THEMA University of Cergy-Pontoise and CREST, France
Co-authors: Christian Gollier

The link between stocks returns Value-at-Risk (VaR) and the state of financial markets cycle is empirically studied across various holding horizons. The econometric analysis is based on a self-exciting threshold autoregression setup. Using quarterly French and US data from 1970Q4 to 2012Q4, it turns out that the k-year VaR of equities is actually dependent on the cycle phase: the expected losses as measured by the VaR are smaller in bear market than in bull market, whatever the horizon. These results strongly suggest that the rules regarding the solvency capital requirements should adapt to the state of the financial market's cycle.

C796: **Financial conditions and density forecasts for US output and inflation**
Presenter: **Haroon Mumtaz**, Queen Mary University of London and Bank of England, United Kingdom
Co-authors: Piergiorgio Alessandri

The predictive power of financial indicators for output and inflation in the US is reassessed by studying predictive densities generated by a rich set of linear and nonlinear forecasting models. We argue that, if the linkage between financial and real economy is state-dependent as implied by standard models with financial frictions, predictive densities should reveal aspects of the comovements between financial and macroeconomic variables that are ignored by construction in an ordinary (central) forecasting exercise. We then study the performance of linear and nonlinear (Threshold and Markov-Switching) VARs estimated on a monthly US dataset including various commonly used financial indicators. We obtain three important results. First, adding financial indicators to an otherwise standard VAR improves both central forecasts and predictive distributions for output, but the improvement is more substantial for the latter. Even in a linear model, financial indicators are more useful in predicting tails, or deviations of output and inflation from their expected paths, than means, namely the expected paths themselves. Second, nonlinear models with financial indicators tend to generate noisier central forecasts than their linear counterparts, but they clearly outperform them in predicting distributions. This is mainly because nonlinear models predict the likelihood of recessionary episodes more accurately. Third, the discrepancies between models are themselves predictable: a Bayesian forecaster can formulate a reasonable real- time guess on which model is likely to be more accurate in the next future. This predictability gives place to a decision problem where the risk preferences of the forecaster take center stage.

C754: **Post-recession US employment through the lens of a non-linear Okun's law**
Presenter: **Laurent Ferrara**, Banque de France, France
Co-authors: Menzie Chinn, Valerie Mignon

The relationship between employment and GDP in the United States is investigated. We disentangle trend and cyclical employment components by estimating a non-linear Okun's law based on a smooth transition error-correction model that simultaneously accounts for long-term relationships between growth and employment and short-run instability over the business cycle. Our findings based on out-of-sample conditional forecasts show that, since the exit of the 2008-09 recession, US employment is on average around 1% below the level implied by the long run output-employment relationship, meaning that about 1.2 million of the trend employment loss cannot be attributed to the identified cyclical factors.

CS33 Room Jessel MODEL ESTIMATION AND PREDICTION IN ART MARKETS AND THE MACROECONOMY Chair: Christian Hafner

C020: **Individual artist career patterns and the hedonic prediction of art prices at auction**
Presenter: **Douglas Hodgson**, UQAM, Canada

Econometric analysis of realized auction prices for paintings shows that career effects as modelled by polynomial age–valuation profiles make a significant contribution to the in-sample fit in hedonic regressions of prices on a variety of explanatory characteristics. The existing research considers almost exclusively the estimation of data sets in which many artists are pooled into a single regression, with estimated age-valuation profiles representing an average over the artists included in the sample. Some attempts at disaggregation of artists, whether by birth cohort or artistic movement, have shown that important differences can occur across groups of artists. More recently, it has been demonstrated that with the aid of recent developments in dimensionality-reduction and model averaging, such profiles can be estimated at the individual- artist level. As the hedonic prediction of possible future sale values of individual paintings is of considerable interest to a number of art market stakeholders, including collectors and insurers, it is consequently also of interest to estimate specifications that will yield the best possible predictive power. We thus compare the out-of-sample predictive ability of a variety of hedonic specifications, focusing on the relative utility of models including age effects at the pooled and individual levels, using a data set on sales for Canadian painters extending from 1968 to 2011.

C792: **Estimation of daily art indices and applications**
Presenter: **Fabian Bocart**, UCL, Belgium
Co-authors: Christian Hafner

A new method is presented to produce hedonic art indices and provide volatility estimators. The method allows us to produce daily indices of prices of art sold at auction houses.

C1107: **Macroeconomic forecasting using VARs with time-varying volatility: A new approach**
Presenter: **Klaus Wohlrabe**, ifo Institute for Economic Research, Germany
Co-authors: Peter Zadrozny, Stefan Mittnik

A Cholesky-factor regularization method for estimating a VAR volatility model for single- or multi-frequency time-series data is described. The method is applied to single- and multi-frequency, financial and macroeconomic data. The key step in the method is adding a positive diagonal matrix to the outer-product matrix of residuals in every period, in order to make the matrix positive definite and, hence, Cholesky factorizable. A main advantage of the method appears to be that it allows prior data analysis, which leads to avoiding unnecessary, costly, nonlinear estimation of models and should make feasible the estimation of multivariate volatility models for more than a few series. We illustrate this approach by forecasting various macroeconomic variables from a large dataset. We compare standard small-scale VARs with the suggested approach. We show that our model is superior in terms of forecast efficiency.

CS39 Room Senate EARLY WARNINGS INDICATORS AND MACRO-PRUDENTIAL POLICY I Chair: Gianluigi Mazzi

C187: **International transmission of financial stress: Evidence from a GVAR**
Presenter: **Bjoern van Roye**, The Kiel Institute for the World Economy, Germany
Co-authors: Jonas Dovern
The international transmission of financial stress and its effects on economic activity are analyzed. We construct country specific monthly financial stress indexes (FSI) using dynamic factor models from 1970 until 2012 for 20 countries. We show that there is a strong co-movement of the FSI during financial crises and that the FSI of financially open countries are relatively more correlated to FSI in other countries. Subsequently, we investigate the international transmission of financial stress and its impact on economic activity in a Global VAR (GVAR) model. We show that i) financial stress is quickly transmitted internationally, ii) financial stress has a lagged but persistent negative effect on economic activity, and iii) that economic slowdowns induce only limited financial stress.

C146: **The impact of preferences on early warning systems: The European commission's scoreboard**
Presenter: **Tobias Knedlik**, Halle Institute for Economic Research, Germany
The European Commission's Scoreboard of Macroeconomic Imbalances is a rare case of a publicly released early warning system. It allows the preferences of the politicians involved to be analyzed with regard to the two potential errors of an early warning system – missing a crisis and issuing a false alarm. These preferences might differ with the institutional setting and for different stakeholders in macro-prudential policy in the European Union. Such an analysis is done for the first time for early warning systems in general by turning the computation of a standard signals approach to set thresholds, including a preference-based optimization approach, upside down. It is shown that, in general, the thresholds of the Commission's Scoreboard are set low (resulting in more alarm signals), as compared to a neutral stand. Based on political economy considerations the result could have been expected.

C1048: **A new indicator for real exchange rate misalignments in Europe**
Presenter: **Gregor von Schweinitz**, Halle Institute for Economic Research, Germany
Co-authors: Markam El-Shagi, Axel Lindner
It is frequently argued that real exchange rate misalignments between periphery countries and the core (in particular Germany) are one of the driving forces of the recent European debt crisis. However, the size of these misalignments is difficult to measure and up to now remains largely speculative. Using a synthetic matching approach, we generate the counterfactual real exchange rate development for several Euro area countries (Euro-12 without Luxembourg) if they had not joined the Euro. Thereby, we are able to evaluate the extent and dynamics of misalignments and can examine their role during the crisis. First evidence for Greece suggests that Greece is 30 percent overvalued compared to its synthetical counterpart.

CS46 Room Court THE ROLE OF THE RISK PREMIUM IN FINANCIAL MARKETS Chair: Jose Olmo

C112: **Unconventional monetary policies and the bank lending channel**
Presenter: **Marcos Sanso-Navarro**, Universidad de Zaragoza, Spain
Co-authors: Jose Olmo
Changes in the monetary policy transmission mechanism derived from appealing to unconventional measures are dealt with. Focusing on the bank lending channel, we study the effect on credit supply of lending facilities to recapitalize banks' balance sheets and of swap programs to exchange banks' toxic assets for government bonds. Our theoretical model shows the outperformance of the first program in decreasing the credit interest rate and the adverse effect the existence of a rate rewarding excess reserves has on the effectiveness of both programs. We also uncover the existence of a limit on these reserves beyond which these programs are counterproductive. The pass-through from the official rate to the credit interest rate also depends on the type of program and the level of excess reserves. An empirical analysis with macroeconomic data for the Euro area and the United States provides evidence of differences in the long-run equilibrium relationships between monetary aggregates and interest rates before and after the collapse of Lehman Brothers. Their extent and scope depend on the type of unconventional measure carried out.

C160: **A statistical test for multifractal and unifractal properties in financial returns**
Presenter: **Mark Hallam**, City University London, United Kingdom
Co-authors: Jose Olmo
Interest in the financial applications of unifractal and multifractal processes has increased over the past decade, with a large number of empirical studies finding that a wide variety of financial asset return series exhibit properties consistent with these classes of process. On the theoretical side, several models for asset returns have been proposed based on the more general class of multifractal processes, which provide methods for modelling and forecasting financial volatility and risk in a flexible and parsimonious manner. The main practical problem with the application of these models is that there is no established method for formally testing whether the properties of a given financial time series are indeed consistent with that of a multifractal process. Therefore a formal statistical procedure is proposed for testing whether a given sample of data is consistent with that of a multifractal process or the more restrictive class of unifractal processes. Several possible test statistics are proposed, all of which have non-standard distributions. Methods for obtaining the relevant distributions using appropriate resampling methods are developed and the resulting tests are shown to have good size and power properties in Monte Carlo exercises.

C1009: **An empirical investigation of the cross-section of interbank funding rates**
Presenter: **Jose Olmo**, Centro Universitario de la Defensa de Zaragoza, Spain
Co-authors: Giulia Iori, Burcu Kapar
Nonparametric methods for studying the cross-section of European interbank funding rates over the recent crisis periods are studied. Before the credit crunch of 2007 the distribution of the cross-section of borrowing and lending spreads is well-behaved. After it, this distribution becomes more disperse and multimodal, reflecting the existence of different bank clubs defined in terms of their creditworthiness. This pattern is reversed after March 2009, coinciding with the recovery of the financial sector. Our analysis also suggests that banks' characteristics such as nationality, operating currency and asset size have an important role in the cross-section of funding rates. These results are statistically confirmed by first order stochastic dominance tests. We observe that banks in the periphery of the European Union face larger borrowing rates than those of core economies; the adoption of the Euro acts as a shield against risk aversion for banks seeking/offering funding; borrowing rates are a decreasing function of banks' asset size. The empirical analysis also suggests that largest banks benefit from being able to set higher than average lending rates.

CS51 Room Woburn LONG MEMORY PROCESSES AND THEIR APPLICATIONS IN ECONOMETRICS Chair: Anne Philippe

C298: **Partial-sum limits for linear processes with changing memory and applications**
Presenter: **Remigijus Leipus**, Vilnius University, Lithuania
Co-authors: Frederic Lavancier, Anne Philippe, Donatas Surgailis
The limit distribution of partial sums of a nonstationary truncated linear process with long memory and changing memory parameter $d_{t,n}$ is studied. Two classes of linear processes are investigated: (i) the class of FARIMA-type truncated moving averages with time-varying fractional integration parameter and (ii) a class of time-varying fractionally integrated processes previously introduced. The cases of fast-changing memory parameter ($d_{t,n} = d_t$ does not depend on n) and slowly changing memory parameter ($d_{t,n} = d(t/n)$ for some function $d(\tau), \tau \in [0,1]$) are discussed. In the applications part, we discuss several test statistics based on the ratio of forward and backward sample variances of the partial sums. We study the behavior of these test statistics for some models with changing memory parameter. The consistency of the tests is proved under a very general setting.

C465: **Statistical inference for multistable processes**
Presenter: **Ronan Le Guevel**, University of Rennes, France
The multistable models for financial data are presented. The multistable processes are alpha-stable processes with an index of stability varying with time, with possibly discontinuous trajectories. The interest is the statistical estimation of two characteristic exponents of these processes: a measure of the regularity of the trajectories and the function of stability. Two consistent estimators of these exponents are introduced, and it is explained how to decide if the financial data arise from a multistable process through a statistical test.

C952: **State space modeling of Gegenbauer processes with long memory**
Presenter: **Gnanadarsha Dissanayake**, University of Sydney, Australia
Co-authors: Tommaso Proietti, Shelton Peiris
Approximation of a Gegenbauer autoregressive moving average (GARMA) process characterised by long memory using a finite order moving average is considered. Using the state space form the parameters are estimated by pseudo maximum likelihood via the Kalman filter. It is comparatively assessed initially with a finite order autoregressive approximation for choice and feasibility in establishing the order of the model. An extensive Monte Carlo experiment is executed to show that the optimal order is not very large (around 35), and rather insensitive to the sample size. A rolling forecasting experiment is performed to validate the choice of the order of the approximation in terms of predictive accuracy. Finally, the proposed state space methodology is applied to the famous yearly sunspots time series, and compared with other conventional and hybrid time series methods.

CS53 Room Montague FILTERS, WAVELETS AND SIGNALS Chair: Stephen Pollock

C789: **Poles of periodic filters with applications to cyclostationary and multivariate time series**
Presenter: **Georgi Boshnakov**, University of Manchester, United Kingdom
The dynamic characteristics of linear filters and associated time series models are determined by their spectral structure. The latter is typically specified by the eigenvalues/vectors of a model matrix (often taken from the state space form of the model) or, equivalently, by the latent roots/vectors of a characteristic polynomial. We show how periodic filters can be parameterised in terms of their eigenvalues and eigenvectors and give applications to simulation and estimation of some cyclostationary and multivariate time series models.

C812: **Anisotropy in random fields**
Presenter: **Sofia Olhede**, University College London, United Kingdom
Analyzing directional structures in images is important in many applications since one-dimensional patterns often correspond to important features such as object contours or trajectories. Most of the work on this topic has dealt with deterministic images; so far, little research has been performed on analyzing directionality in random fields. We propose a measure to quantify the degree of directionality in random fields. This measure can be used to construct statistical tests for the presence of directional structure. Our measure is based on the monogenic signal, which enables a unique decomposition of a 2D signal into local amplitude, local orientation, and local phase. We investigate the second-order statistical properties of the monogenic signal for isotropic, anisotropic, and unidirectional random fields. We analyze our measure of directionality for finite-size sample images, and determine a threshold to distinguish between unidirectional and non-unidirectional random fields.

C050: **Wavelet-based forecasting of chaotic time series**
Presenter: **Chee Kian Leong**, University of Nottingham Ningbo, China
Chaos may be present in economic time series and several tests have been proposed, including the BDS-test, which is based on the correlation integral. However, the estimation of the correlation dimension is not easy to achieve, especially when the dimension of the system increases. Moreover, it is extremely difficult to distinguish between high dimensional chaos and randomness, unless the time series is very long. The inadequacy of current estimation techniques to deal with chaos in time series may be the reason why despite the existence of elaborate forecasting models, forecasters cannot anticipate financial crisis, such as the Asian financial crisis and the recent US sub-prime market crisis. An attempt is made to harness the wavelet technology to develop an approach that can deal with chaos in time series forecasting. The approach can be implemented in software or embedded within applications to allow for real time online estimation and forecasting of time series with chaos.

CS68 Room Chancellor's MODELLING REGIME CHANGES I Chair: Willi Semmler

C895: **Scrapping subsidies in times of global crisis: How long is the shadow of fiscal stimulus**
Presenter: **Bernd Suessmuth**, University of Leipzig - CESifo, Germany
Co-authors: David Leuwer
The intertemporal strategic and unintentional allocative bias associated with a scrapping bonus for old cars is analyzed. To this end, we study the temporary installation of a "Cash for Clunkers" subsidy by the German government in 2009 that coincided with the need for fiscal stimulus in the advent of the global crisis. We use time series intervention analysis techniques for foreign and domestic order book levels to assess stylized idiosyncratic reactions of orders at the sectoral level. Strategic pull-forward effects on sales are found to imply transitory but not permanent structural disruptions in the automobile market (i.e. program reversal). Finally, we use a Kalman Filter approach to assess unintentional windfall profits realized by households due to the program. The actual net effect on vehicle orders confirms that scrappage policies do stimulate the automobile sector in the short run but create the basis of subsequent low activity. Although bearing a substantial fiscal stimulus, we find the scrappage scheme to have prolonged the sectoral business cycle trough by 13 months.

C999: **Leapfrogging and switching of leading and lagging position in economic growth**
Presenter: **Joachim Schnurbus**, University of Passau, Germany
Co-authors: Harry Haupt, Willi Semmler, Verena Meier
Classical growth convergence regressions fail to account for various sources of heterogeneity and nonlinearity. While recent contributions are able to address either the one or the other, we present a simple two-step method to address both issues. Based on a slightly augmented version of a recently proposed algorithm to identify convergence clubs, we formulate a flexible nonlinear framework which allows us to analyze convergence effects on both individual and club level, while alleviating potential misclassification in the club formation process using simultaneous smoothing over the club structure. The merits of the method are illustrated for two data sets. While applying our clubbing algorithm to the two data sets we can also explore the changes in the number, size, and composition of growth clubs over time.

C1150: **Macroeconomic activity, sovereign debt and financial market distress: The importance of non-linearities**
Presenter: **Christian Schoder**, Macroeconomic Policy Institute IMK, Germany
Co-authors: Christian Proano, Willi Semmler
Non-linearities in the relationship between the GDP growth rate, the sovereign debt-GDP ratio and the financial stress index are analyzed. We argue that the sovereign debt's negative effect on growth is more pronounced during episodes of high financial stress. A dynamic growth model is presented demonstrating how the financial market stress affects in a non-linear manner the dynamics of the economy due to severe macroeconomic amplifications which arise when the financial sector comes under distress. For 16 OECD countries the relationship is analyzed empirically for the period 1980-2010 using quarterly data and (panel) threshold regression methods. Our results suggest that both the debt-GDP ratio and the financial stress index reduce the growth rate of the economy only in the high-financial-stress regime. In the low-financial stress regime, the estimated relationship is mostly insignificant.

CS77 Room Bedford MODELLING VOLATILITY AND CORRELATIONS Chair: Helena Veiga

C252: **Covariance estimation of a locally stationary VAR**
Presenter: **Isabel Casas**, University of Southern Denmark, Denmark
Co-authors: Stefano Grassi
The vector autoregressive model (VAR) is a useful alternative to the structural econometrics model when the aim is to study macroeconomic behaviour. It exploits the fact that macroeconomic variables are interrelated and depend on historical data. Although VAR is highly used in practice, it is inflexible and it is not able to capture changes of behaviour in time. We assume that the process is locally stationary propose the time-varying nonparametric local linear estimator as an alternative to the Bayesian approach. The estimation of the covariance matrix, which may also be time-varying, is an important result. This matrix is decomposed to obtained to orthogonalised impulse response function to forecast the effect of a shock on the behaviour of another variable.

C505: **Modelling time-varying volatility in financial returns: Evidence from the bond markets**
Presenter: **Cristina Amado**, University of Minho, Portugal
Co-authors: Helina Laakkonen
The 'unusually uncertain' phase in the global financial markets has inspired many researchers to study the effects of ambiguity (or 'Knightian uncertainty') on the decisions made by investors and its implications for the capital markets. The aim is to contribute by using a modified version of the time-varying GARCH model to analyse whether the increasing uncertainty has caused excess volatility in the US and European government bond markets. In the model, volatility is multiplicatively decomposed into two time-varying conditional components: the first being captured by a stable GARCH(1,1) process and the second driven by the level of uncertainty in the financial market.

C511: **Detection of outliers in multivariate GARCH models**
Presenter: **Belen Martin-Barragan**, The University of Edinburgh, UK
Co-authors: Aurea Grane, Helena Veiga
In multivariate time series, correct estimation of the correlations among the series plays a key role in portfolio selection. In the univariate case, presence of outliers in financial data is known to lead parameter estimation biases, invalid inferences and poor volatility forecasts. The aim is to analyze the impact of outliers in multivariate time series. It is found that the impact in volatility follows a similar pattern to that in univariate time series, but, more interesting, the multivariate approach allows to analyze the impact on correlations. A general outlier detection procedure for multivariate time series is proposed, extending the wavelet-based method proposed in a previous work for the univariate case. The proposal is to extend this procedure to the context of multivariate GARCH models by considering random-projections of multivariate residuals. The method is general enough to deal with different multivariate GARCH models, such as the Constant Conditional Correlation, the Dynamic Conditional Correlation and the Diagonal BEKK. The effectiveness of this new procedure is evaluated through an intensive Monte Carlo study considering isolated and patches of additive level outliers and additive volatility outliers.

CS92 Room Holden CONTRIBUTIONS TO COMPUTATIONAL DECISION THEORY Chair: Richard Hahn

C087: **Emotional games**
Presenter: **Richard Fairchild**, University of Bath, United Kingdom
Behavioural economists are increasingly understanding that humans are not completely self-interested or emotionless, but often exhibit 'other-regarding' behaviour. We develop a game-theoretic approach in which players gain utility from their own material payoffs, but who also develop empathetic emotions towards each other. It has been argued that reciprocal behaviour may depend on the social context, social interaction, and psychological closeness of the players. Motivated by some analysis of sympathy games, we develop a two-stage, extended form, empathy model in which players simultaneously choose empathy levels in one stage, and, in a separate stage, make simultaneous strategy choices in a material game. We consider both conscious (strategic/instrumental), and unconscious (innate) empathy (you simply like someone for who they are). We demonstrate that increasing empathy can have positive or negative effects on welfare in equilibrium, and that these equilibria can be crucially affected by whether empathy is formed consciously, or unconsciously. We tentatively term our modelling approach, 'emotional game theory.'

C643: **Testing and selecting local proper scoring rules**
Presenter: **Emilio Zanetti Chini**, University of Rome Tor Vergata, Italy
In decision theory, an m-local scoring rule is defined as a function that assigns a pre-determined numerical value to a certain density forecast given some statistical model and depends on the predictive density only through its value and the values of its derivatives of order up to m at the realizing event. The most recent econometric literature suggests that these objects are useful for coherent aggregation and assessment of large ensembles of model-based forecasts. The issue of locality is looked as a testable hypothesis to be properly investigated. In particular, the aim is to hold with the problem of the presence of nesting scoring rules by providing a criterion to select between several plausible candidates. Two LM tests are introduced for the null of locality of underlining scoring rule and logarithmic form in case of nesting families. A simulation experiment shows that these tests can be successfully used for applied aims. An application to US business cycle shows that multiplying the recession indicator for different types of

scoring rules can affect the dating algorithm results and the model-based forecast performances in favor of a nonlinear specification, underlining in a such a way the importance of a correct score selection.

C1207: **Stochastic utility in the Ramsey model**
Presenter: **V L Raju Chinthalapati**, University of Greenwich, United Kingdom
Co-authors: Ramakanta Patra, Hans-Werner van Wyk
The aim is to re-investigate the predictions of the Ramsey-Cass-Koopmans model by incorporating individual's uncertainty in realizing utility into the model. We assume that a representatives consumer's derivation of utility varies across time both in terms of functional form and parameters. Thus, we take a hybrid approach to model this behavior into an endogenous growth mechanism. First, following standard approach, the central planner chooses the targeted welfare function by minimaxing the individual's rank preserving preference functions and parameters, both varying over time. One can use the standard techniques of calculus of variation and controls in order to gather sensible general conclusions on the functional form of the welfare function. Then, we take agent based modeling (ABM) approach in order to facilitate the learning of the parameters of the welfare function that the central planner is going to use. Using ABM, we simulate the participating individual's behavior that governs the randomized selection of utility functions at various time steps. From the ABM simulations, we generate the required empirical data for the output values of the welfare function and estimate the required parameters of the welfare function. We show that the growth differences among economies are better accounted for when we assume that individuals' preferences alter in a stochastic manner.

CS121 Room 349 CONTRIBUTIONS IN BANKING AND FINANCIAL MARKETS Chair: Martin M. Andreasen

C114: **On the credibility of the Euro/Swiss Franc floor: A financial market perspective**
Presenter: **Markus Hertrich**, University of Basel, Switzerland
Co-authors: Heinz Zimmermann
The sheer existence of EUR/CHF put options with strike prices below the 1.20 EUR/CHF floor, trading at non-zero cost, puts into question the full credibility of the Swiss National Bank (SNB) enforcing the lower barrier implemented since September 2011. We estimate the risk-neutral probabilities from observed market prices using two different currency pricing models, the Garman-Kohlhagen and an extension of the Veestraeten model, the latter assuming a lower barrier of the underlying security. We show that, in fact, these probabilities are considerably different from zero and how these have evolved over time. Our results contrast the publicly acclaimed credibility of the Swiss exchange rate policy vis-a-vis the Euro.

C131: **Flexible parameterization of economic frontiers: A novel technique and evidence from the US banking industry**
Presenter: **Peter Robejsek**, Durham University, United Kingdom
Co-authors: Panayiotis Andreou, Dennis Philip
A novel technique for estimating economic frontiers in the banking sector is proposed. While previous research assumes that the production function under which banks operate is either parametric or deterministic in nature, we apply the theory of asymmetric loss functions from the forecasting literature to artificial neural networks in order to parameterize the productive frontier. This new technique, referred to as neural frontier analysis (NFA), allows us to model the complex banking sector production function both non-parametrically and stochastically. Using a large sample of US banking sector data encompassing 118,177 bank year observations from 1994 to 2010, we estimate cost efficiency and shareholder value efficiency of the banking industry. We assess the efficiency results of NFA against the backdrop of similar results from the standard approaches of stochastic frontier analysis (SFA) and data envelopment analysis (DEA) using consistency criteria and multivariate regression analysis. We find that US banks are on average 75% shareholder value efficient. The NFA technique compares favorably with the standard approaches in terms of consistency criteria and regression analysis.

C961: **Hong Kong stock market overreaction to US returns: New insights from quantile regression analysis**
Presenter: **Robert Maderitsch**, University of Hohenheim, Germany
Intra daily price reversals in the Hang Seng Index Future following extreme movements in the US market are well documented in the literature for the years 1997 to 2003. Until present, however, it is unclear to what extent this phenomenon is reflected by the underlying Hang Seng Index, too. The reason is that index observations at market opening, required to calculate intraday returns, are unreliable due to stale quotes running into their computation. Resorting to intra daily trading data on all index constituents, however, we take these stale quotes explicitly into account. Based on a sample of daily S&P 500 and intra daily Hang Seng Index returns from 2000 to 2011, we reinvestigate the overreaction phenomenon employing non-parametric tests and quantile regressions. Particularly, we analyse the time variation of the structure and the degree of dependence. We find strong evidence for pronounced overreaction to occur in the index, too. Further we reveal substantial differences in the dependence pattern through time. Compared to recent studies using quantile regression techniques to investigate stock return autocorrelation, we most notably find a stronger dependence for central and upper index return quantiles.

CS124 Room Deller CONTRIBUTIONS IN NONLINEAR TIME SERIES Chair: Efthymios Pavlidis

C353: **On import price responses to depreciations and appreciations: Long lasting or temporary asymmetry**
Presenter: **Raphael Brun-Aguerre**, JPMorgan, United Kingdom
Co-authors: Ana-Maria Fuertes, Matthew Greenwood-Nimmo
The response of import prices to exchange rate shocks is studied in a nonlinear autoregressive distributed lag (NARDL) framework that accommodates sign asymmetry in the short- and long-run. Individual time-series models for 14 emerging and 19 developed importing economies suggest widespread sign asymmetry in the long-run but no general tendency toward asymmetry in the short-run. Depreciations are passed through more vigorously than appreciations, indicating a long-run 'rockets and feathers' effect in import prices. The extension of the proposed model into a nonlinear mean group framework reveals that long-run asymmetry is a robust and pervasive result which is not confined to a specific cohort of countries but which emerges under a wide variety of country groupings. The degree of import dependence is a key determinant of cross-sectional variation in the extent of the asymmetry, suggesting that exporters systematically exploit market power in their pricing decisions. Other significant determinants include importer's freedom to trade internationally and exchange rate volatility, both of which have a moderating effect on the role of import dependence. By contrast, the magnitude of exchange rate changes and the wealth of the importing economy exacerbate the role of import dependence. These findings have important policy implications.

C1163: **Classifying nonlinearities in financial time series**
Presenter: **Diana Mendes**, ISCTE-IUL, Portugal
Co-authors: Vivaldo Mendes, Pedro Ferreira
An elaborate search for the existence of deterministic non-linear dynamics (probable chaos) is carried out in the following financial series: the commodity of carbon allowances underlying the European Union Emission Trading Scheme for mitigation of CO2, the energy commodities from major fossil fuels, and the time series of utilities stock prices of the major European wind-turbine builders, having acknowledged the occurrence of nonlinear relationships (non-deterministic) between the prices of carbon allowances and their fundamentals, such as the prices of energy commodities, macroeconomic risk factors and climate. Thus, it makes sense to try to detect the occurrence of deterministic chaos in the observed signals

due to the persistent increase in fossil fuel prices, the overall growth of the economy, the increasing demand of energy commodities in developing countries, and the additional constraints on supply of fossil fuels, due to time delays in productive capacity investments. Namely, at one hand, the constraints on supply of energy fuels are fundamental for depth studies in deterministic chaotic motions of commodities and other assets in general and, in particular, the financial series of commodities observed. On the other hand, the industrial organizations of the energy commodities and utilities markets for the leading wind-turbine builders in Europe are essentially oligopolistic, which increases the interest in the estimation of deterministic chaos in the observed signals.

C291: **Early warning signals for critical transitions in finance**

Presenter: **Juanxi Wang**, University of Amsterdam, Netherlands

Co-authors: Cees Diks, Cars Hommes

As one of the highest complex nonlinear dynamical systems, financial markets are notably hard to predict, especially when it comes to some extreme circumstance, such as a financial crisis. We introduce the complexity theory of critical slowing down, a leading indicator of critical transitions in complex systems, to financial markets for the first time. We examined four critical transitions in financial systems using six time series. Using the combination indicator of AR(1), ACF, and variance trends, we got early warnings before 1987 Black Monday, 1997 Asia Crisis, 2000 Dot.com, and mixed results before 2008 Financial Crisis. We also checked the reliability of these results and did sensitivity tests. Our analysis showed that the critical slowing down was a fairly good early warning signal for 1987 Black Monday. For the other cases, the results are mixed. Concerning critical transitions in financial markets are hard to predict, a good perspective is given to future research and early warning signals could be thought of as a useful diagnostic tool.

Saturday 14.12.2013 11:20 - 13:00 Parallel Session D – ERCIM

ESI04 Room Beveridge STATISTICAL ALGORITHMS AND SOFTWARE Chair: Stanley Azen

E449: **The peer performance of hedge funds**
Presenter: **Kris Boudt**, KULeuven Lessius VU, Belgium
Co-authors: David Ardia
The modified Sharpe ratio (the ratio between a hedge fund's excess return and its Value-at-Risk estimate computed under the Cornish-Fisher expansion) is becoming a key tool for the practical assessment of the risk-adjusted performance of hedge fund managers. The applicability of this tool is generalized by proposing an equal modified Sharpe ratio performance test, and, an estimator for the fund's equal-performance ratio, defined as the percentage number of funds that have a non-significantly different modified Sharpe ratio. The proposed estimator first estimates the percentage of equal performance from the cross-sectional distribution of p-values of the two-sided test of equal performance of the fund and every peer fund. The remaining proportion is then attributed to the proportion of under- and out-performing funds. These estimates take the uncertainty in the numerator and denominator of the (modified) Sharpe ratio into account. The method is illustrated with an extensive simulation study, an application on real hedge fund data and the R package CompStrat.

E678: **Generalized nonlinear models in R**
Presenter: **Heather Turner**, University of Warwick, UK
Co-authors: David Firth, Ioannis Kosmidis
One way in which the class of generalized linear models can be extended, is to allow nonlinear terms in the predictor function. There are many applications where a nonlinear predictor provides a more parsimonious and interpretable model. However the move from linear to nonlinear presents some technical challenges. The approach taken in the gnm package for R will be presented, focusing on the innovative use of over-parameterized models. This allows models to be represented in a "natural" form, without needing to ensure all parameters are identified before model fitting. This approach relies on a IGLS algorithm, which will be summarized. Methods provided by gnm for subsequent model inference and display will be demonstrated in the examples presented. Further useful features of the gnm package will be highlighted, including the "eliminate" mechanism, which allows efficient estimation of incidental parameters.

E1248: **Statistical algorithms for neuroimaging genetics studies**
Presenter: **Giovanni Montana**, Imperial College London, United Kingdom
Neuroimaging genetics studies make use of imaging techniques to determine how individual genetic differences lead to anatomical and functional differences in the brain. An overview of sparse regression models that have been proposed to detect associations between genetic variants will be given, such as single nucleotide polymorphisms, and quantitative neuroimaging phenotypes. Given the large dimensionality of the data, efficient algorithms and implementations have been developed that exploit modern high-performance computing architectures including GPUs (graphical processing units). Example applications will be discussed.

ES05 Room Russell OPTIMUM EXPERIMENTAL DESIGNS FOR DRUG DEVELOPMENT AND ASSESSMENT Chair: Anthony C. Atkinson

E300: **Optimal design of clinical trials with biologics using dose-time-response models**
Presenter: **Markus Reiner Lange**, Novartis Pharma AG, Switzerland
Co-authors: Heinz Schmidli
Biologics such as monoclonal antibodies are increasingly and successfully used for the treatment of many chronic diseases. Unlike conventional small drug molecules, which are commonly given as tablets once daily, biologics are typically injected at much longer time intervals, i.e. weeks or months. Hence both the dose and the time interval have to be optimized during the drug development process for biologics. Semi-mechanistic nonlinear models are needed to adequately describe the dose-time-response relationship for biologics. We consider optimal experimental designs for estimation and prediction of these complex nonlinear models. Both locally optimal and Bayesian designs will be considered. For finding the optimal designs, we use the particle swarm optimization algorithm and compare this procedure with standard numerical optimizers. Particle swarm optimizers are successfully used in various applications, but have only recently been applied in finding optimal designs. The methodology is illustrated based on results from a clinical study with a monoclonal antibody.

E323: **Covariate-adjusted response adaptive penalized optimal design**
Presenter: **Luc Pronzato**, CNRS - University of Nice, France
Covariate-adjusted treatment allocation is considered, in the situation when a compromise must be made between information (about the dependency of the probability of success of each treatment upon the covariates) and cost (in terms of the number of patients receiving the poorest treatment). Information is measured through a classical design criterion for parameter estimation in the models describing the probability of success of each treatment, the cost is additive and is related to these probabilities. This formulation of the allocation problem leads to the definition of a penalized criterion for optimal design, the penalty term being related to the cost. When the covariates are i.i.d. random variables, the associated optimal design problem is shown to possess strong similarities with the construction of optimal bounded design measures. An equivalence theorem will be given and an adaptive allocation strategy will be proposed and shown to converge to the optimum. The strategy can easily be made response-adaptive for a practical application with unknown model parameters. The case with $T = 2$ treatments will be addressed, but the extension to $T > 2$ will also be considered. The introduction of randomization will be considered as well, inducing a slight modification in the corresponding equivalence theorem. Some illustrative examples will be presented.

E431: **Covariate-adjusted response-adaptive randomization designs for survival trials**
Presenter: **Oleksandr Sverdlov**, Novartis, United States
Co-authors: William Rosenberger, Yevgen Ryeznik
Covariate-adjusted response-adaptive (CARA) randomization is a novel class of randomization procedures for clinical trials where nonlinear and heteroscedastic models determine the relationship between responses, treatment assignments and covariates. Some new CARA randomization designs for survival trials are presented. The proposed designs are studied via extensive simulations under a variety of experimental scenarios, including cases when the model for event times is misspecified. The results show that the proposed CARA designs generate treatment allocations according to covariate-specific treatment differences and can result in fewer events in the trial, while having similar statistical properties (type I error, power, and estimation efficiency) to conventional balanced randomization designs. The impact of delayed responses, recruitment patterns, and the number of covariates in the model on operating characteristics of the designs is discussed. The utility of the proposed CARA randomization methodology is also illustrated by redesigning a survival trial from the literature.

E782: **Optimum dose regimen selection for a target drug concentration**
Presenter: **Barbara Bogacka**, Queen Mary University of London, United Kingdom
Co-authors: Kabir Soeny, Byron Jones
An optimum design is presented for dose regimen selection in clinical trials, where the criterion of optimality meets the requirement that the drug concentration be close to the target drug exposure needed to cure the disease. Furthermore, in cases when the drug is a combination of salts, we find an optimum ratio of the components as well as an optimum dosing regimen. We define new criteria of optimality and present their properties as well as the new optimization algorithm. We compare various dosing regimens and present a sensitivity analysis for the choice of the model parameter values.

ES07 Room Bloomsbury GENERALIZATIONS OF EMPIRICAL LIKELIHOOD Chair: Patrice Bertail

E784: **Empirical likelihood confidence sets for functional parameters**
Presenter: **Soumendra Lahiri**, North Carolina State University, United States
Construction of simultaneous confidence bands is considered for functional parameters using an unbounded number of constraints. The methodology employs a suitable version of the penalized empirical likelihood developed in recent years. Theoretical validity of the proposed methodology is established. Finite sample properties of the proposed method are investigated through a moderate simulation study.

E715: **Empirical Φ-divergence methods: A comparison**
Presenter: **Emmanuelle Gautherat**, CREST and Reims University, France
Co-authors: Patrice Bertail
The empirical Φ-divergences methods are generalized approaches to empirical likelihood method. The empirical Φ-divergences methods are used to test hypotheses and to construct some confidence region in a non-parametric setting. For small samples, the asymptotic chi-square calibration of the empirical ϕ-divergence has a default of coverage. The choice of Φ-divergence is then crucial. Moreover, for all of these methods, calibration of some constants is of prime importance. In the same spirit, for small samples or for non-symmetric distribution or for high dimension setting, different approaches have been proposed. We deal here with a comparison between these methods.

E765: **Targeted learning**
Presenter: **Antoine Chambaz**, Paris Ouest University, France
The targeted minimum loss estimation (TMLE) template for inference in semiparametric models has been quite intensively studied and applied in a variety of contexts. A brief introduction to TMLE inference will be given. The case of the nonparametric variable importance measure, a statistical parameter quantifying the effect of a continous exposure taking into account some relevant auxiliary variables, will serve as an example. Some interesting links with empirical likelihood methods will be discussed.

E696: **Empirical divergence minimizers for Hadamard differential functionals**
Presenter: **Patrice Bertail**, MODAL X Universite Paris-Ouest, France
Co-authors: Emmanuelle Gautherat
Some extensions of the empirical likelihood method are studied when the Kullback distance is replaced by some general convex divergence or $j-$discrepancy. We show that this generalized empirical likelihood method is asymptotically valid for general Hadamard differentiable functionals.

ES43 Room Holden STATISTICAL INFERENCE FOR NETWORKS Chair: Sofia Olhede

E430: **Recent results in statistical network analysis: A gentle introduction**
Presenter: **Patrick Wolfe**, University College London, United Kingdom
An overview of some recent results in statistical network analysis is given. The focus will be on limiting representations of large networks, and asymptotic regimes for inference that are well matched to these representations. This is an exciting and fast-moving area that presents both challenges and opportunities for modern statistical theory and methodology.

E081: **Model-based clustering of large networks**
Presenter: **David Hunter**, Pennsylvania State University, United States
Co-authors: Duy Vu, Michael Schweinberger
A network clustering framework, based on finite mixture models, is described. It can be applied to discrete-valued networks with hundreds of thousands of nodes and billions of edge variables. Relative to other recent model-based clustering work for networks, we introduce a more flexible modeling framework, improve the variational-approximation estimation algorithm, discuss and implement standard error estimation via a parametric bootstrap approach, and apply these methods to much larger datasets than those seen elsewhere in the literature. The more flexible modeling framework is achieved through introducing novel parameterizations of the model, giving varying degrees of parsimony, using exponential family models whose structure may be exploited in various theoretical and algorithmic ways. The algorithms, which we show how to adapt to the more complicated optimization requirements introduced by the constraints imposed by the novel parameterizations we propose, are based on variational generalized EM algorithms, where the E-steps are augmented by a minorization-maximization (MM) idea. The bootstrapped standard error estimates are based on an efficient Monte Carlo network simulation idea. Last, we demonstrate the usefulness of the model-based clustering framework by applying it to a discrete-valued network with more than 131,000 nodes and 17 billion edge variables.

E329: **The blessing of transitivity in sparse and stochastic networks**
Presenter: **Karl Rohe**, University of Wisconsin-Madison, United States
Co-authors: Tai Qin
The interaction between transitivity and sparsity, two common features in empirical networks, has surprising consequences for modeling and inference. Extant research suggests that estimating the Stochastic Blockmodel is more difficult when the edges are sparse. However, this has been confounded by the fact that the asymptotic limit in all of these studies is not only sparse, but also non-transitive. It is straightforward to show that a transitive block in a sparse Stochastic Blockmodel must be small. Previous algorithmic research demonstrates that small local clusters are more amenable to computation, visualization, and interpretation when compared to global graph partitions. We provide the first statistical results that demonstrate how these small transitive clusters are also more amenable to statistical estimation; we show that a local clustering algorithm can, with high probability, detect a transitive stochastic block of a fixed size embedded in a large graph. The only constraint on the ambient graph is that it is large and sparse; it could be generated at random or by an adversary, giving a theoretical explanation for the algorithm's robust empirical performance.

E703: **Maximum likelihood network estimates from social grouping behavior**
Presenter: **Yunpeng Zhao**, George Mason University, United States
Co-authors: Charles Weko
Within the field of network analysis, there is often an important distinction made between physical networks (i.e. highways, router systems, and electrical grids) and social networks (i.e. friendships, movie actors, and citations). In physical networks, the network topology is observable and analysis of the network properties directly informs the means by which the network functions. However, in social networks, the network topology is not explicit and must be inferred from the observed behavior. This effort is often complicated by the use of heuristic techniques for network inference which are not capable of reproducing the original behavior. A network based model to describe the social grouping behavior is defined. A maximum likelihood technique is presented for inferring the network most likely to have produced the observed behavior.

ES47 Room Court ROBUST STATISTICAL MODELING Chair: Alfio Marazzi

E179: **Robust estimation with high finite-sample efficiency through distance-constrained maximum likelihood**
Presenter: **Ricardo Maronna**, University of La Plata, Argentina
Co-authors: Victor Yohai
Good robust estimators can be tuned to combine a high breakdown point and a specified asymptotic efficiency at a central model. This happens in regression with MM- and tau-estimators among others. However, the finite-sample efficiency of these estimators can be much lower than the asymptotic one. To overcome this drawback, an approach is proposed for parametric models. It is based on the distance between parameters induced by the Kullback-Leibler divergence. Given a robust estimator, the proposed one is obtained by maximizing the likelihood under the constraint that the distance is less than a given threshold. For the linear model with normal errors and using the MM estimator, simulations show that the proposed estimator attains a finite-sample efficiency close to one, while its maximum mean squared error is smaller than that of the MM estimator

E322: **Robust estimators of the generalized loggamma regression model with censored observations**
Presenter: **Isabella Locatelli**, University of Lausanne, Switzerland
Co-authors: Alfio Marazzi
Robust estimators for accelerated failure time models with censored observations are proposed. It is assumed that the error model belongs to a location-scale-shape family of distributions on the log scale. In particular, the generalized loggamma distribution is considered. A three steps procedure is proposed. In the first step, the procedure computes a regression high breakdown point (hbp) S estimate $\hat{\beta}$ of the slopes. Then, it calculates a hbp Qtau estimate of the remaining parameters, minimizing a tau scale of the differences between empirical (Kaplan-Meier) and theoretical quantiles of the residuals wrt $\hat{\beta}$. Finally, it computes a weighted likelihood estimator with weights based on a disparity measure between the error model density and a kernel density estimate of the residuals. The final estimate attains full efficiency at the model, with respect to the maximum likelihood estimate, while maintaining the breakdown point of the S and Qtau estimators.

E347: **Robustness issues in linear-circular regression models**
Presenter: **Claudio Agostinelli**, Ca Foscari University, Italy
Co-authors: Alessandro Gagliardi
In many diverse scientific fields some or all the measurements are directions. A circular observation can be regarded as a point on a circle of unit radius, or a unit vector. For the nature of the circular observations, the analysis (descriptive and/or inferential) can not be carried out with standard methods used for observations in Euclidean space. Linear-circular regression models extend the usual linear models to the case where the dependent variable is a circular random variable, with von Mises distribution, and the explanatory variables are on the Euclidean space. We investigate the stability of maximum likelihood estimators for linear-circular regression models in the presence of a fraction of observations not generated by the actual model, i.e., in the presence of outliers. Alternative procedures are also considered.

E421: **Robust generalized method of wavelet moments**
Presenter: **Roberto Molinari**, Research Center for Statistics - University of Geneva, Switzerland
Co-authors: Stephane Guerrier, Maria-Pia Victoria-Feser
The estimation of complex or composite stochastic processes can easily be done via a consistent methodology called the Generalized Method of Wavelet Moments (GMWM). This method takes advantage of the implicit link between the parameters and the Wavelet Variance (WV) of the process. Based on a generalized least squares minimization principle, the GMWM estimates the WV of an observed process and estimates the parameters of the chosen model which minimize a distance between the observed WV and that implied by the model itself. It is intended to provide a robust version of the GMWM by proposing a robust estimation of the WV under Gaussian assumptions and provide its finite sample and asymptotic properties. This allows us to provide an easily computable and consistent robust estimator for time series and state-space models. Simulation studies show that this methodology delivers better results than other methodologies for a considerable number of stochastic processes.

ES92 Room Deller DIMENSION REDUCTION IN MIXTURE MODELLING Chair: Tsung-I Lin

E048: **Maximum likelihood inference for mixtures of common factor analyzers with missing information**
Presenter: **Wan-Lun Wang**, Feng Chia University, Taiwan
Mixtures of common factor analyzers (MCFA), thought of as a parsimonious extension of mixture factor analyzers (MFA), have recently been developed as a novel approach to analyzing high-dimensional data, where the number of observations n is not very large relative to their dimension p. The key idea behind MCFA is to reduce further the number of parameters in the specification of the component-covariance matrices. The occurrence of missing data persists in many scientific investigations and often complicates data analysis. A computationally flexible ECM algorithm for maximum likelihood estimation of the MCFA model with partially observed data is presented. To facilitate the implementation, two auxiliary permutation matrices are incorporated into the estimating procedure for exactly extracting the location of observed and missing components of each observation. Practical issues related to the model-based clustering and discriminant procedure are also discussed. The proposed methodology is illustrated through real and simulated examples.

E049: **Genotype copy number variations using Gaussian mixture models: Theory and algorithms**
Presenter: **Chang-Yun Lin**, National Chung Hsing University, Taiwan
Co-authors: Kenny Ye, Yungtai Lo
Copy number variations (CNVs) are important in the disease association studies, and are usually targeted by the most recent microarray platforms developed for GWAS studies. However, the probes targeting the same CNV regions could vary greatly in performance, with some of the probes carrying little information more than pure noise. We investigate how to best combine measurements of multiple probes to estimate copy numbers of individuals under the framework of Gaussian mixture model (GMM). First, we show that under two regularity conditions and assuming that all the parameters except the mixing proportions are known, optimal weights can be obtained so that the univariate GMM based on the weighted average gives exactly the same classification as the multivariate GMM does. We then develop an algorithm that iteratively estimates the parameters, obtains

the optimal weights, and uses them for classification. The algorithm performs well on simulation data and two sets of real data. This shows a clear advantage over classification based on the equal weighted average.

E091: **Discriminative variable selection with the sparse Fisher-EM algorithm**
Presenter: **Charles Bouveyron**, University Paris Pantheon-Sorbonne, France
The interest in variable selection for clustering has increased recently due to the growing need in clustering high-dimensional data. Variable selection allows us in particular to ease both the clustering and the interpretation of the results. Existing approaches have demonstrated the importance of variable selection for clustering, but turn out to be either very time consuming, or not sparse enough in high-dimensional spaces. It is proposed to perform a selection of the discriminative variables by introducing sparsity in the loading matrix of the Fisher-EM algorithm. This clustering method has been recently considered for the simultaneous visualization and clustering of high-dimensional data. It is based on a latent mixture model which fits the data into a low- dimensional discriminative subspace. Three different approaches are proposed to introduce sparsity in the orientation matrix of the discriminative subspace through L_1-type penalizations. Experimental comparisons with existing approaches on simulated and real-world data sets demonstrate the interest of the proposed methodology. An application to the segmentation of hyperspectral images of the planet Mars is also presented.

E601: **Model based clustering of high-dimensional binary data**
Presenter: **Yang Tang**, University of Guelph, Canada
Co-authors: Ryan P Browne, Paul D McNicholas
The aim is to present a mixture of latent trait models with common slope parameters (MCLT) for high dimensional binary data which few established methods exist. Recent work on clustering of binary data, based on a d-dimensional Gaussian latent variable, is extended by implementing common factor analyzers. The model is extended further by the incorporation of random block effects. The dependencies in each block are taken into account through block-specific parameters that are considered to be random variables. A variational approximation to the likelihood is exploited to derive a fast algorithm for determining the model parameters. BIC is used to select the number of components, the covariance structure as well as the dimensions of latent variables. The approach is demonstrated on U.S. Congress voting data and a data set describing the sensory properties of orange juice. The model gives a good fit given that the number of observations n is not very large relative to their dimension p and yield intuitive clustering results. Additionally, this dimensionality-reduction method allows the data to be displayed in low-dimensional plots.

ES59 Room Chancellor's ANALYSIS OF FUNCTIONAL DATA Chair: Gil Gonzalez-Rodriguez

E854: **Unifying amplitude and phase analysis: An analysis of Mandarin Chinese**
Presenter: **John Aston**, University of Warwick, UK
Co-authors: Pantelis Hadjipantelis, Hans Mueller, Jonathan Evans
Mandarin Chinese is characterised by being a tonal language and as such the pitch (or F0) of its utterances carries considerable linguistic information. However, speech samples from different individuals are subject to changes in amplitude and phase which must be accounted for in any analysis which attempts to provide a typology of the language. A joint model for amplitude, phase and duration is presented which combines elements from Functional Data Analysis, Compositional Data Analysis and Linear Mixed Effects Models. By decomposing functions via a functional principal component analysis, and connecting registration functions to compositional data analysis, a joint multivariate mixed effect model can be formulated which gives insights into the relationship between the different modes of variation as well as their dependence on linguistic and non-linguistic covariates. The model is applied to the COSPRO-1 data set, a comprehensive database of spoken Taiwanese Mandarin, containing approximately 50 thousand phonetically diverse sample F0 contours, and reveals that phonetic information is jointly carried by both amplitude and phase variation.

E452: **The Mahalanobis distance for functional data with applications to classification**
Presenter: **Rosa Lillo**, Universidad Carlos III de Madrid, Spain
Co-authors: Esdras Joseph, Pedro Galeano
The aim is to present a general notion of Mahalanobis distance for functional data that extends the classical multivariate concept to situations where the observed data are points belonging to curves generated by a stochastic process. Specifically, a new semi-distance for functional observations that generalize the usual Mahalanobis distance for multivariate datasets is introduced using a regularized square root inverse operator in Hilbert spaces. Some of the main characteristics of the functional Mahalanobis semi-distance are shown. In order to illustrate the applicability of this measure of proximity between functional observations, new versions of several well-known functional classification procedures are developed using the Mahalanobis distance. A Monte Carlo study and the analysis of two real examples indicate that the methods used in conjunction with the Mahalanobis distance give better results than other well known functional classification procedures.

E1268: **Spatial clustering for the temporal evolution of country population pyramids**
Presenter: **Pedro Delicado**, Universitat Politecnica de Catalunya, Spain
Consider n individuals having a geographical reference s, for which a functional random variable χ is observed at T consecutive instants of times, where $\chi : I \to \mathbf{R}$, and $I = [a,b] \subseteq \mathbf{R}$. The data are organized as a functional data set, each functional data having two arguments, one of them being time: $\{(s_i, \chi_i(t,x)), 1 \leq i \leq n\}$, with $\chi_i(t,x) = \chi_i^t(x)$, $t \in \{1,\ldots,T\}$, $x \in [a,b]$. We want to perform a cluster analysis, respecting geographical nearness, on the set of these dynamic functional data. The motivating example is the set of population pyramids from year 1950 to year 2010 for each country in the world ($n = 196$, $T = 61$). Our objective is to find clusters for the set of "movies" of population pyramids, that show their evolution from 1950 to 2010. An important point in this example is that clustering must take into account the geographical distribution of countries. That is, clusters with geographical continuity are desirable. In this context there are two main problems, namely, how to do the clustering and how to visualize the results, that are equally important.

E327: **Categorical functional data**
Presenter: **Cristian Preda**, University of Lille, France
Categorical functional data is considered as sample path of a continuous-time stochastic jump process with finite set of states. Spectral properties of the transition probability matrix are studied. As an extension of the multiple correspondence analysis, we give the expression of the eigenvalues and the optimal transformation of states for the Markovian case when the stationary distribution is reversible. The case of two states is detailed.

ES65 Room Woburn ADVANCES OF RARE VARIANT ANALYSIS IN LARGE SCALE GENETIC ASSOCIATION STUDIES Chair: Taesung Park

E077: **Genomic MB-MDR for rare variant analysis**
Presenter: **Kristel van Steen**, University of Liege, Belgium
Co-authors: Kyrylo Bessonov, Francois van Lishout, Jason Moore
The advances in whole-genome sequencing technologies have led to the availability of data on rare variants, with a large body of evidence for their involvement in human complex disease. A growing number of analysis techniques exist, usually focusing on specific trait types (either dichotomous or quantitative) or particular study designs (either using seemingly unrelated individuals or small numbers of extended pedigrees). When handling the genetic variants, most methods either adopt a single locus, a multiple locus or a collapsing strategy and either only consider rare variants or incorporate in the analysis common variants along with rare variants. Here, an alternative analysis technique is proposed that builds upon Model-Based Multifactor Dimensionality Reduction, an existing methodology that was primarily developed for GWAI studies. At the basis of the method lies a data organization step that involves a (natural) clustering of individuals according to genomic variants of interest, such as SNPs in MB-MDR 2D. In the context of rare variant analysis, individuals are clustered according to similarity based on rare and/or common variants linked to genes or genomic regions of interest, after which the MB-MDR machinery can be applied to investigate genomic associations with disease, irrespective of the study design or trait type.

E673: **Residual-based MDR for detecting gene-gene interaction associated with the survival phenotype**
Presenter: **Seungyeoun Lee**, Sejong University, Korea, South
Co-authors: Youngkang Kim, Jinseok Oh, Min-Seok Kwon, Taesung Park
While most of statistical methods in genome-wide association studies (GWAS) have been developed to identify genetic variants based on a binary phenotype for case-control studies, there is an interest in identifying genetic variants associated with the survival phenotype in cohort studies. A previous work first proposed a novel method, called Surv-MDR, for detecting gene-gene interactions associated with the survival time by using the log-rank test statistic in the frame of the multifactor dimensionality reduction (MDR) method. However, Surv-MDR has some limitations that it cannot adjust the covariates and requires rather intensive computations. Another work alternatively proposed Cox-MDR which uses a martingale residual from Cox model to classify multi-level genotypes into high and low risk groups. The big advantage of Cox-MDR is to allow the covariate effect and to need much less computations than Surv-MDR. The purpose is to generalize the use of the residual score in identifying gene-gene interaction associated with the survival time. For an accelerated failure time model, both a standardized residual and a martingale residual can be used to classify multi-level genotypes into high and low risk groups while keeping the rest of the MDR procedure unchanged. Through out simulation studies, the proposed method is compared with Cox-MDR as well as Surv-MDR in detecting two-way interactions associated with the survival phenotype. In addition, the effect of adjusting covariates on the power is investigated by considering the correlations between the survival time and covariates.

E746: **Gene-gene interaction analysis for rare variants**
Presenter: **Taesung Park**, Seoul National University, Korea, South
Co-authors: Min-Seok Kwon
With a rapid advance of array-based genotyping techniques, genome-wide association studies (GWAS) have successfully identified common genetic risk factors associated with common complex diseases. However, it was shown that genetic etiology of complex diseases could be rarely explained by the genetic factors identified from GWAS. This missing heritability could be partly due to gene-gene interaction and rare variants. There has been an exponential growth of gene-gene interaction analysis for common variants in terms of methodological developments and practical applications. Also, the recent advance of high-throughput sequencing technologies makes it possible to conduct rare variant analysis. However, little progress has been made in gene-gene interaction analysis for rare variants. Here, we propose a new gene-gene interaction method for the rare variants in the framework of the multifactor dimensionality reduction (MDR) analysis. The proposed method consists of two steps. The first step is to collapse the rare variants in a specific region such as gene. The second step is to perform MDR analysis for the collapsed rare variants. The proposed method is illustrated with 1080 whole exome sequencing data of Korean population to identify causal gene-gene interaction for rare variants for type 2 diabetes.

E1231: **Network-based omics integration framework: Overcoming the obstacle of high dimensional data**
Presenter: **Francesco Gadaleta**, University of Liege, Belgium
Co-authors: Kyrylo Bessonov, Kridsadakorn Chaichoompu, Silvia Pineda, Kristel van Steen
In genome-wide association studies, data integration, as the practice of combining evidence from different data sources, represents a challenging activity, due to the unattainable task of merging large and heterogeneous data sets. A practical framework that fulfills the needs of integration is provided by means of networks. Due to the manifest risks of introducing bias when integrating heterogeneous data and because of the curse of dimensionality, which is a common aspect in computational biology, preliminary procedures of variable selection are essential. We investigate two approaches that capture the multidimensional relationships between different data sets, namely conditional inference trees and linear regression methods. A strategy that is common to both methods consists in using every expression trait as a dependent variable and the remaining expression traits as covariates. The aforementioned strategy will derive variable importance scores that, in turn lead to selecting important variables. We subsequently proceed with the construction of weighted networks that integrate evidence gathered from different data sets with the purpose of detecting pathways and interactions among genetic compounds and better describing the complex mechanisms behind the traits under study.

ES67 Room Senate BAYESIAN SEMI- AND NONPARAMETRIC MODELLING I Chair: Antonio Lijoi

E376: **Multi-resolution adaptive shrinkage in nonparametric inference**
Presenter: **Li Ma**, Duke University, United States
An extension to the Polya tree and optional Polya tree processes for generating random probability distributions is introduced. By incorporating multiple shrinkage states and the transition among these states on different parts of the sample space in a Markovian fashion, the new construction achieves multi-resolution adaptive shrinkage in posterior inference. The inference has multi-resolution in the sense that the prior applies a varying amount of shrinkage to data structures of different scales. It is also adaptive in that the appropriate amount of shrinkage is determined through the local behavior of the data and so structures of the same scale may receive different amounts of shrinkage. The theoretical properties of the process are studied. In particular, it is shown that it possesses large support, posterior conjugacy, and posterior consistency. Analytic recipes for marginalization and for computing the posterior through recursion are provided. It is also shown that the process can be centered around any parametric family, and is invariant under change-of-variable. The last two properties are used to derive a convenient strategy for inference using mixtures of the process. We present a finite version of the process based on truncating the partition of the sample space at a given resolution, and establish its posterior consistency. We illustrate how to achieve multi-resolution adaptive shrinkage using the process through two numerical examples in density estimation.

E443: **Random function priors for exchangeable databases**
Presenter: **James Lloyd**, University of Cambridge, United Kingdom

The task of statistical inference for relational data is considered with multiple relations between heterogenous sets of objects, i.e. any relational database. A commonly applicable assumption is exchangeability of the objects upon which the relations act. When the data are encoded in the form of arrays this assumption corresponds to invariance to the ordering of rows and columns, i.e. exchangeable arrays. Recent work has connected representation theorems to the modeling of single exchangeable arrays. This work is extended by deriving corollaries of the representation theorems aplicable to collections of arrays with shared exchangeability properties. This allows for further clarification of the similarities between different models in the literature. The corollaries also elucidate how to extend models for single relations to the case of multiple relations and design principles for model classes are discussed that will likely extend well in practice.

E219: **Bayesian modeling of temporal dependence in large sparse contingency tables**
Presenter: **Tsuyoshi Kunihama**, Duke University, United States
Co-authors: David Dunson

In many applications, it is of interest to study trends over time in relationships among categorical variables, such as age group, ethnicity, religious affiliation, political party and preference for particular policies. At each time point, a sample of individuals provide responses to a set of questions, with different individuals sampled at each time. In such settings, there tends to be abundant missing data and the variables being measured may change over time. At each time point, one obtains a large sparse contingency table, with the number of cells often much larger than the number of individuals being surveyed. To borrow information across time in modeling large sparse contingency tables, we propose a Bayesian autoregressive tensor factorization approach. The proposed model relies on a probabilistic Parafac factorization of the joint pmf characterizing the categorical data distribution at each time point, with autocorrelation included across times. Efficient computational methods are developed relying on MCMC. The methods are evaluated through simulation examples and applied to social survey data.

E599: **Some thoughts on Bayesian inference for semiparametric regression models with shape-constraints**
Presenter: **Taeryon Choi**, Korea University, South Korea
Co-authors: Peter Lenk, Sooyeon Lee

The aim is to consider Bayesian inferential procedures for the semiparametric regression models when the constraint of the regression function needs to be incorporated in modeling. For this purpose, some aspects of Bayesian inference with constraints are provided and several Bayesian approaches to the regression model with functional constraints are discussed that include the shape-restricted function estimation problems using orthogonal series expansion and regression problems subject to uncertainty about the functional constraints. Empirical performance of the proposed method is illustrated based on synthetic data and real data applications by comparison with other existing competitors, and the asymptotic properties are investigated.

ES69 Room Gordon ROBUST ANALYSIS OF HIGH-DIMENSIONAL DATA Chair: Juan Romo

E136: **Local depth for functional data**
Presenter: **Davy Paindaveine**, Universite libre de Bruxelles, Belgium
Co-authors: Germain van Bever

Statistical depth allows us to measure centrality of an arbitrary location with respect to a probability distribution. This concept, which has been introduced in the context of finite-dimensional Euclidean spaces, has many possible inferential applications, in particular for classification. Recently, various concepts of depth have been introduced for functional data. Parallel to their finite-dimensional antecedents, however, these depth concepts can only deal with unimodal convexly-supported distributions. We therefore introduce a 'local' extension of depth that can get rid of such limitations. The construction applies both for the finite- and infinite-dimensional cases. We derive some basic properties of the proposed local depth concept and discuss possible inferential applications. In particular, we propose classification procedures based on local depth.

E448: **Some geometric tools in functional data analysis and nonparametrics**
Presenter: **Antonio Cuevas**, Autonomous University of Madrid, Spain
Co-authors: Alejandro Cholaquidis, Ricardo Fraiman

In the setting of binary classification with functional data, the aim is to analyze the theoretical and practical consequences of using a "visual" metric, defined in terms of the Hausdorff distance between the hypographs. The use of such a metric is particularly meaningful in those cases where the involved functions are assumed to fulfil a mild convexity-type condition called ρ-cone-convexity. Some applications outside functional data analysis (e.g. in the classical problem of estimation of the efficient boundary) are also briefly outlined.

E556: **Supervised learning from multivariate functional data**
Presenter: **Francesca Ieva**, Politecnico di Milano, Italy
Co-authors: Rachele Biasi, Anna Paganoni, Nicholas Tarabelloni

Nowadays data coming from a biomedical context are frequently functions produced by medical devices. This calls for the identification of suitable inferential techniques for managing their complexity. The aim is to present some inferential tools for dealing with multivariate functional data, i.e., data where each observation is a set of possibly correlated functions. It aims at modeling the binary outcome of interest (the presence of cardiovascular ischaemic event), estimating the probability of each patient to be affected by acute myocardial infarction. First, the 8-leads ECG trace of each patient and its first derivative are used as multivariate functional predictor in a generalized regression model, reducing their dimensionality via multivariate functional PCA, in order to identify a robust base of coefficients that enables a reliable prediction for a new patient entering the study. On the other hand, data are summarized using an extension of Depth Measure to the multivariate functional setting. In particular, to define a multivariate depth measure, it weights the contribution of each component according to the estimated correlation among data components themselves. Data depths are then used within the regression model aimed at predicting patient's status. Finally, the performances of the two approaches are compared in terms of patients classification.

E582: **DD-classification of functional data**
Presenter: **Karl Mosler**, Universitaet zu Koeln, Germany
Co-authors: Pavlo Mozharovskyi

Functional data have become increasingly important in many fields of application. To classify such data into $k \geq 2$ classes they are first transformed into a so-called DD-plot in the k-dimensional unit cube and then classified within this cube by a proper efficient procedure like the α-procedure or KNN. Each function is transformed into a vector that consists of the depth values of it regarding the k classes. Several depth notions for functional data are discussed and employed for this task, and the respective classifiers are compared. Results are given for simulated as well as real data.

ES70 Room Athlone SPATIAL ECONOMETRICS Chair: Anne Ruiz-Gazen

E417: **Permutation tests for labeled network analysis**
Presenter: **Nathalie Villa-Vialaneix**, SAMM, France
Co-authors: Thibault Laurent
Graphs (also called networks) are widely used to model data in many application fields to model interaction data (e.g., social network, regulation network in biology, Internet network...). In a number of real world situations, additional information is available, that describes each node of the network. For instance, the nodes in a social network can be labeled by the individual's membership to a given social group; the node of protein interaction networks can be labeled with the protein family... In this context, it is important to understand if the labels of the nodes are somehow related to the network topology (which can give an insight on the reason why two nodes are connected). This issue is addressed and exploratory tools are presented, widely used in spatial statistics, that can be used to answer this question. More precisely, we will focus on permutation tests and explain how these tests should be performed depending on the nature of the labels (numeric labels, factors, spatial labels...). The methods are illustrated on a real world dataset which is a social network extracted from a large corpus of medieval notarial acts.

E561: **Spatial concentration of a marked point pattern with application to industrial location**
Presenter: **Christine Thomas-Agnan**, Universite Toulouse, France
Co-authors: Florent Bonneu
A unified framework for defining measures of industrial concentration based on micro-geographic data is proposed. These encompass the indices introduced previously. The basic requirements for such previous measures are discussed and four additional requirements are proposed. The null assumptions classically used for testing aggregation of a particular sector are also discussed. The proposed mathematical framework is based on some second order characteristics of marked spatial point processes discussed previously. The general index measure involves a cumulative and a non-cumulative version. This allows us to propose an alternative version of the Duranton-Overman index with a proper baseline as well as a cumulative version of this same index. An additional requirement is then introduced about the inhomogeneity of the spatial location of firms and several types of concentration are described depending on the second order characteristics of the marginal patterns of locations and of firms' sizes (marks) but also on their mutual dependence.

E596: **Accuracy of areal interpolation methods**
Presenter: **Van Huyen Do**, University Toulouse, France
Co-authors: C. Thomas-Agnan, A. Vanhems
The analysis of socio-economic data often implies the combination of data bases originating from different administrative sources so that data have been collected on several different partitions of the zone of interest into administrative units. It is therefore necessary to allocate the data from the source spatial units to the target spatial units. A particular case of that problem is when one uses a common regular grid and re-allocate everything to this single target partition. This option is currently under study in France at INSEE and in Europe with the EU directive 'INSPIRE', or INfrastructure for SPatial InfoRmation. There are three main types of such techniques: proportional weighting schemes, smoothing techniques and regression based interpolation. A model based on Poisson point patterns is proposed to consider the accuracy of these techniques for the regular grid targets. The error depends on the nature of variable of interest and its correlation with the auxiliary variable. The focus is on proportional weighting schemes and regression based methods. There is no technique which always dominates.

E652: **Using indirect and feedback effects to measure geographic knowledge spillovers in French regions**
Presenter: **Thibault Laurent**, Toulouse School of Economics, France
Co-authors: Ines Moussa
The purpose is to measure more directly the role of geographical proximity in the innovation process on the French NUTS 3 regions (94 French metropolitan departments) over the 1997-2008 period. The innovation process (the dependent variable) is measured by the number of patents per department. A spatial decomposition coefficient of the independent variables (including the total outsourced R&D per capita, the internal R&D expenses, the population density, etc) is proposed to measure more explicitly geographical spillovers. In addition, the role of innovation networks in the diffusion of knowledge is studied. Classical panel data econometrics tools and spatial panel econometrics model are used to assess spillovers, measured by feedback effects and indirect effects. In this perspective, it is proposed to model a knowledge production function using two autoregressive models (SAR and SDM). The estimation results confirm spatial dependency and positive spillovers effects.

ES73 Room Torrington RECENT ADVANCES IN NON- AND SEMIPARAMETRIC REGRESSION Chair: Melanie Schienle

E192: **Nonparametric instrumental regression with non-convex constraints**
Presenter: **Anne Vanhems**, University of Toulouse- TBS and TSE, France
Co-authors: Markus Grasmair, Otmar Scherzer
The nonparametric regression model with an additive error that is dependent on the explanatory variables is considered. As is common in empirical studies in epidemiology and economics, it also supposes that valid instrumental variables are observed. A classical example in microeconomics considers the consumer demand function as a function of the price of goods and the income, both variables often considered as endogenous. In this framework, the economic theory also imposes shape restrictions on the demand function, like integrability conditions. Motivated by this illustration in microeconomics, we study an estimator of a nonparametric constrained regression function using instrumental variables by means of Tikhonov regularization. We derive rates of convergence for the regularized model both in a deterministic and stochastic setting under the assumption that the true regression function satisfies a projected source condition including, because of the non-convexity of the imposed constraints, an additional smallness condition.

E303: **Nonparametric estimation of a heterogeneous demand function under restrictions of consumer theory**
Presenter: **Matthias Parey**, University of Essex, United Kingdom
Co-authors: Richard Blundell, Joel Horowitz
Economic theory rarely provides a parametric specification for a model, but it often provides shape restrictions. We consider nonparametric estimation of the heterogeneous demand for gasoline in the U.S. subject to the Slutsky inequality restriction of consumer choice theory. We derive conditions under which the demand function can be estimated consistently by nonparametric quantile regression subject to the Slutsky restriction. The estimated function reveals an important variation in price responsiveness across the income distribution. A new method for estimating quantile instrumental variables models is also developed to allow for the endogeneity of prices. In our application, shape-constrained quantile IV estimates show similar patterns of demand as shape-constrained estimates under exogeneity. The results illustrate the improvements in the finite-sample performance of a nonparametric estimator that can be achieved by imposing shape restrictions based on economic theory.

E396: **Significance testing in quantile regression**
Presenter: **Holger Dette**, Ruhr-Universitaet Bochum, Germany
Co-authors: Melanie Birke, Natalie Neumeyer, Stanislav Volgushev
The problem of testing significance of predictors in multivariate nonparametric quantile regression is considered. A stochastic process is proposed, which is based on a comparison of the responses with a nonparametric quantile regression estimate under the null hypothesis. It is demonstrated that under the null hypothesis this process converges weakly to a centered Gaussian process and the asymptotic properties of the test under fixed and local alternatives are also discussed. In particular it is shown that - in contrast to the nonparametric approach based on estimation of L2-distances - the new test is able to detect local alternatives which converge to the null hypothesis with parametric rate. A small simulation study is also presented illustrating the finite sample properties of a bootstrap version of the corresponding Kolmogorov-Smirnov test.

E463: **Model selection via Bayesian information criterion for quantile regression models**
Presenter: **Eun Ryung Lee**, University of Mannheim, Germany
Co-authors: Hohsuk Noh, Byeong U. Park
Bayesian Information Criterion (BIC) is known to identify the true model consistently as long as the predictor dimension is finite. Recently, its moderate modifications have been shown to be consistent in model selection even when the number of variables diverges. Those works have been done mostly in mean regression, but rarely in quantile regression. The best known results about BIC for quantile regression are for linear models with a fixed number of variables. We investigate how BIC can be adapted to high-dimensional linear quantile regression and it is shown that a modified BIC is consistent in model selection when the number of variables diverges as the sample size increases. We also discuss how it can be used for choosing the regularization parameters of penalized approaches that are designed to conduct variable selection and shrinkage estimation simultaneously. Moreover, the results are extended to structured nonparametric quantile models with a diverging number of covariates. The theoretical results are illustrated via some simulated examples and a real data analysis on human eye disease.

ES80 Room Montague ESTIMATION FOR DIFFUSION PROCESSES Chair: Frank van der Meulen

E783: **Non-linear lasso and drift estimation for multivariate diffusions**
Presenter: **Niels Richard Hansen**, University of Copenhagen, Denmark
The lasso penalty is used in combination with non-linear least squares estimation of the drift in multivariate and high-dimensional diffusions. The result is sparse and computationally attractive models. Theoretical results on the generalization error are presented together with an algorithm for the non-linear lasso and the automatic tuning of the lasso penalization. The methods are available in the R package smd.

E432: **Non-parametric drift estimation for SDEs**
Presenter: **Yvo Pokern**, University College London, United Kingdom
A Gaussian prior measure is chosen in the space of drift functions by specifying its mean function and its precision operator as an appropriate differential operator. A Bayesian Gaussian conjugate analysis for the drift of one-dimensional non-linear diffusions is feasible using high-frequency data, by expressing the log-likelihood as a quadratic function of the drift, with sufficient statistics given by the local time process and the end points of the observed path. Computationally efficient posterior inference is carried out using a finite element method. This technology is then embedded in partially observed situations adopting a data augmentation approach whereby missing data paths and draws from the unknown functional are iteratively generated. Extensions to dimension two are presented and the methodology is illustrated using one and two dimensional examples from molecular dynamics, financial econometrics and movement ecology as well as high and low frequency observations.

E718: **Exact simulation of jump diffusions and unbiased estimation of irregular barrier crossing probabilities**
Presenter: **Murray Pollock**, University of Warwick, United Kingdom
Co-authors: Adam Johansen, Gareth Roberts
Recent work will be discussed which extends methodology for simulating finite dimensional representations of jump diffusion sample paths over finite intervals, without discretisation error (exactly), in such a way that the sample path can be restored at any finite collection of time points. We consider an application of the methodology developed which allows the simulation of upper and lower bounding processes which almost surely constrain (jump) diffusion sample paths to any specified tolerance. We demonstrate the efficacy of our approach by showing that with finite computation it is possible to determine whether or not sample paths cross various irregular barriers, simulate to any specified tolerance the first hitting time of the irregular barrier and simulate killed diffusion sample paths.

E262: **Simulation of diffusion bridges**
Presenter: **Frank van der Meulen**, Delft University of Technology, Netherlands
Co-authors: Moritz Schauer
Computing Bayesian estimators for diffusion processes based on discrete time observations can be dealt with by data-augmentation. Within this approach, the paths of a diffusion between the observations are treated as missing data. Draws from the posterior are obtained by employing a Gibbs sampler that targets the distribution of the missing data and parameter, conditional on the observed data. Especially in high dimensions, simulation of the missing data, which are diffusion bridges, can be difficult. We propose a method to deal with this problem using guided proposals generated by linear processes.

ES84 Room 349 STATISTICAL ANALYSIS OF NEUROIMAGING DATA Chair: Mattias Villani

E534: **An fMRI-based biomarker for physical pain**
Presenter: **Martin Lindquist**, Johns Hopkins University, United States
Persistent pain is central to many disorders, but can only be measured by self-report, which hampers diagnosis and treatment. The aim is to introduce a biomarker based on distributed patterns of functional Magnetic Resonance Imaging (fMRI) activity that a) accurately predicts pain experience and b) differentiates physical pain from other related experiences. It is illustrated using a number of validation studies that tested the biomarker's sensitivity and specificity for pain at the level of the individual person. Throughout it focuses on the statistical issues involved in developing the biomarker and outlines potential extensions moving forward.

E711: **Bayesian connectomics**
Presenter: **Marcel van Gerven**, Radboud University Nijmegen, Netherlands
In recent years, neuroscientists have become increasingly interested in the question of how cognitive function is mediated by distributed patterns of activity across spatially remote brain regions. Research focuses on elucidating which brain regions are physically connected with each other (structural connectivity) and to what extent these brain regions show covarying patterns of brain activity (functional connectivity). However, established approaches to whole-brain structural and functional connectivity estimation provide point estimates that do not properly reflect uncertainty in the estimates. We will present our recent work on Bayesian methods that do allow quantification of this uncertainty. We refer to this research line as Bayesian connectomics. The Bayesian connectomics framework makes use of generative models that explain observed neuroimaging data as

acquired by different imaging modalities. Bayesian connectomics provides an elegant theoretical framework where structural and functional connectivity can be jointly estimated using a Bayesian data fusion approach. We will end by discussing implications and by providing some pointers to future research directions.

E867: **Modelling nonlinear parametric responses in fMRI data**
Presenter: **William Penny**, University College London, United Kingdom
Co-authors: Sukhbinder Kumar
A model of fMRI data comprising a parametric neurodynamic component, a hemodynamic component and a BOLD observation model is described. The hemodynamic component corresponds to a previously established 'Balloon model' and is described using differential equations relating neuronal activation to blood flow, volume and deoxygenation. The BOLD observation model is a static nonlinear function that has been well studied in the literature and relates volume and deoxygenation to the BOLD signal. Our novel contribution is to employ a model of neuronal activity that allows one to infer parameters governing nonlinear parametric relationships among neuronal responses. We demonstrate the method using an event-related fMRI study of Repetition Suppression (RS) in the auditory system and use our method to estimate RS time constants along Heschl's Gyrus. Previous approaches to estimating parametric responses, or 'parametric modulations', in fMRI data have relied on linear models and so have been constrained to problems in which BOLD responses are linearly related to model parameters. Our new method is shown to present a significant advance.

E1194: **Spatial Bayesian point process modelling of neuroimaging meta-analysis data**
Presenter: **Thomas Nichols**, University of Warwick, United Kingdom
Co-authors: Jian Kang, Timothy Johnson
Functional magnetic resonance imaging (fMRI) is now over 20 years old, and the vast number of published fMRI studies are now frequently summarised with meta-analysis. Unfortunately, the only data widely available for neuroimaging meta-analysis are the [x,y,z] coordinates of peak activation foci. Most of the existing methods convert these coordinate data into images and then apply traditional imaging modelling and inference procedures. However, these methods are mass-univariate, make inference on individual voxels (or clusters) and require fixed tuning parameters. Such methods do not provide an interpretable fitted model, and cannot produce spatial confidence intervals on location of activation. Using a Bayesian hierarchical point process approach, modelling brain activation as a mixture of latent activation centers, and these activation centers are in turn modelled as off-spring from latent population centers. When our fully Bayesian model fit to multiple groups it is trivial to make predictions of the group label for new data and, through an importance sampling trick, leave-one-out cross validation accuracy for the sample studied.

ES88 Room Bedford GENERALIZED LINEAR MODELS AND BEYOND Chair: Achim Zeileis

E377: **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. The aim is the extension of the Beta regression model that can naturally accommodate zero and one observations avoiding 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.

E492: **Exposure-adjusted binary response models in a large portfolio of car insurance policies**
Presenter: **Linda Elin Walter**, University of Basel, Switzerland
The aim is to model the claim incidence in a Swiss motor third-party liability insurance portfolio by applying extensions of generalized linear models (GLM). The unique data set is characterized by a large number of contracts, different lengths of risk exposure and changes made to contracts during the insurance period. In order to investigate the relationship between reporting a claim and various covariates, an exposure-adjusted logit model is employed and, additionally, keep track of the changes of the covariates over one calendar year. This results in clustering effects with a highly skewed distribution of cluster sizes. Among other results, it is found that leasing a car is associated with the probability of reporting a claim, whereas the gender of the main driver shows no statistically significant effect. It can be shown that the benefit from keeping track of the changes in the contract is limited, whereas it is important to adjust for the exposure time. Furthermore, various robustness checks are conducted and alternatives provided - modified non-canonical link functions and weighted likelihood procedures - to correct for exposure in a binary response setting.

E642: **Beyond Poisson and negative binomial hurdle models**
Presenter: **Christian Kleiber**, Universitaet Basel, Switzerland
In many applications of count data regression, a large number of zero counts effectively precludes the use of the classical Poisson and negative binomial regression models that are conveniently discussed within the framework of generalized linear models (GLMs). One solution is to employ a hurdle model, which combines a binary component and a zero-truncated count component. Although hurdle models themselves are not GLMs, it is argued that systematic use of the GLM framework offers methodological as well as computational advantages. For example, the count part of a hurdle model is typically specified either as a (truncated) Poisson or a (truncated) negative binomial distribution. Both distributions belong to the large family of generalized power series distributions (GPSDs) and hence can be conveniently handled within the GLM setting. The aim is to explore certain non-classical parametric hurdle models given by different count distributions. For distributions that are not GPSDs the GLM framework can sometimes still be exploited to some extent.

E773: **Beyond binary GLMs: Exploring heteroskedastic probit and parametric link models**
Presenter: **Achim Zeileis**, Universitaet Innsbruck, Austria
The R package glmx (currently available from R-Forge) for estimation of generalized linear models (GLMs) with extra parameters is introduced and discussed. The motivation for development of the package was in particular to provide various kinds of binary response models with additional flexibility that are somewhat popular in the social sciences: e.g., heteroskedastic probit models or models with parametric link functions. Using the R implementation, these models are explored concerning parameter recovery, identifiability, usefulness in practice, and connections to other related models.

ES104 Room Jessel DEPENDENCE MODELS IN ENVIRONMENTAL SCIENCES Chair: Fabrizio Durante

E224: **Distortions of multivariate distribution functions and associated level curves: Applications in multivariate risk theory**
Presenter: **Elena Di Bernardino**, Universite Lyon, France
Co-authors: Didier Rulliere

A parametric model for multivariate distributions is proposed. The model is based on distortion functions, i.e. some transformations of a multivariate distribution which permit us to generate new families of multivariate distribution functions. We derive some properties of considered distortions. A suitable proximity indicator between level curves is introduced in order to evaluate the quality of candidate distortion parameters. Using this proximity indicator and properties of distorted level curves, we give a specific estimation procedure. The estimation algorithm is mainly relying on straightforward univariate optimizations, and we finally get parametric representations of both multivariate distribution functions and associated level curves. Our results are motivated by applications in multivariate risk theory. The methodology is illustrated on simulated and real examples. In particular, we focus on meteorological data in order to estimate the multivariate return period for the environmental event studied (eruption of geysers, rainfall etc etc). Several possible developments are in progress. For instance, the possibility to estimate non-parametrically the generator of an Archimedean copula using the notion of self-nested copula.

E237: **Non-parametric copulas for circular-linear and circular-circular data: An application to wind directions**
Presenter: **Michael Wiper**, Universidad Carlos III de Madrid, Spain
Co-authors: Jose Antonio Carnicero, Concepcion Ausin

A nonparametric method for estimating the dependence relationships between circular variables and other circular or linear variables using copulas is proposed. The method is based on the use of Bernstein copulas which are a very flexible class of non-parametric copulas which allows for the approximation of any kind of dependence structure, including non-symmetric relationships. In particular, we present a simple procedure to adapt Bernstein copulas to the circular framework and guarantee that the constructed bivariate distributions are strictly continuous. We provide two illustrative case studies, the first on the relation between wind direction and quantity of rainfall in the North of Spain and the second on the dependence between the wind directions in two nearby buoys at the Atlantic ocean.

E242: **Physically coherent probabilistic weather forecasts via ensemble copula coupling**
Presenter: **Roman Schefzik**, Heidelberg University, Germany
Co-authors: Thordis Thorarinsdottir, Tilmann Gneiting

State-of-the-art weather forecasts depend on ensemble prediction systems, which consist of multiple runs of dynamical numerical weather prediction models differing in the initial conditions and/or the parameterized numerical representation of the atmosphere. Statistical postprocessing of the ensemble forecasts can address biases and dispersion errors. However, current postprocessing approaches are mostly univariate and apply to a single weather quantity at a single location and for a single prediction horizon only. Such methods do not account for dependencies which are crucial in many applications. To address this, we propose a tool called ensemble copula coupling (ECC), in which the postprocessed forecast ensemble inherits the spatial, temporal and inter-variable dependence pattern of the unprocessed raw ensemble. Essentially, this is achieved by using the empirical copula of the raw ensemble to aggregate samples from predictive distributions obtained by univariate postprocessing techniques. We study several variants of ECC, discuss relationship to discrete copulas, and assess the predictive performance in an application to weather forecasts over Germany.

E375: **Multivariate return period calculation via survival functions**
Presenter: **Gianfausto Salvadori**, Universita Del Salento, Italy

The concept of return period is fundamental for the design and the assessment of many engineering works. In a multivariate framework, several approaches are available to its definition, each one yielding different solutions. A consistent theoretical framework for the calculation of return periods in a multidimensional environment is outlined, based on survival copulas and the corresponding survival Kendall's measures. The approach solves the problems raised in previous publications concerning the coherent foundation of the notion of return period in a multivariate setting. As an illustration, practical hydrological applications are presented.

Saturday 14.12.2013 11:20 - 13:00 Parallel Session D – CFE

CS01 Room B35 MULTIVARIATE GARCH MODELS WITH APPLICATIONS Chair: Niklas Ahlgren

C130: **Conditional correlation models of autoregressive conditional heteroskedasticity with nonstationary GARCH equations**
Presenter: **Timo Terasvirta**, Aarhus University, Denmark
Co-authors: Cristina Amado
The effects of careful modelling the long-run dynamics of the volatilities of stock market returns on the conditional correlation structure are investigated. To this end we allow the individual unconditional variances in Conditional Correlation GARCH models to change smoothly over time by incorporating a nonstationary component in the variance equations. The modelling technique to determine the parametric structure of this time-varying component is based on a sequence of specification Lagrange multiplier-type tests. The variance equations combine the long-run and the short-run dynamic behaviour of the volatilities. The structure of the conditional correlation matrix is assumed to be either time independent or to vary over time. We apply our model to pairs of seven daily stock returns belonging to the S&P 500 composite index and traded at the New York Stock Exchange. The results suggest that accounting for deterministic changes in the unconditional variances considerably improves the fit of the multivariate Conditional Correlation GARCH models to the data. The effect of careful specification of the variance equations on the estimated correlations is variable: in some cases rather small, in others more discernible. In addition, we find that portfolio volatility-timing strategies based on time-varying unconditional variances often outperform the unmodelled long-run variances strategy in the out-of-sample. As a by-product, we generalize news impact surfaces to the situation in which both the GARCH equations and the conditional correlations contain a deterministic component that is a function of time.

C105: **A Lagrange multiplier test for testing the adequacy of the CCC-GARCH model**
Presenter: **Paul Catani**, Hanken School of Economics, Finland
Co-authors: Timo Terasvirta, Meiqun Yin
A Lagrange multiplier test for testing the parametric structure of a constant conditional correlation generalized autoregressive conditional heteroskedasticity (CCC-GARCH) model is proposed. The test is based on decomposing the CCC-GARCH model multiplicatively into two components, one of which represents the null model, whereas the other one describes the misspecification. A simulation study shows that the test has good finite sample properties. We compare the test with other tests for misspecification of multivariate GARCH models. The test has high power against alternatives where the misspecification is in the GARCH parameters and is superior to other tests. The test has also power against misspecification in the conditional correlation structure.

C119: **A latent dynamic factor approach to forecasting multivariate stock market volatility**
Presenter: **Bastian Gribisch**, University of Cologne, Germany
A latent dynamic factor model for potentially high-dimensional realized covariance matrices of stock returns is proposed. The approach is based on the matrix logarithm and allows for flexible dynamic dependence patterns by combining common latent factors driven by HAR dynamics and idiosyncratic AR(1) factors. The model accounts for symmetry and positive definiteness of covariance matrices without imposing parametric restrictions. Simulated Bayesian parameter estimates as well as positive definite covariance forecasts are easily obtained using basic Markov Chain Monte Carlo (MCMC) methods. An empirical application to 5-dimensional and 30-dimensional realized covariance matrices of daily New York Stock Exchange (NYSE) stock returns shows that the model outperforms other approaches of the extant literature both in-sample and out-of-sample.

C104: **Multivariate finite-sample bootstrap tests for ARCH in vector autoregressive models**
Presenter: **Niklas Ahlgren**, Hanken School of Economics, Finland
Co-authors: Paul Catani
Multivariate finite-sample bootstrap tests for ARCH effects in the errors of vector autoregressive (VAR) models using Monte Carlo testing techniques are proposed. The tests under consideration are combined equation-by-equation Lagrange multiplier (LM) tests, multivariate LM tests and LM tests of constant error covariance matrix. The tests are based on standardised multivariate residuals. We use a parametric bootstrap to circumvent the problem that the test statistics in VAR models are not free of nuisance parameters under the null hypothesis. The tests are evaluated in simulation experiments and are found to have excellent size and power properties. The LM tests of constant error covariance matrix outperform the combined and multivariate tests in terms of power. The tests are applied to VAR models estimated on credit default swap (CDS) prices and Euribor interest rates data.

CS73 Room B33 MODELLING REGIME CHANGES II Chair: Willi Semmler

C1222: **Markov-switiching structural BVAR with Dynare**
Presenter: **Michel Juillard**, Bank of France, France
Dynare has a new module for Markov-Switching Bayesian structural VAR analysis. The aim is to present the syntax to represent Markov-switching structural BVAR models in Dynare and illustrate the functioning of the toolbox with some simple examples.

C1228: **Rationality in a switching environment**
Presenter: **Junior Maih**, Norges Bank, Norway
Evidence from both theory and the data is presented which suggests that the dynamics of various macroeconomic variables is subject to regime switches. It then presents new algorithms for solving rational expectations models subject to regime switches. In particular, unlike it is common, the switching probabilities can be endogenous. The performance of the algorithms is compared to previous ones. The new algorithms are implemented in a toolbox with the name RISE, which in addition to solving and simulating Markov Switching Rational Expectations models, also estimates those models.

C1164: **Overleveraging and regime change in the banking-macro link**
Presenter: **Willi Semmler**, New School for Social Research, United States
Co-authors: Stefan Mittnik
The implications of different types of financial–sector stress for real economic activity are investigated. The analysis is motivated by a theoretical banking–macro model, in which adverse asset–price movements and their impact on risk premia and credit spreads can induce instabilities in the banking sector. We employ a nonlinear, multi–regime vector autoregression (MRVAR) approach, in order to assess whether and how industrial production is affected by the individual risk indicators that form the Financial Stress Index (FSI) the IMF constructs for various countries. The impact is assessed for eight economies—the U.S., Canada, Japan and the UK, and for the four largest euro–zone economies, Germany, France, Italy, and Spain—, using Granger–causality and nonlinear impulse–response analysis. Our results strongly suggest that financial–sector stress exerts a

strong, nonlinear influence on economic growth, but that the different financial–risk factors affect economic activity rather differently across stress regimes and across countries.

C788: **Estimating the financial stress-output link using a vector smooth transition autoregressive model**
Presenter: **Frauke Schleer**, Centre for European Economic Research, Germany
Co-authors: Willi Semmler
The aim is to analyze the feedback mechanisms between economic downturns and financial stress and the role of the heterogeneity of the euro area countries based on a new, broader data set for extracting financial condition indices. For this purpose, we apply a non-linear Smooth Transition Vector Autoregressive model, an appropriate tool for investigating instabilities arising in the financial sector-output relation. This model has not been used within this literature, but may be more appropriate to model asymmetric dynamics since it allows for smooth regime changes. We show that the typically found negative output effect is not always present, particularly before the Lehman collapse, although we are in a model-defined high stress regime. Events leading to a strong economic breakdown are rather related to a financial cycle which has low frequency and hence, they occur rarely. The difficulty is to find out what kind of high stress regime leads to such negative consequences. One has to differentiate not only between high and low stress regimes but also to identify relevant (for the real economy) stress regimes.

CS83 Room G16 NONLINEAR TIME SERIES I Chair: Frederique Bec

C439: **Likelihood based inference in a dynamic mixture cointegrated VAR model**
Presenter: **Paolo Paruolo**, University of Insubria, Italy
Co-authors: Emil Nejstgaard, Anders Rahbek
Likelihood based inference is considered in a general class of regime-switching cointegrated vector error correction (VECM) models. This framework allows for epochs of non-stationary behavior, asymmetric error correction and switching error covariance. Consequences for inference are discussed when cointegration vectors are estimated and when a constant is included in the cointegration relations. These extensions provide additional flexibility to a broad class of regime switching vector error correction models in the literature.

C257: **Parameter identification in the logistic STAR model**
Presenter: **Emil Nejstgaard**, University of Copenhagen, Denmark
A simple reparameterization of the so-called 'speed of transition parameter' of the logistic smooth transition autoregressive (LSTAR) model is proposed. We show how the original parameterization of the speed of transition parameter is impractical and, as a consequence, estimation can result in an arbitrarily large and imprecise parameter estimate. The new parameterization highlights that a consequence of the well-known identification problem of the speed of transition parameter is that the threshold autoregression (TAR) is a limiting case of the LSTAR process. We show that information criteria can be used as a tool to choose between an LSTAR model and a TAR, a choice previously based primarily on economic theory. Finally, two empirical applications illustrate the usefulness of the reparameterization.

C209: **Frequentist evaluation of small DSGE models**
Presenter: **Luca Fanelli**, University of Bologna, Italy
Co-authors: Gunnar Bardsen
The aim is to propose a new evaluation approach of the class of small-scale 'hybrid' New Keynesian Dynamic Stochastic General Equilibrium (NK-DSGE) models typically used in monetary policy and business cycle analysis. The novelty of our method is that the empirical assessment of the NK-DSGE model is based on a conditional sequence of likelihood-based tests conducted in a Vector Autoregressive (VAR) system in which both the low and high frequency implications of the model are addressed in a coherent framework. The idea is that if the low frequency behaviour of the original time series of the model can be approximated by unit roots, stationarity must be imposed by removing the stochastic trends. This means that with respect to the original variables, the solution of the NK-DSGE model is a VAR that embodies a set of recoverable unit roots/ cointegration restrictions, in addition to the cross-equation restrictions implied by the rational expectations hypothesis. The procedure is based on the sequence 'LR1!LR2 !LR3', where LR1 is the cointegration rank test, LR2 the cointegration matrix test and LR3 the cross-equation restrictions test: LR2 is computed conditional on LR1 and LR3 is computed conditional on LR2. The type-I errors of the three tests are set consistently with a pre-fixed overall nominal significance level and the NK-DSGE model is not rejected if no rejection occurs. We investigate the empirical size properties of the proposed testing strategy by a Monte Carlo experiment and illustrate the usefulness of our approach by estimating a monetary business cycle NK-DSGE model using U.S. quarterly data.

C273: **The stochastic stationary root model**
Presenter: **Andreas Hetland**, University of Copenhagen, Denmark
The stochastic stationary root (SSR) time-series model is proposed and studied. This multivariate nonlinear state-space model decomposes the p-dimensional observation vector into r stationary, self-exciting directions, and $p-r$ nonstationary, random walk directions. This allows for e.g. a nonstationary homoskedastic level and stationary self-exciting spread between two variables, such as interest rates or asset prices. By applying sequential Monte Carlo methods, otherwise known as particle filters, we propose a part closed-form, part simulated filter that efficiently approximates the intractable model filter density and likelihood. By applying Markov chain Monte Carlo, the simulated likelihood is then used to conduct Bayesian inference on the model parameters. An application to the 10-year government bond rates in Germany and Greece in the period from January 2000 to May 2013 illustrates the usefulness of the SSR model.

CS14 Room B34 NEW DEVELOPMENTS IN TIME SERIES Chair: Richard Gerlach

C044: **Bayesian analysis of periodic unit roots**
Presenter: **Alexander Vosseler**, Institute for employment research IAB and University of Bamberg, Germany
A Bayesian testing approach for a periodic unit root in quarterly and monthly data is presented. Further, a Bayesian $F-$test is introduced to test for unit roots at (non)seasonal spectral frequencies. All procedures admit one structural break in the periodic trend function, where the occurrence of a break as well as the associated timing are treated as additional model parameters. Instead of resorting to a model selection approach by choosing one particular model specification for inference, a Bayesian model averaging (BMA) approach is proposed to capture the model uncertainty associated with a specific parameterization of the test regression. In order to analyze the properties of the presented Bayesian unit root tests, the respective power functions are computed. Overall the results indicate that the BMA periodic unit root test exhibits quite favorable power properties even in small samples, whereas the nonperiodic BMA unit root ($F-$) test has some power problems in small samples. In an empirical application the presented testing procedures are finally used to test for (non)seasonal forms of unemployment persistence among OECD countries using monthly unadjusted unemployment rates.

C060: **Statistical estimation for CAPM with long-memory dependence**
Presenter: **Tomoyuki Amano**, Wakayama University, Japan
Co-authors: Tsuyoshi Kato, Masanobu Taniguchi

The Capital Asset Pricing Model (CAPM) with time dimension is considered. By using time series analysis, we discuss the estimation of CAPM when market portfolio and the error process are long memory process and correlated with each other. We give a sufficient condition for the return of assets in the CAPM to be short-memory. In this setting, we propose a two-stage least squares estimator for the regression coefficient, and derive the asymptotic distribution. Some numerical studies are given. They show an interesting feature of this model.

C444: **Estimation of autocopulas**
Presenter: **Hiroaki Ogata**, Waseda University, Japan
The autocorrelation is a basic, widely used measure for describing serial dependence of a stationary process. However, if the process has a heavy-tailed distribution, and does not have a finite second order moment, the autocorrelation cannot be defined. One remedy for overcoming this drawback is to use the autocopulas. The problem of estimating the autocopulas is considered and some Monte Carlo studies are given. An illustration to real data is also given.

C602: **Bayesian semi-parametric expected shortfall forecasting incorporating intra-day range data**
Presenter: **Richard Gerlach**, University of Sydney, Australia
Co-authors: Cathy Chen
Bayesian methods have proven effective for quantile and tail conditional expectation estimation, including for financial Value at Risk (VaR) and Expected Shortfall (ES) forecasting. Intra-day range data has added efficiency and accuracy to return volatility and VaR forecasting. The class of conditional expectile models is extended to directly incorporate intra-day range data as an input to forecast ES. Adaptive Markov chain Monte Carlo sampling schemes are employed for estimation. The proposed models are favoured in an empirical study forecasting multiple financial return series: evidence of more accurate ES forecasting, compared to a range of competing methods, is found.

CS21 Room G15 MULTIVARIATE MODELS FOR FINANCIAL RISK ASSESSMENT Chair: Silvia Figini

C099: **Banks' default probabilities and credit rating**
Presenter: **Mario Maggi**, University of Pavia, Italy
Co-authors: Dean Fantazzini
A sample of large European banks, computing the default probabilities as their risk indicator, is analyzed. We compare the results obtained applying the structural KMV-Merton model, the recently proposed zero price probability (ZPP) model with the expected default frequency by rating agencies. We show the differences among the three approaches from the empirical point of view and we discuss their suitability during the last decade.

C516: **Graphical network modelling of cross-border systemic risk**
Presenter: **Paolo Giudici**, University of Pavia, Italy
Co-authors: Alessandro Spelta
The late-2000s financial crisis has stressed the need of understanding financial systems as networks of countries, where cross-border financial linkages play the fundamental role. Deriving macro (global) situations from micro (local) scenarios has been a recurrent topic in economics. The way to link the macro and the micro levels hinges on graph theory. The adoption of a network approach is recommended not only because of the proper emphasis on the financial interdependencies but also due to the possibility of describing how the structure of these interdependencies evolves in time. The aim is to present a Bayesian dynamic graphical model that provides an advanced solution to the problem of specifying a correct network model class, aimed at studying financial interdependences that may lead to cross-border systemic risk. The proposed model contributes to financial network modelling as: i) it provides a measure of distance based not only on marginal but also on partial correlations; ii) it provides a measure of variability of the connectedness measures, from which confidence intervals can be derived; iii) it provides a dynamic extension that is particularly suited for longitudinal, cyclic data. The proposal has been applied on a Bank of International Settlements database that contains the financial interflows between several national banking systems. The obtained results show that the proposed model provides, with a limited number of assumptions, a general tool for selecting the most relevant connection between national banking systems, that should be monitored to prevent cross-border contagion and systemic risk at the interbank level.

C658: **The extremal index for stochastic volatility models with state space representation**
Presenter: **Fabrizio Laurini**, University of Parma, Italy
When modelling financial time series is nowadays a standard goal to explore the extremal characteristics of the data, which have strong connection with risk management and prediction. To characterize the extremal behaviour of such a process a single key-parameter is required, known as the extremal index. Loosely, the extremal index represents the average clusters size of values which exceed a high-level threshold. Available results for simpler special cases, like the ARCH(1), cannot be extended to obtain the extremal index of more general processes. Recently, a successful approach was introduced to obtain the extremal index for the GARCH(1,1), which exploits the bivariate regular variation of GARCH(1,1) representation. Extending to the higher dimensional case for the GARCH(p,q) requires multivariate regular variation, hence such a generalization is not mathematically straightforward, and closed-form analytical expression for the extremal index is not usable for its numerical evaluation. Additionally, there are non-linear models, which belong to the parameter driven class of stochastic volatility, and often embedded into the state space representation, for which clustering of extreme values does not occur. More complicated state space models, with non-linear hidden volatility, are explored and their clustering behaviour is discussed. A simulation-based technique is introduced that turns out to be computationally efficient for the calculation of the extremal index, as it requires only the simulation of clusters of extreme values, and not of the full process.

C751: **Predicting intra–daily volume shares for VWAP–based trading strategies: A GAS approach**
Presenter: **Giampiero Gallo**, Universita di Firenze, Italy
Co-authors: Fabrizio Cipollini, Francesco Calvori
The Volume Weighted Average Price (VWAP) mixes volumes and prices at intra-daily intervals and is a benchmark measure frequently used to evaluate a trader's performance. Under suitable assumptions, splitting a daily order according to *ex ante* volume predictions is a good strategy to replicate the VWAP. To bypass possible problems generated by local trends in volumes, we propose a novel Generalized Autoregressive Score (GAS) model for predicting volume shares (relative to the daily total), inspired by the empirical regularities of the observed series (intra-daily periodicity pattern, residual serial dependence). An application to six NYSE tickers confirms the suitability of the model proposed in capturing the features of intra– daily dynamics of volume shares.

CS34 Room B18 EMPIRICAL DYNAMICS OF CREDIT MARKETS Chair: Florian Ielpo

C380: **Fiscal policies and credit regimes: A TVAR approach**
Presenter: **Andrea Roventini**, University of Verona, Italy
Co-authors: Tommaso Ferraresi, Giorgio Fagiolo
The aim is to investigate how the state of credit markets non-linearly affects the impact of fiscal policies. A Threshold Vector Autoregression (TVAR) model is estimated on U.S quarterly data for the period 1984-2010. The spread between BAA-rated corporate bond yield and 10-year treasury constant maturity rate is employed as a proxy for credit conditions. It is found that the response of output to fiscal policy shocks are stronger and more persistent when the economy is in a tight credit regime. The fiscal multipliers are abundantly and persistently higher than one when firms face increasing financing costs, whereas they are feebler and often lower than one in the normal credit regime. On the normative side, the results suggest policy makers to carefully plan fiscal policy measures according to the state of credit markets.

C1239: **Implied liquidity risk premium in the term structure of sovereign credit default swap and bond spreads**
Presenter: **Lara Cathcart**, Imperial College, United Kingdom
Co-authors: Saad Badaoui, Lina El Jahel
The focus is on the dynamic properties of the risk-neutral liquidity risk premium specific to the sovereign credit default swap (CDS) and bond markets. We show that liquidity risk has a non-trivial role and participates directly to the variation over time of the term structure of sovereign CDS and bond spreads for both the pre- and crisis periods. Secondly, our results indicate that the time-varying bond and CDS liquidity risk premium move in opposite directions which imply that when bond liquidity risk is high, CDS liquidity risk is low (and vice versa), which may in turn be consistent with the substitution effect between CDS and bond markets. Finally, our Granger causality analysis reveals that, although the magnitude of bond and CDS liquidity risk are substantially different, there is a strong liquidity information flow between the CDS and the bond markets, however no market seems to consistently lead the other.

C1189: **Sovereign bond risk premiums**
Presenter: **Engelbert Dockner**, Vienna University of Economics and Business, Austria
Co-authors: Manuel Mayer, Josef Zechner
Credit risk has become an important factor driving government bond returns. We introduce an asset pricing model which exploits information contained in both forward interest rates and forward CDS spreads. Our empirical analysis covers euro-zone countries with German government bonds as credit risk-free assets. We construct a market factor from the first three principal components of the German forward curve as well as a common and a country-specific credit factor from the principal components of the forward CDS curves. We find that predictability of risk premiums of sovereign euro-zone bonds improves substantially if the market factor is augmented by a common and an orthogonal country-specific credit factor. While the common credit factor is significant for most countries in the sample, the countryspecific factor is significant mainly for peripheral euro-zone countries. Finally, we find that during the current crisis period, market and credit risk premiums of government bonds are negative over long subintervals, a finding that we attribute to the presence of financial repression in euro-zone countries.

C606: **Sector spillovers in credit market**
Presenter: **Florian Ielpo**, BCV Asset Management, Switzerland
Co-authors: Peng-Han Fan
Credit spreads have a tendency to drop and burst together along financial crises. Despite their tendency to covariate, significant differences of behavior can be observed between sectors when it comes to volatility. It is shown how the aggregated spreads of credit sectors contribute differently to volatility bursts and drops over the past 12 years. The banking sector has been a net contributor to credit spreads's volatility over the period when consumer cyclicals have had a tendency to receive volatility shocks rather than giving them. A Markov Switching VAR estimated over the period confirms this feature.

CS52 Room B20 RECENT ADVANCES IN FINANCIAL ECONOMETRICS Chair: Christian Pigorsch

C687: **Fast algorithms for estimating the probability of informed trading from tick data**
Presenter: **Stefan Kloessner**, Saarland University, Germany
Co-authors: Andreas Recktenwald
The methodology to estimate the probability of informed trading from tick data is redefined. In our model, market participants (informed or uninformed) trade equities with a market maker. However, informed traders only enter the market on days with relevant information: on good news they buy otherwise they sell. Buys and sells are modelled as latent point processes with waiting times following Weibull distributions whose scales come from autoregressive conditional duration (ACD) models, allowing for time-varying intensities. There are three (latent) types of trading days (no news, good news, bad news) whose evolving over time is modelled by a Hidden Markov Model (HMM). Using four years (2007 - 2010) of tick data for 12 firms of the automobile sector about 20 parameters are estimated by ML. Because of the huge number of durations fast algorithms are required as well as parallel computing on the nowadays widely available multicore processors. Empirically we find that with the HMM approach a clear classification for most trading days is possible and that the usually used exponential distribution is not flexible enough.

C706: **Futures pricing in electricity markets based on stable CARMA spot models**
Presenter: **Gernot Mueller**, University of Oldenburg, Germany
Co-authors: Fred Benth, Claudia Klueppelberg, Linda Vos
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 modeling and estimation of electricity prices are an important issue for the risk management of electricity markets. We consider a new model for the electricity spot price dynamics, which is able to capture seasonality, low-frequency dynamics and the extreme spikes in the market. Instead of the usual purely deterministic trend we introduce a non-stationary independent increments process for the low-frequency dynamics, and model the large fluctuations by a non-Gaussian stable CARMA process. We suggest an estimation procedure, where we fit the non-stationary trend using futures data with long time until delivery. The procedure also involves the estimation of the empirical and theoretical risk premiums.We apply the procedure to base load and peak load data from the German electricity exchange EEX.

C842: **Analysis of discrete dependent variable models with spatial correlation**
Presenter: **Jan Vogler**, University of Cologne, Germany
Co-authors: Roman Liesenfeld, Jean-Francois Richard
ML estimation is considered for a broad class of parameter-driven models for discrete dependent variables with spatial correlation. Under this class of models, which includes spatial discrete choice models, spatial Tobit models and spatial count data models, the dependent variable is driven by a latent stochastic state variable which is specified as a linear spatial regression model. The likelihood is a high-dimensional integral whose dimension depends on the sample size. For its evaluation we propose to use efficient importance sampling (EIS). The specific spatial EIS implementation

we develop exploits the sparsity of the precision (or covariance) matrix of the errors in the reduced-form state equation typically encountered in spatial settings, which keeps numerically accurate EIS likelihood evaluation computationally feasible even for large sample sizes. The proposed ML approach based upon spatial EIS is illustrated with estimation of a spatial probit for US presidential voting decisions and spatial count data models (Poisson and Negbin) for firm location choices.

C1247: **Predicting large covariance matrices using a characteristic-based conditionally heteroskedastic factor model**
Presenter: **Uta Pigorsch**, University of Mannheim, Germany
Co-authors: Enno Mammen

The aim is to predict the covariance matrix of excess returns, y_{it} with $i = 1,\ldots,n$ and $t = 1,\ldots,T$, of a large number of financial assets ($n >> T$). To this end, we extend a previous characteristic-based factor model by introducing conditionally heteroskedastic factors. The model is given by$y_{it} = f_{ut} + \sum_{j=1}^{J} g_j(X_{ji}) f_{jt} + \varepsilon_{it}$ where f_{ut} denotes the return of a factor that is relevant to all assets and has unit factor loadings. The factor returns $f_{jt},\ j = 1,\ldots,J$, instead, are related to J asset characteristics via the characteristic-based betas $g_j(X_{ji})$, where X_{ji} denotes the time-invariant and continuously distributed characteristic j of asset i. The function g is a smooth and time-invariant function of the jth characteristic. The asset-specific returns, ε_{it}, are assumed to have zero mean and to be cross-sectionally and temporally uncorrelated. We incorporate conditional heteroskedasticity into this model by assuming that the factors follow a multivariate GARCH process. We provide some asymptotic results and present an empirical application of the model using data on US securities.

CS54 Room B29 FORECAST ACCURACY Chair: Pilar Poncela

C460: **Forecasting inflation: On the measure of uncertainty to be used**
Presenter: **Eva Senra**, Universidad de Alcala, Spain
Co-authors: Pilar Poncela

Uncertainty may play a key role in forecasting inflation. The Survey of Professional Forecasters (SPF) provides point and density forecasts of inflation. The SPF has been widely used to forecast point inflation through combination of forecasts and to extract measures of uncertainty. There are several measures for uncertainty that can be classified as direct (based on individual uncertainty), indirect (based on disagreement among forecasters or alternatively time series models on the original inflation data) as well as ex-post (based on forecast errors). The aim is to seek an alternative procedure that aggregates individual density forecasts to obtain a new aggregated density forecast. From there, the alternative measure of uncertainty is extracted. Finally, this aggregate measure is compared with other coming from factor models on individual uncertainties.

C730: **Multivariate sparse partial least squares for macroeconomic forecasting**
Presenter: **Julieta Fuentes**, Universidad Carlos III de Madrid, Spain
Co-authors: Julio Rodriguez

In macroeconomics, finance and many other fields the joint prediction of a set of related variables constitutes an important issue. The available information contained in large data sets can lead to more accurate univariate and multivariate forecasts. The usefulness of sparse partial least squares techniques for multivariate forecast is investigated. This approach takes into account the target forecast variables to construct -from an informative subset of predictors- the jointly factors. We use the Stock and Watson database in order to compare the forecast performance of multivariate and univariate spls approach and other widely used factor methods in macroeconomic forecasting.

C733: **Dynamic factor models in forecasting inflation in Poland**
Presenter: **Grzegorz Szafranski**, National Bank of Poland and University of Lodz, Poland

In a framework of dynamic factor analysis the researchers study correlations among hundreds of variables searching for the common forces driving macroeconomic processes. With a similar motivation the central bankers construct factor-based leading indicators to forecast inflation or output in real time. Simple methods for estimating factor scores like a principal component analysis perform quite well in large datasets. Other, quasi maximum likelihood methods based on state space formulation and Kalman filter need more attention, because they are naturally dedicated to deal with important empirical problems like unsynchronous data releases and mixing the data frequency. From the practical perspective also the data preparation, the pre-selection of variables and the calculation of their contributions to forecasts are also of a big importance. In a Monte Carlo simulation we analyze short-sample properties of different methods for estimating common factors under: model misspecification, factor block structure, missing data and non-normality. In the empirical part we analyse different strategies for estimating and forecasting with dynamic factor models. In a real-time exercise for the Polish economy we compare their abilities to predict inflation and its uncertainty. We also describe changes in informational content of consecutive data vintages.

C805: **Inflation uncertainty revisited: A proposal for robust measurement**
Presenter: **Steffen Henzel**, Ifo Institute, Germany
Co-authors: Christian Grimme, Elisabeth Wieland

Any measure of unobserved inflation uncertainty relies on specific assumptions which are most likely not fulfilled completely. This calls into question whether an individual measure delivers a reliable signal. To reduce idiosyncratic measurement error, we propose using common information contained in different measures derived from survey data, a variety of forecast models, and volatility models. We show that all measures are driven by a common component which constitutes an indicator for inflation uncertainty. Moreover, our results suggest that using one individual disagreement measure only may be misleading particularly during turbulent times. Finally, we study the Friedman-Ball hypothesis. Using the indicator, it turns out that higher inflation is followed by higher uncertainty. By contrast, we obtain contradictory results for the individual measures. We also document that, after an inflationary shock, uncertainty decreases in the first two months which is traceable to the energy component in CPI inflation.

CS67 Room B36 MODELING UNIVARIATE AND MULTIVARIATE VOLATILITY Chair: Christos Savva

C1014: **Estimation and empirical performance of non-scalar dynamic conditional correlation models**
Presenter: **Lyudmila Grigoryeva**, Universite de Franche-Comte, France
Co-authors: Luc Bauwens, Juan-Pablo Ortega

The Dynamic Conditional Correlation (DCC) model is widely used for the description of the dynamics of time-varying asset correlations. Despite the convenience of the representation of the conditional covariances through univariate GARCH models and the conditional correlations using the GARCH standardized residuals, the two-stages non-scalar DCC system estimator is difficult to handle in practice. The quadratic dependence of the number of parameters on the dimension of the problem together with the nonlinear constraints make the estimation process complicated. We focus on the Bregman-proximal trust-region method that was already successfully used to handle the VEC prescription and we extend it to the class of DCC models with different types of parametrizations: scalar DCC, Hadamard DCC, rank deficient DCC with arbitrary rank, and a new type of Almon polynomial-parametrized models (Almon DCC, Almon shuffle DCC, Almon PC DCC) that we introduce. For each of these cases we provide a general formulation of the estimation problem using the Bregman-proximal method to handle the constraints and show the efficiency of

the proposed optimization method. Moreover, we compare the empirical out-of-sample performance of different types of estimated models with the standard O-GARCH and EWMA prescriptions in the forecasting of equity-based conditional covariances and correlations.

C560: **Dynamic cointegration and diversification benefits among the Greater China, the UK and the US stock markets**
Presenter: **Francesco Guidi**, University of Greenwich, United Kingdom
Co-authors: Christos Savva

The aim is to investigate the level of integration among the Greater China (i.e Hong Kong, mainland China and Taiwan), the UK and US stock markets in the last ten years. Using both static and dynamics cointegration approaches, it is found that the level of integration has gone through intermittent periods. The dynamic comovements among these markets are then modeled. The results show that there are only one way spillover effects; from the US and UK to TWSE and Hang Seng markets, whilst there are not any effects at any direction for the mainland China stock market. Secondly, positive, low and insignificant conditional correlations among UK and all Greater China markets and US with the Mainland China market are observed. This finding indicates that diversification opportunities exist, since the low correlation allows investors to use these markets in their hedging strategies. Hence a third question addressed is whether portfolio diversification benefits exist during the period that encompasses the recent global financial crisis. Using different measures of risk such as standard deviation as well as lower partial moments, evidence is found that diversification benefits did exist from October 2008 to November 2011.

C151: **Multivariate FIAPARCH modeling with dynamic correlation analysis of financial markets with structural breaks**
Presenter: **Stavroula Yfanti**, Brunel University, United Kingdom
Co-authors: Menelaos Karanasos, Michail Karoglou

The Multivariate model which combines an AR(1) mean specification with cross effects, the Fractionally Integrated GARCH with the Asymmetric Power ARCH and the Dynamic Conditional Correlation, is applied. We analyse the applicability of the AR-FIAPARCH-DCC model that allows in the mean specification for cross effects between the dependent variables of each bi(tri)variate specification and in the variance for long memory, power, leverage and time-varying correlations on eight national stock market indices daily returns from 1988 to 2010, taking into account the structural breaks of each time series linked to two financial crises, the Asian (1997) and the recent Global (2007-08) crisis. We find significant cross effects as well as long range volatility dependence, asymmetric volatility response to positive and negative shocks and the power of returns that best fits to the volatility pattern. The time-varying conditional correlations are found to be highly persistent. The main finding of the DCC model analysis is the higher dynamic correlations of the stock markets after a crisis event, which means increased contagion effects between the markets. The fact that during the crisis the conditional correlations remain on a high level can indicate a continuous herding behavior during these periods of increased market volatility. Finally, it is remarkable that during the recent Global financial crisis the correlations remain in a much higher level than during the Asian financial crisis.

C390: **Risk-return trade-off for European stock markets**
Presenter: **Christos Savva**, Cyprus University of Technology, Cyprus
Co-authors: Nektarios Aslanidis, Charlotte Christiansen

Dynamic factor models with macro-finance predictors are adopted to revisit the intertemporal risk-return relation in five large European stock markets. Country-specific, Euro-area, and global factors are identified to determine the conditional moments of returns considering the role of higher-order monents as additional measures of risk. The preferred combination of factors varies across countries. In the linear regression, there is a strong but negative relation between conditional returns and conditional volatility. A Markov switching model describes the risk-return trade-off better. A number of variables have explanatory power for the regime of the European stock markets.

CS75 Room B30 FINANCIAL VOLATILITY AND COVARIANCE MODELLING Chair: Genaro Sucarrat

C325: **Testing for nonlinearity in covariances**
Presenter: **Bilel Sanhaji**, Aix Marseille University, France

Two Lagrange multiplier tests for nonlinearity in covariances are proposed. The null hypothesis is the scalar BEKK model in which covolatilities of time series are driven by a linear function of its own lags and lagged squared innovations. The alternative hypothesis is an extension of that model in which covolatilities are modelled by a nonlinear function in the lagged squared innovations. The nonlinearity is represented by an exponential and a logistic transition function. We define the asymptotic properties of the scalar BEKK and determine the moment conditions of the test statistics. We investigate the size and the power of these tests through Monte Carlo experiments, and we show empirical illustrations.

C808: **Stochastic leverage models via iterated filtering**
Presenter: **Carles Breto**, Universidad Carlos III de Madrid, Spain

Novel models are proposed for leverage effects found in financial return series where leverage is assumed to be stochastic instead of deterministic and constant as it is often assumed. To evaluate the value of such models, we maximize their likelihood and study the behavior of the stochastic leverage implied by them. To do this, we take advantage of a plug-and-play approach to inference: iterated filtering, which we implement via particle filters. This plug-and-play property allows us to explore new models by simply writing computer code for the new model equations, without the need of model-specific work like analytical derivations, approximations or algorithmic refinements.

C793: **Gaussian QML estimation of the log-GARCH model via the ARMA representation: The asymptotic variance-covariance matrix**
Presenter: **Steffen Groenneberg**, BI Norwegian Business School, Norway
Co-authors: Genaro Sucarrat

The aim is to derive the full joint variance-covariance matrix of the Gaussian Quasi Maximum Likelihood Estimator (QMLE) of the log-GARCH model via the ARMA representation, and to verify the result in Monte Carlo simulations.

C641: **An exponential chi-squared QMLE for log-GARCH models via the ARMA representation**
Presenter: **Genaro Sucarrat**, BI Norwegian Business School, Norway
Co-authors: Christian Francq

Estimation of log-GARCH models via the ARMA representation is attractive because zero errors can be handled in a satisfactory manner. We propose an exponential Chi-squared QMLE for log-GARCH models via the ARMA representation. The advantage of the estimator is that it corresponds to the theoretically and empirically important case where the conditional error of the log-GARCH model is normal. The consistency and asymptotic normality of the estimator are proved, and the results are verified on Monte Carlo simulations. An empirical application illustrates the estimator in practice.

Saturday 14.12.2013 14:30 - 16:10 Parallel Session E – ERCIM

ESI03 Room Beveridge ROBUST PENALIZED METHODS Chair: Elvezio Ronchetti

E478: **High-dimensional robust statistics through additive over-parameterization with shrinkage**
Presenter: **Yiyuan She**, Florida State University, United States

Recently high dimensional data applications have attracted a lot of attention from statisticians. Most high dimensional approaches are however not robust in the sense that they are extremely sensitive to anomaly points. The classical robustification through designing a robust loss function or observation weights has some inherent difficulties to meet the needs of modern high-dimensional models (possibly non-Gaussian), both in theory and in computation. A new robustification methodology is proposed via additive over-parameterization with shrinkage. It applies to multivariate generalized linear models in possibly high dimensions, and includes the conventional psi estimators and least trimmed squares as particular instances. Rather than making an asymptotic study assuming that the sample size goes to infinity, it is possible to prove finite-sample theoretical results and reveal the minimax optimal rates. Moreover, a family of prediction accuracy based information criteria is proposed that can achieve the optimal rates for the purpose of parameter tuning. Real data examples are shown to demonstrate the efficiency and efficacy of the computational algorithms in high dimensional robust estimation.

E646: **The robust nonnegative garrote for generalized linear and additive models**
Presenter: **Marco Avella-Medina**, University of Geneva, Switzerland
Co-authors: Elvezio Ronchetti

Generalized linear and additive models are popular statistical methods for modelling continuous and discrete data both parametrically and nonparametrically. In this general framework the problem of variable selection through penalized methods is considered by focusing on resistance issues in the presence of outlying data and other deviations from the stochastic assumptions. In particular, a robust version of the nonnegative garrote is proposed, the asymptotic normality of the resulting estimator and its consistency for variable selection are shown. This extends the available theory from linear models to generalized linear and additive models as well as from L2-based estimation to robust estimation. Finally, the finite sample performance of the method is illustrated by a simulation study in a Poisson regression setting and its robustness properties are discussed.

E752: **Robust regularized estimation in multivariate statistics**
Presenter: **Christophe Croux**, Leuven, Belgium

An overview of some recent work on sparse and robust estimation is given. In regression analysis, the Lasso estimator can be shown to have a zero breakdown point. This means that one single outlier can drive the estimate up to infinity. By using a least trimmed objective function combined with a penalty term, one obtains the sparse least trimmed squares estimator. The sparse least trimmed squares estimator is implemented in R, and can deal with high dimensional data sets. Robust sparse methods for (inverse) covariance matrix estimation, principal components analysis, and factor analysis will also be considered. Robust selection of the regularization parameter as well as computational speed turns out to be important.

ES02 Room B30 STATISTICAL ALGORITHMS AND SOFTWARE IN R Chair: Andreas Alfons

E489: **Model selection for graphical models using the focused information criterion: R-based software implementations**
Presenter: **Eugen Pircalabelu**, Katholieke Universiteit Leuven, Belgium
Co-authors: Gerda Claeskens, Lourens Waldorp

A new method for model selection for Gaussian directed acyclic graphs (DAG) and undirected graphs (UG), is constructed to have good mean squared error properties, for both low and high dimensional settings. The method is based on the focused information criterion (FIC) which allows for selecting individual models, tailored to a specific purpose. Under a local misspecification, which assumes that the true model is in a 'neighborhood' of the least complex model one is willing to assume, a *focus parameter* is defined, i.e. $\mu(\theta_0, \gamma_0 + \delta\sqrt{n})$ as a function of the parameters of the model density, and potentially of a user-specified vector of covariate values, for any particular node in the graph $G(E,V)$. Subsequent steps involve specifying a collection of models and a (un-)penalized objective function which is optimized for parameters corresponding to the prespecified models. The FIC estimates MSE($\hat{\mu}_S$) for each of a collection of models S, and selects the model with the smallest value. The focus directs the selection and different focuses may lead to different selected models. Thus, one can obtain better selected models in terms of MSE, than obtained from a global model search. For this application of FIC for structure learning in graphs, the focus is the expected value of a variable, reflecting interest in discovering a topology of the graph that produces a low MSE for this focus. Two software implementations are presented: a hill-climbing and a nodewise approach. The current implementations make use of the R facilities for using C++ linear algebra libraries and also multicore parallellization.

E504: **Robust estimation of linear mixed effects models with the robustlmm package in R**
Presenter: **Manuel Koller**, University of Bern, Switzerland

A robust method of estimating linear mixed effects models is presented. So far, most other robust estimation methods had to make strong assumptions on the structure of the data, e.g., they required the data to be hierarchical or allowed only one variance component. By splitting random components and treating contamination at its source - the random components - a robust method is developed that requires only a few assumptions to be applicable. There are no assumptions about the structure of the data aside from the regular assumptions made for linear mixed effects models. The method can estimate data with crossed random effects, i.e., non-hierarchical data. It is implemented in a software package in R called "robustlmm".

E622: **Highly parallelized algorithms for exploratory analysis of large functional MRI datasets in R**
Presenter: **Roland Boubela**, Medical University of Vienna, Austria
Co-authors: Klaudius Kalcher, Wolfgang Huf, Peter Filzmoser, Ewald Moser

In the analysis of functional magnetic resonance imaging (fMRI) data exploratory approaches are well established. Independent Component Analysis (ICA) is the most popular method to analyze resting-state data, where no specific a priori information about activities in the brain is required. With recent advances in fMRI acquisition techniques allowing for higher temporal and spatial resolution, the amount of data per measurement increases up to a factor of 100. To be able to perform exploratory analyses on such large datasets, a graphical processing unit (GPU) accelerated pipeline for temporal and spatial ICA is presented. The pipeline is incorporated in a framework for fMRI data analysis in R handling out-of-memory operations and parallelization with R's facilities. The first step of ICA, the dimensionality reduction and pre-whitening via principal component analysis (PCA), is performed via irlba: Fast partial SVD by implicitly-restarted Lanczos bidiagonalization and parallelized on multiple GPUs, which might also be of general interest for the computation of PCA on very large datasets. The actual computation of ICA is done with the fastICA algorithm.

E719: **Dual scaling methods for response style detection and purging in R**
Presenter: **Pieter Schoonees**, Erasmus University Rotterdam, Netherlands

The occurrence of response styles in multivariate categorical data, such as those collected from surveys, can severely bias statistical analyses. A new method, derived from dual scaling, can be used to detect and purge response styles from such data. The method and its optimization criterion are introduced and an alternating nonnegative least squares algorithm is devised to optimize it. In this talk a new R package, `constrDS`, which implements this method, is introduced and illustrated with empirical applications. The strength and weaknesses of the method for purging response styles from data are discussed and illustrated.

ES06 Room B36 SOFTWARE FOR DESIGN AND ANALYSIS OF EXPERIMENTS WITH MANY FACTORS Chair: Rosemary A. Bailey

E066: **Desmond Patterson and the design key**
Presenter: **Rosemary Bailey**, University of St Andrews and Queen Mary University of London, United Kingdom

H. Desmond Patterson, who died in March 2013, introduced the design key in a paper in 1965. It is a way of constructing designs for experiments with many factors. The algorithm will be explained and some of the early work on it will be discussed. Various pieces of software that implement it will be mentioned. Two recent applications will be considered: multiphase experiments and strip-block experiments.

E124: **Automatic and flexible generation of regular factorial designs and its implementation in R**
Presenter: **Herve Monod**, INRA, France

Regular factorial designs are a very useful class of orthogonal designs based on precise confounding relationships between factorial effects. Algorithms for their construction, using the designkey concept, were developed as early as 1965 and some of them implemented in statistical packages. Based on a previous framework, we present a generalised version of the designkey algorithm, allowing for symmetric and asymmetric designs and for fractional as well as split-plot or criss-cross designs. We give an overview of the algorithm and of its implementation as a R package called planor.

E623: **An algorithm for constructing a mixed model using the experimental design**
Presenter: **Marion Chatfield**, University of Southampton GlaxoSmithKline, United Kingdom
Co-authors: Simon Bate

In many areas of scientific research complex experimental designs with many factors are now routinely used. With the advent of mixed model algorithms, implemented in many statistical software packages, the analysis of data generated from such experiments has become more accessible to scientists. However, failing to correctly identify the experimental design used can lead to incorrect model selection and misleading inferences. A procedure will be described that identifies the structure of the experimental design and, given the randomisation, generates a maximal mixed model. This model is determined before the experiment is conducted and provides a starting point for the final statistical analysis. The whole process will be illustrated using a generalisation of the Hasse diagram called the Terms Relationships diagram. The model selection process is simplified through use of a systematic procedure and most parts of the algorithm have been implemented in a program written in R.

E640: **A tool for automating the analysis of variance of orthogonal designs**
Presenter: **Heiko Grossmann**, Queen Mary, University of London, UK

The analysis of variance for experiments with complex block and treatment structures usually requires that F tests of treatment effects need to be performed in different strata. When the underlying design is orthogonal, the correct anova can be easily obtained by using Hasse diagrams. It is shown how the Hasse diagrams and the anova table can be automatically derived from the experimental design. A software tool is presented and the use of the program is illustrated with several examples.

ES18 Room G16 STATISTICAL METHODS FOR NON-PRECISE DATA I Chair: Ana B. Ramos-Guajardo

E420: **Lasso estimation of a multiple linear regression model for interval data**
Presenter: **Marta Garcia-Barzana**, University of Oviedo, Spain
Co-authors: Ana Colubi, Erricos Kontoghiorghes

A multiple linear regression model which accounts for cross-relationships of mids and spreads has been recently introduced. The least squares estimation of the parameters can be done by transforming a quadratic optimization problem with inequality constraints into a linear complementary problem and solving it by means of Lemke's algorithm. However, not all the cross-relationships are always relevant. Thus, alternative estimators via Lasso are proposed. A comparative study is developed.

E647: **Adjusting the analytic hierarchy process for fuzziness: A risk assessment context**
Presenter: **Arnold Shapiro**, Penn State University, United States

The Analytic Hierarchy Process (AHP) is a theory of measurement through pair-wise comparisons that relies on judgment to derive priority scales. During its implementation, one constructs hierarchies, then makes judgments or performs measurements on pairs of elements with respect to a criterion to derive preference scales, which are then synthesized throughout the structure to select the best alternative. Depending on the decision-making context, however, problems can arise because decision-making often is hindered by data limitations and ambiguities, such as incomplete or unreliable data, and vague and subjective information owing to a reliance on human experts and their communication of linguistic variables. Since fuzzy logic (FL) is an effective tool in such circumstances, there has been considerable research based on adjusting the AHP for fuzziness (AHP-FL). The aim is to discuss the state of the art with respect to AHP-FL in a risk assessment context. It begins with a brief overview of the AHP and its limitations when confronted with a fuzzy environment. This is followed with a discussion of FL modifications of the AHP. The work ends with a commentary on the findings.

E1146: **Effects of discretization on the construction of confidence sets for geometric problems**
Presenter: **Raffaello Seri**, University of Insubria, Italy

A method to obtain a confidence set for the Aumann mean of a random closed set has been recently proposed. This requires the computation of a quantile of the supremum of a Gaussian stochastic process defined on the sphere. Unfortunately this distribution is not known but an approximation to it can be obtained discretizing the process along a set of directions. We analyze the error involved in this discretization using some chaining inequalities for the distribution of the supremum of a Gaussian process. Several other examples of application of the same result are considered.

E1159: **Testing the equality of two-paired distributions through the fuzzy representation**
Presenter: **Angela Blanco-Fernandez**, University of Oviedo, Spain
Co-authors: Ana Ramos-Guajardo, Ana Colubi

The fuzzy representation of a real-valued random variable is used to visualize or/and characterize its distribution through fuzzy sets. Various fuzzy representations, which differ on their usefulness to explore or test on different characteristics of the real distribution, are available. New inferential strategies for the equality of two-paired distributions based on bootstrap techniques are presented. A particular definition of fuzzy representation

for the real random variables is employed, whose expected value fully characterizes the distribution of the variables. The technique is theoretically and empirically analyzed. The consistency of the bootstrap testing procedure is shown. Besides, some simulation studies and real applications are developed in order to compare its empirical behaviour with asymptotic methods as well as with usual nonparametric alternatives.

ES53 Room B20 ON SOME ISSUES IN CLUSTER ANALYSIS Chair: Geoff McLachlan

E041: **Measurement of quality in cluster analysis**
Presenter: **Christian Hennig**, UCL, United Kingdom
There is much work on benchmarking in supervised classification, where quality can be measured through misclassification probabilities. In unsupervised classification (cluster analysis), the measurement of quality is much more difficult, because in reality there is no true class label. Furthermore, there is no guarantee that the 'true' classification in benchmark data sets from supervised classification or simulated data is unique. There can be a number of different reasonable clusterings of the same data, depending on the research aim. The use of statistics for the assessment of clustering quality that can be computed from classified data without making reference to 'true clusters' will be discussed. Such statistics have traditionally been called 'cluster validation indexes'. Most of the traditional statistics (e.g., average silhouette width) try to balance various aspects of a clustering (such as within-cluster homogeneity and between-cluster separation). In order to characterize advantages and disadvantages of a clustering, different aspects of cluster quality can be formalized separately. This can also be used to explain misclassification rates in cases where true classifications exist, as function of the features of these clusterings.

E373: **A study comparing different (mixture and non-mixture) approaches to clustering high-dimensional data**
Presenter: **Cristina Tortora**, University of Guelph, Canada
Co-authors: Paul McNicholas
Cluster analysis aims at identifying homogeneous groups of units. Mixture models are receiving renewed attention within the literature. However, their performance can deteriorate as the number of variables increases. Factor clustering, or subspace clustering, refers to methods that integrate variable extraction and clustering into one iterative algorithm, which optimizes a unique criterion to find the groups in a reduced subspace. An alternative to the model-based paradigm is introduced: distance-based clustering. A comparison of some methods that belong to distance- and model-based families is presented. Their overall performance is compared using simulated data. The methods used in the comparison, based on the mixture of Gaussian, are: mixture of factor analyzers, and mixture of common factor analyzers, mixture of parsimonious Gaussian mixture models, high dimensional Gaussian mixture models, and discriminative latent mixture models. The comparison includes a non-Gaussian based method, mixture of generalized hyperbolic factor analyzers and a distance based method, and factor probabilistic distance clustering.

E401: **Mixtures of dynamic factor models**
Presenter: **Jia Chen**, University of York, United Kingdom
Co-authors: Geoffrey McLachlan
A dynamic factor model is proposed, where the common factors are a mixture of vector autoregressive processes. Due to its dynamic nature, the model can be used for modelling multivariate time series data. The flexibility and the capacity of the model to capture various shapes of distributions imply its applicability in a wide range of fields, in particular, in finance, where data are often found to have multi-modes, heavy-tails, and leptokurtosis which can not be properly described by commonly used distributions such as normal and student-t distributions. An EM algorithm is developed for the estimation of model parameters with the conditional expectations in the E-step of the iterative procedure calculated via the Kalman filter. The presented model and methods can also be used for time series clustering as an alternative to distance measure based methods. The performance of the methods in time series clustering analysis is shown through simulation experiments. An application to stock returns data is also presented.

E740: **New advances related to the application of trimming and restrictions in model based clustering**
Presenter: **Agustin Mayo-Iscar**, Universidad de Valladolid, Spain
Co-authors: Luis Angel Garcia Escudero, Alfonso Gordaliza, Carlos Matran
An approach based on trimming and restrictions in clustering is considered. Trimming methodology provides robustness properties to the estimators and the implementation of restrictions allows us to get a well-posed estimation problem and to avoid spurious solutions. We show the computational feasibility and the statistical properties for several different proposals that incorporate these add-ins.

ES55 Room B29 ESTIMATION AND TESTING IN CONDITIONALLY HETEROSCEDASTIC TIME SERIES Chair: Simos Meintanis

E301: **Tests for time series of counts based on the probability generating function**
Presenter: **Marie Huskova**, Charles University, Czech Republic
Co-authors: Sarka Hudecova, Simos Meintanis
Testing procedures are proposed for the hypothesis that a given set of discrete observations may be formulated as a particular time series of counts with a specific conditional law. The proposed test statistics are based on empirical probability generating functions. The focus is on integer autoregression and Poisson autoregression. Results on asymptotic behavior both under the null hypothesis as well as under alternatives will be presented. Finite sample properties based on a simulation study will accompany theoretical results.

E413: **Fourier inference for stochastic volatility models with heavy-tailed innovations**
Presenter: **Bruno Ebner**, Karlsruhe Institute of Technology, Germany
Co-authors: Bernhard Klar, Simos Meintanis
Estimation of stochastic volatility models is considered, which are driven by a heavy-tailed innovation distribution. Exploiting the simple structure of the characteristic function of suitably transformed observations we propose an estimator which minimizes a weighted L2-type distance between the theoretical characteristic function of these observations and an empirical counterpart. A related goodness-of-fit test is also proposed. Monte-Carlo results are presented which facilitate comparison with existing methods. The procedures are applied to real data from the financial markets.

E615: **Fourier-type inference for GARCH models with heavy-tailed innovations**
Presenter: **Violetta Dalla**, National and Kapodistrian University of Athens, Greece
Co-authors: Yannis Bassiakos, Simos Meintanis
Estimation of GARCH models is considered with innovations following a heavy-tail and possibly asymmetric distribution. Although the method is fairly general and applies to GARCH models with arbitrary innovation distribution, it is considered as a special instance of the stable Paretian, the variance gamma and the normal inverse Gaussian distribution. Exploiting the simple structure of the characteristic function of the observations driven by these models, minimum distance estimation based on the empirical characteristic function of properly standardized errors is proposed. A related goodness-of-fit test is also proposed. The finite-sample results presented facilitate comparison with existing methods, while the procedures are also applied to real data from the financial markets.

E304: **Test for spherical symmetry in multivariate GARCH models**
Presenter: **Simos Meintanis**, University of Athens, Greece
Co-authors: Christian Francq
Tests for spherical symmetry of the innovation distribution are proposed in multivariate GARCH models. The new tests are of Kolmogorov-Smirnov and Cramer-von Mises-type and make use of the common geometry underlying the characteristic function of any spherically symmetric distribution. The asymptotic properties of the test statistics are investigated under general conditions. Since the finite-sample and the asymptotic null distribution depend on the unknown distribution of the Euclidean norm of the innovations, a conditional Monte Carlo procedure is used to actually carry out the tests.

ES91 Room B35 COMPUTATIONAL METHODS IN BIOMEDICAL RESEARCH Chair: Joyce Niland

E849: **Time-to-event endpoint definition and required sample size**
Presenter: **Martina Mittlboeck**, Medical University of Vienna, Austria
Co-authors: Harald Heinzl
Sample size calculation for prospective clinical studies should ensure that existing relevant group differences in the primary endpoint can be detected with high probability (power). In studies with a time-to-event endpoint several sensible endpoint definitions might be considered. E.g. in cancer studies, overall-survival, event-free survival or recurrence-free survival are potential candidates among others. After sample size calculation the endpoint definition with the highest number of expected events will often be chosen to keep the anticipated study duration at a minimum. Consequences of several time-to-event endpoints for the required study sample size, especially with respect to disease and non-disease related events, will be compared and discussed. A further focus will be put on the ensuing data analyses and result interpretations, in particular on the comparison and interpretation of hazard ratios when the proportions of non-disease related events differ between studies.

E881: **Implementing broader notions of optimality for statistical planning of biomedical studies**
Presenter: **Jeffrey Longmate**, City of Hope, United States
The notion of optimality embedded in common software tools often limits the design of biomedical studies. Some techniques and tools are presented that depart from minimizing sample size subject to a power constraint, and exhibit multiple designs so that trade-offs can be evaluated. These tools deal with discreteness, multiple stages, multiple objectives, and prospectively specified models for retrospective designs.

E889: **How to find consolidated top elements in omics ranked lists: Novel computational and graphical techniques**
Presenter: **Michael Georg Schimek**, Medical University of Graz, Austria
Co-authors: Eva Budinska, Vendula Svendova
For almost a decade an increasing interest in the statistics of ranked lists can be seen, primarily stimulated by new biotechnologies such as highthroughput analysis. Typically, such lists comprise tens of thousands of elements (e.g. genes in comparative genomics). However, only a comparably small subset of k top-ranked elements is informative. These elements are characterized by a strong overlap of their rank positions when they are assessed several times (e.g. when biological experiments are replicated). The statistical task is to identify these top-k elements. Now there are some approaches to this decision problem, but all are limited to just two complete ranked lists (i.e. no missing assignments). Here we propose novel computational and graphical techniques that allow us to extend and apply recent theoretical findings to the typical omics-situation of multiple ranked lists. The handling of three to ten very long lists with missing assignments is a common requirement. Given this scenario, the new techniques are compared and evaluated in extensive simulations.

E1028: **Effect of length biased sampling and overdiagnosis on survival from screen-detected disease**
Presenter: **Karen Kafadar**, Indiana University, United States
Co-authors: Philip C. Prorok
Screen-detected cases form a length-biased sample among all cases of disease, since longer sojourn times afford more opportunity to be screen-detected. Typically the length-biased (LB) sampled variable is observed. In screening, the LB sampled variable (sojourns) are not observed, but their subsequent clinical durations are. We investigate the effect of length-biased sampling of sojourn times on the distribution of the observed clinical durations. When they are positively correlated, the mean clinical duration can be substantially inflated, even in the absence of any benefit on survival from screening. We discuss this effect and other biases, such as overdiagnosis, both in screening and in other scenarios where length-biased sampling occurs.

ES63 Room B34 DIMENSION REDUCTION Chair: Hannu Oja

E159: **Estimation of dimension reduction subspace in PCA**
Presenter: **Eero Liski**, University of Tampere, Finland
Co-authors: Klaus Nordhausen, Hannu Oja, David Tyler
Dimensionality is major concern in high dimensional data analysis to which dimension reduction often offers a solution. In dimension reduction it is hoped that a projection of the data to a subspace of lower dimension would contain all the interesting structures within the data. In practice, the true subspace and its dimension are not known and have to be estimated. In our approach based on principal component analysis (PCA), we assume that the observations come from an elliptical distribution and the eigenspace corresponding to few largest eigenvalues contains all the information. The smallest eigenvalues are assumed to be equal and the corresponding vector of principal variables is then spherically distributed. There exists a number of tests for subsphericity or equality of eigenvalues, often with strict assumptions. We consider tests that are based on robust scatter matrices with more relaxed assumptions. The tests are then used in the estimation of the subspace dimension and finally, the resulting subspace estimates are considered. The theory is illustrated with simulation studies.

E562: **Separation of uncorrelated stationary time series using SOBI**
Presenter: **Klaus Nordhausen**, University of Tampere, Finland
Co-authors: Jari Miettinen, Hannu Oja, Sara Taskinen
Assume that the observed p time series are linear combinations of p latent uncorrelated weakly stationary time series. The problem is then to find an estimate for an unmixing matrix that transforms the observed time series back to uncorrelated time series. The so called SOBI (Second Order Blind Identification) estimate achieves this by jointly diagonalizing the covariance matrix and several autocovariance matrices with varying time lags. Different ways on how to compute SOBI and the statistical properties of the corresponding estimates are discussed.

E739: **Sufficient dimension reduction for dynamic factor models**
Presenter: **Efstathia Bura**, The George Washington University, United States
Co-authors: Alessandro Barbarino
Factor models are useful for summarizing a large dataset with few underlying factors and have been used in building time series forecasting models of economic variables. When the objective is to forecast a target variable y with a large dataset of predictors X, the construction of the summary of

the Xs must be driven by how informative on y it is. Most existing literature has tried to blend the task of reducing the dimension of the predictors X with the only partially related task to forecast y using dynamic factor models. The procedure contains two potentially unrelated steps: 1) first summarize X using PCA or some variant, then 2) use the extracted factors to forecast y. We offer an alternative and potentially more attractive alternative: 1) summarize X as it relates to y, using Sufficient Dimension Reduction methodology such as Principal Fitted Components, 2) estimate the forecasting equation.

E899: **Robust estimators in Gaussian graphical models**
Presenter: **Daniel Vogel**, Ruhr Universitaet Bochum, Germany
Co-authors: David Tyler
The classical statistical theory of Gaussian graphical models is likelihood-based and highly non-robust, with the sample covariance matrix being its key ingredient. Our aim is to robustify the statistical analysis. We study robust scatter estimators that obey the zero pattern in the inverse which is imposed by a given graph. In particular, we compare two ways of constructing M-estimators in this situation: one can either start from an unconstrained, affine equivariant M-estimator and apply the Gaussian methodology (the plug-in method) or directly consider M-estimators (i.e. the argmax of a suitable criterion function) in a Gaussian graphical model. We call the latter approach graphical M-estimation. We show that both approaches are asymptotically equivalent and derive the asymptotic distribution of the estimators. However, the finite sample properties of both types of M-estimators may be substantially different and are generally favorable for the graphical M-estimators. They turn out to be more efficient and more robust in finite samples, and fewer observations are necessary for the estimator to be computable. On the other hand, graphical M-estimators hold considerably more theoretical challenges than the plug-in M-estimators. The latter are a combination of two well developed techniques: classical graphical modeling and affine equivariant scatter estimation.

ES90 Room G15 PREDICTIVE LEARNING Chair: Roberta Siciliano

E631: **Combining clustering of preference rankings with unfolding and vector models**
Presenter: **Sonia Amodio**, University of Naples Federico II, Italy
Co-authors: Antonio D'Ambrosio, Willem Heiser
As a multitude of real data problems involve preference ranking data, in the last decades analysis of rankings has become increasingly important and interesting in several fields such as behavioral sciences, machine learning and decision making. A combination of cluster analysis of preference ranking data with vector models and unfolding technique is presented. Vector models and unfolding techniques are used to analyze preference ranking data and are based on badness of fit functions. Both methods map the preferences in a joint low-dimensional space. For the clustering part the K-Medians Cluster Component Analysis method is used. Real data sets are analyzed to illustrate the proposed approach.

E691: **Model-free probability distance clustering of time series**
Presenter: **Gianluca Frasso**, Universite de Liege, Belgium
Co-authors: Antonio D'Ambrosio, Roberta Siciliano
A new model-free clustering approach for time series is introduced. We combine a probability distance clustering method with a penalized smoothing procedure of time series. We take advantage from the class of probability distance clustering. It allows for probabilistic allocation of series to groups and it is based on the principle that probability and distance are inversely related. On the other hand the class of penalized smoothing procedure allows us to efficiently separate the signal and the noise component of a generic time series. Using the smoothed version of the raw data as an input of our procedure makes the clustering algorithm more flexible due to the fact that we do not need to specify a parametric model to summarize the observed dynamics. Our approach can be easily generalized for the analysis of multivariate time series. We evaluate the performances of the proposed clustering procedure using different distance measures. Comparisons of our proposal with already known procedures are provided. We discuss the applicability of the proposed clustering method to analyse general classes of time series emphasizing its fuzzy nature.

E829: **The latent class distance association model**
Presenter: **Mark de Rooij**, Leiden University, Netherlands
Categorical variables are often collected in the social and behavioral sciences. Analysis of such data typically uses loglinear or (multinomial) logit models. With many variables, however, such statistical analyses fail because parameter estimates do not exist due to the sparseness of the data. Dimension reduction techniques such as the RC(M) association model or the Distance Association model might be employed to deal with the problem. The corresponding visualization of these models is, however, troublesome due to the large number of points, cluttering the display. A latent class distance association model is proposed to simultaneously cluster the profiles and reduce the dimension of a loglinear model. The methodology will be illustrated using an empirical data set from the Dutch Parliamentary Election studies.

E1191: **Sequential automatic search of subsets of classifiers in multiclass classification**
Presenter: **Claudio Conversano**, University of Cagliari, Italy
Co-authors: Francesco Mola
Multiclass Learning (ML) requires a classifier to discriminate instances (objects) among several classes of an outcome (response) variable. Most of the proposed methods for ML do not consider that analyzing complex datasets requires the results to be easily interpretable. We refer to the Sequential Automatic Search of Subset of Classifiers (SASSC) algorithm as an approach able to find the right compromise between knowledge extraction and good prediction. SASSC is an iterative algorithm that works by building a taxonomy of classes in an ascendant manner: this is done by the solution of a multiclass problem obtained by decomposing it into several r-nary problems ($r >> 2$) in an agglomerative way. We consider the use of different classification methods as base classifiers and evaluate the performance of SASSC with respect to alternative classes aggregation criteria which allow us to compare either the predictive performance or the interpretation issues related to the use of each set of classifiers.

ES85 Room B18 DIRECTIONAL STATISTICS I Chair: Christophe Ley

E363: **Parallel transport on manifolds**
Presenter: **John Kent**, University of Leeds, United Kingdom
For many purposes the simplest non-Euclidean manifold is the usual sphere in three-dimensional space. Statistical interest is generally focused just on the location of points on the sphere, but sometimes can include tangent directions at these points as well. For example, for a point lying at the north pole, a tangent direction is specified by a line of longitude. Parallel transport provides the mathematical machinery needed to connect tangent planes at different points. The relevance of a tangent direction can arise in a variety of statistical contexts, including (i) as part of the data such as an edgel in shape analysis, (ii) as a distributional component of variability analogous to the orientation of the ellipse of a covariance matrix in multivariate analysis or (iii) as a direction of systematic variation in a regression analysis.

E196: **Probabilistic models of protein structure: From theory to applications**
Presenter: **Thomas Hamelryck**, University of Copenhagen, Denmark

The so-called protein folding problem is the loose designation for an amalgam of closely related, unsolved problems that include protein structure prediction, protein design and the simulation of the protein folding process. We adopt a unique probabilistic approach to modeling bio-molecular structures, based on graphical models and directional statistics. Notably, we develop a generative probabilistic model of protein structure in full atomic detail. We will give an overview of how rigorous probabilistic models of something as complicated as a protein's atomic structure can be formulated, focusing on the use of graphical models and directional statistics to model angular degrees of freedom. We will also discuss the reference ratio method, a novel statistical method of general interest that can be used to 'glue' different probabilistic models of protein structure together in a consistent way. This method also sheds new light on the potential of mean force, widely used in protein structure prediction, and which was up to now poorly understood and justified. Finally, we will describe some applications, including probabilistic protein structure prediction and the inference of protein structure from experimental data.

E094: **Optimal tests for circular reflective symmetry about the median direction**
Presenter: **Thomas Verdebout**, University Lille Nord de France, France
Co-authors: Christophe Ley

Optimal tests for reflective circular symmetry about a fixed median direction are proposed. The distributions against which optimality is achieved are the so-called k-sine-skewed distributions. We first show that sequences of k-sine-skewed models are locally and asymptotically normal in the vicinity of reflective symmetry. Following the Le Cam methodology, we then construct optimal (in the maximin sense) parametric tests for reflective symmetry, which we render semi-parametric by a studentization argument. These asymptotically distribution-free tests happen to be uniformly optimal, under any reference density, and are moreover of a very simple and intuitive form. They furthermore exhibit nice small sample properties, as we show through a Monte Carlo simulation study.

E524: **Inference on circular orders with application to cell cycle gene expression data**
Presenter: **Sandra Barragan**, University of Valladolid, Spain
Co-authors: Cristina Rueda, Miguel A. Fernandez, Shyamal D. Peddada

In recent years there has been considerable interest in drawing inferences regarding order relationships among angular parameters. In particular, in biology the interest is to understand genes participating in cell cycle across multiple species and whether they are functionally conserved. The time to peak expression, known as phase angle, of such a gene can be mapped onto a unit circle. Biologists are not only interested in estimating the phase angles but in determining the relative order of expression of various genes. The final aim is to know whether the order of peak expression among cell cycle genes is conserved evolutionarily across species. These questions are challenging due to large variability among studies and to the circular nature of the data. A methodology to find the underlying circular order in a population is presented. We also propose a solution for the problem of testing equality of circular orders among two or more populations. Unbalanced samples and differences in distributions are taken into consideration. The proposed methodology is illustrated by analyzing data sets from three species: Schizosaccharomyces Pombe, Schizosaccharomyces Cerevisiae and Humans. As a result a set of genes is presented where the circular order is conserved across these three species.

ES123 Room B33 BAYESIAN SEMI- AND NONPARAMETRIC MODELLING II Chair: Li Ma

E785: **An optimal transport based theory of hierarchical Bayesian inference**
Presenter: **Long Nguyen**, University of Michigan, United States

Hierarchical Bayesian models present a powerful tool in statistics and machine learning. Statistical dependence can be easily expressed via latent variables, which may also be objects of inference. In a hierarchical model, the latent hierarchy enables the "borrowing of strength" between different data sets through shared parameters in the latent hierarchy. We will discuss an asymptotic theory for hierarchical model-based inference, taking a view that places latent variables at the center of the inference. By analyzing posterior concentration behaviors of the latent hierarchy that arise in a number of Bayesian nonparametric models, including the Dirichlet process mixture and the hierarchical Dirichlet processes, we show how to quantify in a precise sense the benefits of borrowing strength in a hierarchical model. We will also discuss the roles of transportation distances in the theory.

E790: **Bayesian semi-parametric estimation of the long-memory parameter under FEXP-priors**
Presenter: **Willem Kruijer**, Wageningen University, Netherlands
Co-authors: Judith Rousseau

For a Gaussian time series with long-memory behavior, we use the FEXP-model for semi-parametric estimation of the long-memory parameter d. The true spectral density f_o is assumed to have long-memory parameter d_o and a FEXP-expansion of Sobolev-regularity $\beta > 1$. We prove that when k follows a Poisson or geometric prior, or a sieve prior increasing at the rate $n^{1/(1+2\beta)}$, d converges to d_o at a suboptimal rate. When the sieve prior increases at the rate $n^{1/(2\beta)}$ however, the minimax rate is almost obtained. Our results can be seen as a Bayesian equivalent of the result which Moulines and Soulier obtained for some frequentist estimators.

E234: **Confidence in nonparametric Bayesian credible sets**
Presenter: **Botond Szabo**, Eindhoven University of Technology, Netherlands
Co-authors: Aad van der Vaart, Harry van Zanten

Adaptive techniques for nonparametric estimation have been widely studied and many rate-adaptive results have been provided for a variety of statistical problems. However, an adaptive estimator without any knowledge of its uncertainty is rather uninformative. In the Bayesian framework credible sets are intended to visualize the remaining uncertainty of the estimator. The coverage of the credible sets describes to what extent credible sets can be viewed as a quantification of uncertainty of the estimator from a frequentist point of view. Consider the problem of constructing Bayesian based confidence sets that are adaptive in L^2-loss over a continuous scale of Sobolev classes ($\beta \in [B, 2B]$, for some fixed $B > 0$) in the Gaussian White noise model. It is shown that both the hierarchical Bayes and marginal likelihood empirical Bayes approaches lead to credible sets with asymptotic coverage zero for certain oddly behaving functions. Then a new empirical Bayes method based on risk estimation is introduced, which provides uniform and adaptive confidence sets over the collection of Sobolev classes.

E184: **Posterior consistency of nonparametric location-scale mixtures for multivariate density estimation**
Presenter: **Pierpaolo de Blasi**, University of Torino and Collegio Carlo Alberto, Italy
Co-authors: Antonio Canale

Multivariate density estimation represents one of the most successful applications of Bayesian Nonparametrics. In particular, Dirichlet process mixtures of normal kernels are the golden standard for density estimation and their asymptotic properties have been studied extensively, especially in the univariate case. However, a gap between practitioners and the current theoretical literature is present. So far, posterior asymptotic results in the multivariate case are available only for location mixtures of Gaussian kernels with independent prior on the common covariance matrix, while in practice as well as from a conceptual point of view a location-scale mixture is often preferable. We address posterior consistency for such general mixture models by adapting a convergence rate result which combines the usual low-entropy, high-mass sieve approach with a suitable

summability condition. Specifically, we establish consistency for Dirichlet process mixtures of Gaussian kernels with various prior specifications on the covariance matrix including priors that parsimoniously model the covariance via a sparse factor model.

EP01 Room Macmillan POSTER SESSION I Chair: Ayse Ulgen

E162: **Socio-economic, demographic and health behaviour heterogeneity as determinants of self-medication**
Presenter: **Sofia Vale**, ISCTE-IUL, Portugal
Co-authors: Francisco Camoes

Self-medication is defined as the choice by individuals of medicines which are available without prescription to treat their self-diagnosed conditions. It is recognized for having benefits and risks. The benefits of self-medication are identified with greater independence of the patient, the improvement of his/her general condition and a reduction in government expenditures with healthcare. There are, however, several non-negligible risks of self-medication such as incorrect diagnosis, incorrect administration and over dosage or over duration of the treatments and delay in the search for indispensable medical advice. This is translated into benefits and costs for society, and is consequently an important piece of public health and public policy debate. Self-medication is frequently associated with health self-assessment and therefore it can be explained by individual practices towards risk, being susceptible of reporting heterogeneity. In this context, we use Portuguese micro data to identify determinants of self-medication. Using data from 2005 National Survey on Health we relate self-medication with socio-economic and demographic status but also with adverse health behaviour. Among our explanatory variables in a discrete dependent variables model, we include income, education, civil status, the incidence of chronic diseases, age, gender, the prevalence of risky health behaviour (smoking, drinking alcohol and being obese) and the habits of medical prevention. Our expected results are an important relevance of income and education, the individual attitude towards risky health behaviour, and significant gender-differences.

E308: **Structural variable selection via nested spike and slab priors**
Presenter: **Tso-Jung Yen**, Academia Sinica, Taiwan
Co-authors: Yu-Min Yen

Grouped variable selection problems are selected. We propose a specified prior, called the nested spike and slab prior, to model collective behavior of regression coefficients. At the group level, the nested spike and slab prior puts positive mass on the probability that the l_2-norm of the grouped coefficients is equal to zero. At the individual level, it assigns each coefficient a spike and slab prior. To estimate the model, we adopt a maximum a posteriori approach. The estimation problem involves an approximate objective function modified with the majorization-minimization technique. Simulation studies show that the proposed estimator performs relatively well when both the true and redundant covariates are included in the same group. Asymptotic analysis further shows that the proposed estimator can achieve a smaller l_2 estimation error if groups that include the true covariates do not include too many redundant covariates. In addition, given some regular conditions hold, the proposed estimator can achieve model selection consistency regardless of the irrepresentable-type conditions.

E407: **Global hypothesis test to compare the likelihood ratios of multiple binary tests with ignorable missing data**
Presenter: **Ana Eugenia Marin**, University of Granada, Spain
Co-authors: Jose Antonio Roldan

Positive and negative likelihood ratios are parameters to assess the accuracy of a binary diagnostic test, and only depend on the sensitivity and the specificity of diagnostic test. The likelihood ratios quantify the increase in terms of knowledge of the disease presence through the application of the diagnostic test. In the presence of partial disease verification the comparison of the likelihood ratios of two or more binary diagnostic tests in relation to the same gold standard cannot be made excluding those individuals not verified with the gold standard, since the results obtained will be affected by verification bias. A global hypothesis test, based on chi-squared distribution, is studied to simultaneously compare the likelihood ratios of more than two binary diagnostic tests when in the presence of partial disease verification the missing data mechanism is ignorable.

E736: **Plain MCMC forecasts and data augmentation imputed data for item non-response for endogenously allocated categories**
Presenter: **Maria Odejar**, Kansas State University, USA

Bayesian Markov chain Monte Carlo (MCMC) algorithms for a supervised learning process are developed to provide estimates and inference about missing data by forecasting without including them in the parameter estimation and straight data augmentation that treats generated data as part of the whole sample. The MCMC algorithms are applied to bivariate ordered threshold model and nested mixed probit model. The models are estimated also with and without incorporating random utility maximization. Parameter estimates are compared based on sum of log conditional predictive ordinate goodness-of-fit criterion of the holdout sample data, and predictive ability of the model in and out of the estimation sample based on apparent error rate of misclassification of the estimation data set and the actual error rate of misclassification of the holdout sample. Results show that inclusion of the non-ignorable missing responses through data augmentation is not detrimental to the quality of parameter estimates, as reflected in their standard deviations, sum of predictive ordinate goodness-of-fit criterion and predictive ability of the model in and out of the estimation sample, as well as in the mixing and convergence properties of the Markov chains generated by the MCMC Bayesian algorithm

E801: **Scaling human choice behaviour using Bayesian multidimensional scaling model**
Presenter: **Tomoya Okubo**, The National Center for University Entarance Examinations, Japan

A multidimensional model that enables us to model inconsistent human choice behaviour is discussed. Generally, multidimensional models assume a fixed preference (or dominance) order for a subject or group during paired comparison sessions, but this is unable to describe intransitive judgements. However, it has been reported that some respondents systematically violate their transitive judgement in multi-attribute choice situations. Following these influential researches, a number of theoretical and empirical studies on the intransitivity of preference judgements have been presented. Our proposing model, the Bayesian mixed-preference model is an extended MDS model that allows a subject or group to have multiple preference orders, and permits subjects to change their preference (dominance) order when judging paired comparisons. The proposed model can be considered as an extension of the latent class MDS models, and Bayesian solution of the model is also provided in this presentation. Further, in this presentation, we will provide some results based on the proposed model.

E870: **Using inverse probability weighting and generalized linear models to estimate cost with time censored data**
Presenter: **Patricia Ziegelmann**, Federal University of Rio grande do Sul, Brazil

The collection of cost data attached to randomized clinical trials has become a common practice with the increasing use of economic evaluation for decision making. As a result, cost data are observed longitudinally with informative censoring, since there is a positive correlation between costs accumulated to, for example, death and costs accumulated to censoring. The inverse probability weighted (IPW) least squares regression is one of the methodologies proposed to analyse this kind of data. However, common assumptions of linear regression such as normality and homoscedasticity are usually violated, since the distribution of health care costs can be heavily right skewed. We explore the use of generalized linear regressions (with IPW) to estimate mean total cost as it is a more flexible model to fit cost data. We assess the accuracy and precision of the estimators through a set of numerical simulations where different censoring scenarios, including or not dependency between censoring distribution and covariates, are incorporated.

E893: **Optimality in small-sample inference for models with correlated observations**

Presenter: **M. Jesus Rivas-Lopez**, University of Salamanca, Spain

Co-authors: Juan Manuel Rodriguez-Diaz, Sandra Martin-Chaves

The precision of parameter estimators in the Gaussian linear model is traditionally accounted by the variance-covariance matrix of the asymptotic distribution. However, under different conditions this measure can underestimate the true variance, specially for small samples. Traditionally, optimal design theory pays attention to this variance through its relationship with the model's information matrix, but it seems convenient that, at least in some cases, this deviation should be taken into account and the optimality criteria adapted in order to get the best designs for the actual variance structure. Different examples will be studied, and comparison of the designs obtained using the new approach with those computed employing the traditional one will be performed.

Saturday 14.12.2013 14:30 - 16:10 Parallel Session E – CFE

CS88 Room Montague STATISTICAL METHODS AND APPLICATIONS IN HEALTH AND FINANCE Chair: Luisa Cutillo

C106: **Regulating prostitution: Theory and evidence from Italy**
Presenter: **Francesco Russo**, University of Naples Federico II, Italy
Co-authors: Giovanni Immordino

The aim is to build an equilibrium model of prostitution where clients and sex workers choose to demand and supply sex under three legal regimes: prohibition, regulation and laissez-faire. The key feature is the endogenous evolution of the risk as a consequence of policy changes. We calibrate the model to empirical evidence from Italy and then compare the effect of different policies on the equilibrium quantity of prostitution and on the harm associated with it. A prohibition regime that makes it illegal to buy sex, but not to sell it is more effective than the opposite regime in reducing quantity. Taxes are one inducement to go illegal and prevent some of the less risky individuals from joining the market, leaving it smaller but riskier. A licensing system that prevents some infected individuals from joining the legal market reduces the risk and is therefore associated with an increase in quantity. While prohibition is preferable to minimize quantity, regulation is best to minimize harm.

C645: **A factor analysis for the European economy**
Presenter: **Albina Orlando**, Italian National Research Council CNR, Italy
Co-authors: Maria Francesca Carfora, Luisa Cutillo

A dynamic factor model is presented to estimate and forecast the rate of growth of the European economy in the very short term following a previous methodology. The indicators selected for model estimation are business and consumer surveys. As well known the Directorate General for Economic and Financial Affairs DG ECFIN conducts regular harmonised surveys for different sectors of the economies in the European Union and in the applicant countries. They are addressed to representatives of the industry (manufacturing), the services, retail trade and construction sectors, as well as to consumers. These surveys allow comparisons among different countries' business cycles and have become an indispensable tool for monitoring the evolution of the EU and the euro area economies, as well as monitoring developments in the applicant countries. The model is finally integrated with a Markov switching model to estimate the probabilities of expansion and recession of the European business cycle phases.

C509: **Patient and gene clustering using NetSel on time course expression data**
Presenter: **Annamaria Carissimo**, Fondazione Telethon - TIGEM, Italy
Co-authors: Luisa Cutillo

An algorithm for time series gene expression analysis is presented that perform clustering of patients on the basis of gene clusters comparison. In this light, the published Network Selection method (NetSel) is adapted to time course gene expression data. NetSel is an algorithm inspired by graph theory that, given a heterogeneous group of rankings, discovers the different communities of homogeneous rankings. In the proposed approach, for each patient, genes are treated as ranked lists across time points. NetSel is then applied to cluster genes with similar temporal profiles. As second step, patients are grouped according to gene partitions similarity calculated using the Variation of Information criterion. In order to highlight the strength of the proposal, an application is shown both on simulated and on a biological dataset. Experimental results on simulated data show that the method can significantly outperform existing related methods. Furthermore, the outcome on the real biological dataset demonstrates that the algorithm can correctly detect patients with the same infection or disease.

C458: **Clustering of count data using a negative binomial model**
Presenter: **Luisa Cutillo**, University of Naples Parthenope, Italy
Co-authors: Annamaria Carissimo

Counts are a special kind of data that take only non-negative integer values. They arise in many contexts and generally in event-counting applications. In economics, examples include the number of applicants for a job, or the number of labour strikes during a year. Recently in biology high-throughput sequencing, such as ribonucleic acid sequencing (RNA-seq) and chromatin immunoprecipitation sequencing (ChIP-seq) analyses, enables various features of organisms to be compared through tag counts. Such data cannot be modelled adequately by means of standard regression analysis. This is mainly due to the necessity of accounting for discreteness and overdispersion. The main difficulty to overcome would be the unavailability of normality assumption based techniques. The Poisson distribution is the most popular distribution for count data, yet it is constrained by its equi-dispersion assumption. Hence a more compelling model for over-dispersed count data is the Negative Binomial distribution whose parameters are uniquely determined by mean and variance. The aim is to propose an approach for clustering count data using a dissimilarity measure that is based upon the Poisson model. The performances of the approach are shown in a simulation study and on publicly available RNA-seq datasets.

CS38 Room Senate TIME-SERIES ECONOMETRICS Chair: Robert Kunst

C137: **LM and Wald tests for changes in the memory and level of a time series**
Presenter: **Heiko Rachinger**, University of Vienna, Austria
Co-authors: Juan Jose Dolado, Carlos Velasco

Lagrange Multiplier (LM) and Wald tests for breaks in the memory and the level of a time series are proposed. The contribution is to consider both types of breaks simultaneously to solve a potential confounding problem between long-memory changes and breaks in the level and in persistence. Identifying different sources of breaks is relevant for several reasons, such as improved forecasting, shock identification and avoiding spurious fractional cointegration. We derive the asymptotic distribution for both known and unknown break fraction as well as for known and unknown parameters under the null and a local break hypothesis. Further, we extend the proposed testing procedures by allowing for potentially breaking short-run dynamics. Monte Carlo results are provided.

C593: **On the usefulness of cross-validation for directional forecast evaluation**
Presenter: **Mauro Costantini**, Brunel University, United Kingdom
Co-authors: Christoph Bergmeir, Jose Benitez

The purpose is to investigate the usefulness of a predictor evaluation framework which combines a blocked cross-validation scheme with directional accuracy measures. The advantage of using a blocked cross-validation scheme with respect to the standard out-of-sample procedure is that cross-validation yields more precise error estimates since it makes full use of the data. In order to quantify the gain in precision when directional accuracy measures are considered, a Monte Carlo analysis is provided using univariate and multivariate models. The experiments indicate that more precise estimates are obtained with the blocked cross-validation procedure. An application is carried out on forecasting UK interest rate for illustration purposes. The results show that the cross-validation scheme has considerable advantages over the standard out-of-sample evaluation procedure as it makes possible to compensate for the loss of information due to the binary nature of the directional accuracy measures.

C324: **Testing for bubbles**
Presenter: **Philip Hans Franses**, Erasmus School of Economics, Netherlands
Time series with bubble-like patterns display an unbalance between growth and acceleration, in the sense that growth in the upswing is too fast and then there is a collapse. In fact, such time series show periods where both the first differences (1-L) and the second differences (1-L)2 of the data are positive-valued, after which period there is a collapse. For a time series without such bubbles, it can be shown that 1-L2 differenced data should be stable. A simple test based on one-step-ahead forecast errors can now be used to timely monitor whether a series experiences a bubble and also whether a collapse is near. Illustration on simulated data and on two housing prices and the Nikkei index illustrates the practical relevance of the new diagnostic. Monte Carlo simulations indicate that the empirical power of the test is high.

C321: **A combined nonparametric test for seasonal unit roots**
Presenter: **Robert Kunst**, Institute for Advanced Studies, Austria
Nonparametric unit-root tests are a useful addendum to the toolbox of time-series analysis. They tend to trade off power for enhanced robustness features. We consider combinations of the RURS (seasonal range unit roots) test statistic and a variant of the level-crossings count. This combination exploits two main characteristics of seasonal unit-root models, the range expansion typical of integrated processes and the low frequency of changes among main seasonal shapes. The combination succeeds in achieving power gains over the component tests. Simulations explore the finite-sample behavior relative to traditional parametric tests.

CS48 Room Gordon CHANGES IN VOLATILITY AND CORRELATION DYNAMICS Chair: Edoardo Otranto

C070: **A new approach to high-dimensional volatility modelling**
Presenter: **Christian Hafner**, UCL Louvain-la-Neuve, Belgium
A new approach to model multivariate volatilities based on a Kronecker decomposition of the conditional covariance matrix is proposed. We discuss properties of the model, suggest an estimation algorithm, and illustrate the performance by applying it to a high-dimensional financial time series.

C202: **The individual and global determinants of dynamic correlations**
Presenter: **Hassan Malongo**, Paris Dauphine University, France
Co-authors: Jean-David Fermanian
The drivers of the correlations between asset returns are studied in a portfolio framework. The relevance of individual GARCH volatilities, switching regime probabilities and vectors of macro-economic variables will be discussed through the introduction of new DCC type specifications. The performances of these new models are evaluated, particularly to risk-manage financial portfolios dynamically.

C213: **Realized dynamic conditional correlation model**
Presenter: **Denis Pelletier**, North Carolina State University, United States
Co-authors: Aymard Kassi
The Realized GARCH framework is extended to multivariate returns and realized measures of volatility and co-volatility. We introduce an additional measurement equation, based on a regime switching model with Wishart-distributed states, that relates the realized correlations to the conditional correlations of the returns.

C513: **Modeling covariance breakdowns in multivariate GARCH**
Presenter: **John Maheu**, McMaster University, Canada
Co-authors: Xin Jin
The aim is to propose a flexible way of modeling heterogeneous breakdowns in the conditional covariance matrix dynamics of multivariate financial time series within the framework of multivariate GARCH (MGARCH) models. During periods of normal market activities, volatility dynamics are modeled by a MGARCH specification. It refers to any significant temporary deviation of the conditional covariance matrix from its implied MGARCH dynamics as a covariance breakdown, which is captured through a flexible stochastic component that allows for changes in the conditional variances, covariances and implied correlation coefficients. Bayesian inference is used and an efficient posterior sampling procedure is proposed. By applying the modeling approach to daily stock and bond data, it is shown that the model is flexible to capture complex and erratic temporary structural change/deviations from the long run dynamics which is otherwise difficult to account for comprehensively. Modeling covariance breakdowns leads to a significant improvement in the marginal likelihood.

CS50 Room Bedford MARKET LEARNING, RISK AND ASSET PRICING Chair: Valerio Poti

C843: **Learning in a heterogeneous economy**
Presenter: **Francisco Vazquez-Grande**, Federal Reserve Board, United States
The aim is to document business-cycle properties and a significant increase in the average level of risk-prices in the presence of learning in economies with learning agents with heterogeneous beliefs. A model is solved with long-run risk where both, the level and persistence of expected consumption growth are unobserved. A new methodology is introduced to quantify the effects of learning about parameter uncertainty and latent variables. The average historical maximum Sharpe ratio increases significantly in the learning economy when compared to the full-information case, and the difference between the subjective risk-prices of learning and non-learning agents is shown to be counter-cyclical. The agent facing parameter uncertainty chooses state variables that are sufficient statistics of the learning problems and, conditional on her information set, forms posterior distributions of the states and future consumption growth.

C725: **The trading profits and economic losses from false reporting to the libor bank panel**
Presenter: **Conall O'Sullivan**, University College Dublin, Ireland
Co-authors: Gregory Connor, Brian OKelly
An analytical model of Libor rate setting is built in which the Libor quote-setting process serves to aggregate bank-specific information on market-clearing interest rates. We use the model to analyse the impact on the Libor rate of false (and undetected) reporting by one or more banks intending to manipulate the Libor rate. We analyse the potential trading profits to interest rate derivatives traders aware of the false reporting, and the economic costs associated with the loss of Libor rate integrity after false reporting is publicly detected.

C821: **Revealing downturns**
Presenter: **Sergey Zhuk**, University of Vienna, Austria
Co-authors: Martin Schmalz
When investors are risk-averse, cash flow news is more informative about firm value in downturns than in upturns. Good performance in good times can come from desirable good idiosyncratic performance or undesirable high correlation with a market-wide factor. In contrast, good performance in bad times can come from desirable good idiosyncratic performance or desirable low correlation with a market-wide factor. Therefore, good

performance in bad times is a more strongly positive signal about firm value than good performance in good times. Similarly, the signal from bad news is weaker in upturns than in downturns. As a result, (i) stocks react more sensitively to news about performance in downturns, and (ii) negative market returns predict cross-sectional dispersion of stock returns. Both are novel predictions for which we provide support.

C472: **GMM-based tests of efficient market learning and an application to testing for a small firm effect in equity pricing**
Presenter: **Valerio Poti**, DCU, Italy
Co-authors: Akhtar Siddique
The aim is to extend a previous analysis of the implications of the efficient learning market hypothesis (ELM) for asset prices by reformulating it in a GMM setting. Our representation is more amenable to widespread application and allows the econometrician, in testing ELM, to make use of the full range of specification tests that have been developed by the empirical literature in the context of tests of the more restrictive Efficient Market Hypothesis (EMH). We apply this framework to test for efficient learning in the pricing of small capitalization stocks. We find evidence of mispricing of small stocks but we cannot rule out that, in spite of possibly incorrect priors about the future payoffs of small firms, the market efficiently processes information as it becomes available over time. That is, our evidence contradicts the Efficient Market Hypothesis (EMH) but it is not incompatible with efficient learning.

CS57 Room Deller BAYESIAN ECONOMETRICS Chair: Francesco Ravazzolo

C500: **On the stability of standard VARs since the crisis**
Presenter: **Knut Are Aastveit**, Norges Bank, Norway
Co-authors: Andrea Carriero, Todd E. Clark, Massimiliano Marcellino
Small or medium scale VARs are commonly used in applied macroeconomics for forecasting and evaluating the shock transmission mechanism. This requires the VAR parameters to be stable over the evaluation and forecast sample, or to explicitly consider parameter time variation. The earlier literature focused on whether there were sizable parameter changes in the early'80s, in either the conditional mean or variance parameters, and in the subsequent period till the beginning of the new century. A similar analysis is conducted but focusing on the effects of the recent crisis. Using a range of techniques, substantial evidence against parameter stability is provided. The evolution of the unemployment rate seems particularly different from the past. Alternative methods to handle parameter instability are then discussed and evaluated, both in a forecasting and in a policy evaluation context. While none of the methods clearly emerges as best, some techniques turn out to be useful to improve the forecasting performance and obtain shock response functions more in line with the predictions of economic theory.

C499: **Efficient Gibbs sampling for Markov switching GARCH models**
Presenter: **Ayokunle Anthony Osuntuyi**, University Ca Foscari of Venice, Italy
Co-authors: Monica Billio, Roberto Casarin
Efficient simulation techniques for Bayesian inference on switching GARCH models are developed. The contribution to existing literature is manifold. First, different multi-move sampling techniques for Markov Switching (MS) state space models are discussed with particular attention to MS-GARCH models. The multi-move sampling strategy is based on the Forward Filtering Backward Sampling (FFBS) applied to an approximation of MS-GARCH. Another important contribution is the use of multi-point samplers, such as the Multiple-Try Metropolis (MTM) and the Multiple trial Metropolize Independent Sampler, in combination with FFBS for the MS-GARCH process. Finally, it is suggested to further improve the sampler efficiency by introducing the antithetic sampling within the FFBS. The simulation experiments on MS-GARCH model show that the multi-point and multi-move strategies allow the sampler to gain efficiency when compared with single-move Gibbs sampling.

C608: **A direct sampler for G-Wishart variates**
Presenter: **Alex Lenkoski**, Norwegian Computing Center, Norway
The G-Wishart distribution is the conjugate prior for precision matrices that encode the conditional independencies of a Gaussian graphical model. While the distribution has received considerable attention, posterior inference has proven computationally challenging, in part due to the lack of a direct sampler. This situation is rectified. The existence of a direct sampler offers a host of new possibilities for the use of G-Wishart variates. One such development is discussed by outlining a new transdimensional model search algorithm-which terms double reversible jump-that leverages this sampler to avoid normalizing constant calculation when comparing graphical models. It concludes with two short studies meant to investigate the algorithm's validity.

C1140: **Financial frictions and monetary policy tradeoffs**
Presenter: **Paolo Gelain**, Norges Bank, Norway
Co-authors: Francesco Furlanetto, Marzie Taheri Sanjani
The recent global financial crisis illustrates that financial frictions are a significant source of volatility in the economy. The recent generation of estimated New Keynesian models that include financial frictions is considered. We investigate monetary policy stabilization in an environment where financial frictions are a relevant source of macroeconomic fluctuation. We make two contributions. First, we derive a measure of output gap that accounts for financial frictions in the data. Second, we compute the trade-offs between nominal and real stabilization that arise when the monetary policy authority behaves optimally. The presence of financial frictions implies that the central bank faces an additional source of inefficiency, besides the presence of monopolistic competition and nominal rigidities in the goods and labor markets. We find that policy trade-offs are substantial. Price and wage inflation are significantly less volatile under the optimal policy, but stabilization policies fail to counteract the fluctuation in output gap.

CS58 Room Athlone NOWCASTING Chair: Francesca Monti

C135: **Nowcasting Norway**
Presenter: **Matteo Luciani**, Universite libre de Bruxelles, Belgium
Co-authors: Lorenzo Ricci
Predictions of the previous, the current, and the next quarter of Norwegian GDP are produced. To this end, we estimate a Bayesian Dynamic Factor model on a panel of 14 variables (all followed closely by market operators) ranging from 1990 to 2011. By means of a real time forecasting exercise we show that the Bayesian Dynamic Factor Model outperforms a standard benchmark model, while it performs equally well than the Bloomberg Survey. Additionally, we use our model to produce annual GDP growth rate nowcast. We show that our annual nowcast outperform the Norges Bank's projections of current year GDP.

C260: **Nowcasting China real GDP**
Presenter: **Silvia Miranda Agrippino**, London Business School, United Kingdom
Co-authors: Domenico Giannone, Michele Modugno
The aim is to nowcast mainland China real GDP estimating a Dynamic Factor Model on a set of mixed frequency indicators to exploit the information content of more timely data in producing early forecasts of GDP figures. Estimating the model via quasi maximum likelihood allows

us to efficiently handle complications arising from the dissemination pattern and transformations of Chinese data types; Moreover, model-based news can be used to sequentially update GDP nowcasts allowing for a meaningful interpretation of the revision process. The model is evaluated by means of a pseudo real time forecasting exercise over the period 2008 – 2013. We show how it efficiently processes new information improving in terms of forecasting accuracy as more data releases are taken into account; This ultimately delivers a forecast which is comparable to the best judgemental forecasts available within financial markets, while significantly outperforming the standard AR benchmark. We also evaluate model performance against institutional forecasters in terms of annual growth rates: We show how model-implied annual rates become extremely accurate at Q4, beating virtually all the surveyed benchmarks, with performance expected to further improve towards the end of the year.

C953: **Incorporating conjunctural analysis in structural models**
Presenter: **Francesca Monti**, Bank of England, United Kingdom
Co-authors: Lucrezia Reichlin, Domenico Giannone

A methodology to incorporate monthly timely information in estimated quarterly structural DSGE models is developed. We derive the monthly dynamics consistent with the model and discuss under which conditions this mapping is unique. We then augment its state space with auxiliary monthly variables typically used in conjunctural analysis. In this way, we can consider quarterly DSGE based estimates of the parameters as given, while obtaining increasingly accurate early estimates of the quarterly variables - both observed and unobserved - as new data are released throughout the quarter. We illustrate our method on the basis of a previous neo-classical synthesis model, augmented with a panel of thirteen monthly indicators for the U.S. economy.

C1184: **Real-time nowcasting with a Bayesian mixed frequency model with stochastic volatility**
Presenter: **Andrea Carriero**, Queen Mary University of London, United Kingdom

The purpose is to develop a method for producing current-quarter forecasts of GDP growth with a (possibly large) range of available within-the-quarter monthly observations of economic indicators, such as employment and industrial production, and financial indicators, such as stock prices and interest rates. In light of existing evidence of time variation in the variances of shocks to GDP, we consider versions of the model with both constant variances and stochastic volatility. We also evaluate models with either constant or time-varying regression coefficients. We use Bayesian methods to estimate the model, in order to facilitate providing shrinkage on the (possibly large) set of model parameters and conveniently generate predictive densities. We provide results on the accuracy of nowcasts of real-time GDP growth in the U.S. from 1985 through 2011. In terms of point forecasts, our proposal is comparable to alternative econometric methods and survey forecasts. In addition, it provides reliable density forecasts, for which the stochastic volatility specification is quite useful, while parameter time-variation does not seem to matter.

CS60 Room Russell FORECASTING A LARGE DIMENSIONAL COVARIANCE MATRIX Chair: Eduardo Rossi

C368: **Forecasting vast realized covariance matrices and portfolio choice**
Presenter: **Anders Kock**, Aarhus University and CREATES, Denmark
Co-authors: Marcelo Medeiros, Laurent Callot

Precise forecasts of the covariance matrix are vital for the construction of portfolios and the last decades have witnessed a great deal of research in this area. Among the major innovations are the use of ultra high frequency data to compute more reliable daily measures of variance and covariance, as for example, the realized variance and realized covariance. The purpose is threefold. Firstly, estimators based on the Lasso and the adaptive Lasso of high-dimensional realized covariance matrices are proposed. These sparse estimators are used to deal with the curse of dimensionality and to obtain reliable inference in settings where the number of predictors is larger than the sample size. Secondly, upper bounds are proved on the forecast errors of the procedure which are valid even in finite samples. Finally, forecasts of returns and covariances are used for portfolio selection and the economic gain of our method is compared with the one from benchmark models in the literature. The data is composed of 1495 daily realized covariance matrices of the returns of 437 stocks observed from January 2006 to December 2011. It covers the recent crisis which gives us the opportunity to investigate its effects on covariance matrices.

C411: **A fast model averaging method to forecast high dimensional realized covariance matrices**
Presenter: **Paolo Santucci de Magistris**, Aarhus University, Denmark
Co-authors: Eduardo Rossi

The forecast of the realized covariance matrix of financial assets is considered by means of Choleski factorization. Since each factorization depends on the particular ordering of the asset returns, it turns out that the forecast of the realized covariance matrix depends on the chosen ordering. This could potentially induce a bias which can be attenuated by taking average forecasts, each computed for a different sorting. However, when the number of assets is large (e.g. $n > 10$), the number of possible assets orderings, $n!$, is huge. A methodology is proposed to deal with the determination of the minimum number of orderings to be considered to minimize the bias. Bayesian model averaging is also employed to evaluate the relative contribution of each forecast to the average. An empirical analysis is carried out on a set of NYSE stocks; statistical and economic criteria are adopted to compare alternative models.

C820: **Testing for instability in covariance structures**
Presenter: **Lorenzo Trapani**, Cass Business School, United Kingdom
Co-authors: Giovanni Urga, Chihwa Kao

A test for the stability over time of the covariance matrix of multivariate time series is proposed. The analysis is extended to the eigensystem to ascertain 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, and we develop tests for the stability of the loadings in a panel factor model context. 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 AutoRegression applied to exchange rates.

C428: **Risk premia and time-varying higher-order conditional moments: An independent factor model**
Presenter: **Eduardo Rossi**, University of Pavia, Italy
Co-authors: Alexios Ghalanos

The effect that time-varying higher-order conditional moments have on stock risk premia is studied. The model assumes that the returns have a linear factor representation. The factors are assumed and estimated as independent, by means of the Independent Component Analysis. The stock returns have a conditionally multivariate Generalized Hyperbolic distribution with time-varying skewness and kurtosis. The factor representation allows for closed-form expressions of the conditional skewness and kurtosis of the return of any portfolio. As a consequence, the return on market portfolio has time-varying skewness and kurtosis. A large cross-section of NYSE stocks is employed in the analysis. A three-moment asset pricing model is estimated and compared with nested restrictions. An out-of-sample forecasting analysis is carried out. Several alternative models are considered in the comparison.

CS61 Room Woburn MEDIUM-SCALE FACTOR MODELS FOR GDP AND INFLATION FORECASTS Chair: Rosa Ruggeri-Cannata

C128: **Structural FECM: Cointegration in large-scale structural models**
Presenter: **Anindya Banerjee**, University of Birmingham, United Kingdom
Co-authors: Massimiliano Marcellino, Igor Masten
Starting from the dynamic factor model for non-stationary data we derive the factor-augmented error correction model (FECM) and, by generalizing the Granger representation theorem, its moving-average representation. The latter is used for the identification of structural shocks and their propagation mechanism. Besides discussing contemporaneous restrictions, we show how to implement classical identification schemes based on long-run restrictions in the case of large panels. The importance of the error correction mechanism for impulse response analysis is analysed by means of both empirical examples and simulation experiments. Our results show that the bias in estimated impulse responses in a FAVAR model is positively related to the strength of the error-correction mechanism and the cross-section dimension of the panel. We observe empirically in a large panel of US data that these features have a substantional effect on the responses of several variables to the identified real shock.

C293: **The disaggegation map approach to identify non-pervasive common features in a large set of disaggregates**
Presenter: **Antoni Espasa**, Carlos III, Spain
Co-authors: Guillermo Carlomagno
The econometric analysis on prices and inflation has been expanded to consider not only aggregated variables but also their disaggregates. During the last years it has been suggested that applying dynamic factor models (DFM) to subsets of series could provide better forecasting results than when applied to the whole dataset and propose ad hoc procedures for doing it. Recently the disaggregation maps strategy, based on bivariate analyses on all pairs of disaggregates, has been proposed as an objective procedure for constructing these subsets and its results suggest that the factors can be less pervasive than previously thought. We show that DFM fail in these cases. The main contribution is threefold. First, it is shown that disaggregation maps work for different data structures. Second, the disaggregation maps strategy is extended in two directions; i) providing a treatment of outliers, ii) generalizing the search for common features. Third, the properties of the disaggregation maps procedure are studied. Finally, we compare the forecasting performance of this procedure with standard DFM and DFM applied to the subsets coming from a disaggregation map.

C620: **Variable reduction and variable selection methods using small, medium and large datasets: A forecast comparison for the PEEIs**
Presenter: **Fotis Papailias**, Queens University Belfast, United Kingdom
Co-authors: George Kapetanios, Massimiliano Marcellino
The work is concerned with the forecasting performance of variable reduction and variable selection methods using medium and large datasets. The variable reduction methods include Principal Components, Partial Least Squares and Bayesian Shrinkage Regression. The variable selection methods choose the appropriate predictors by minimising an information criterion. Heuristic optimisation algorithms are used that are computationally feasible and include the Genetic Algorithm, the Simulated Annealing, the Sequential Testing and the MC^3. The medium sets are subsets of the original dataset and are created using: (i) the best representative (most correlated) variable in each category and (ii) estimates of the recently introduced exponent of cross-sectional dependence. The forecasting performance of the above methods is assessed in forecasting four principal European economic indicators for the Euro Area economy. These are the quarterly GDP growth, the quarterly final consumption expenditure growth, the monthly industrial production growth and the monthly inflation. The empirical exercise suggests that smaller datasets highly improve the forecasting performance of variable selection models and slightly enhance the variable reduction models.

C493: **Euromind: A flexible framework to estimate euro-area and member states monthly indicators of economic activity**
Presenter: **Rosa Ruggeri Cannata**, European Commission, Luxembourg
Co-authors: Cecilia Frale, Srefano Grassi, Massimiliano Marcellino, Gian Luigi Mazzi, Tommaso Proietti
Timely monitoring the economic situation, at least monthly, is one of the main goals of analysts; unfortunately official statistics present significant gaps at monthly frequency. The aim is to present Euromind, a euro area monthly indicator of economic activity essentially based on official statistics and derived by means of a robust procedure combining single factor models with temporal disaggregation techniques within an unobserved components model. The proposed approach produces estimates for each individual component of the expenditure and output side of GDP which are then aggregated to produce the two different estimates. The final version of Euromind is then obtained as a weighted average of expenditure and output side indicators. We also present Euromind-C, an extension of Euromind allowing for the simultaneous estimates of euro area and some of its largest economies monthly indicators thanks to a medium size factor model. Structure and restrictions of the extended model are presented as well as some issues related with its computationally efficient estimation. Finally, it is shown in details the usefulness of Euromind in business cycle analysis and short term forecasting. A comparative analysis of the behaviour of Euromind and Euromind-C for the euro area closes the work.

CS66 Room Holden SYSTEMIC RISK TOMOGRAPHY Chair: Silvia Figini

C173: **Financial network systemic risk contributions**
Presenter: **Julia Schaumburg**, VU University Amsterdam, Germany
Co-authors: Nikolaus Hautsch , Melanie Schienle
The "realized systemic risk beta" is proposed as a measure for financial companies' contribution to systemic risk given network interdependence between firms' tail risk exposures. Conditional on statistically pre-identified network spillover effects and market and balance sheet information, we define the realized systemic risk beta as the total time-varying marginal effect of a firm's Value-at-risk (VaR) on the system's VaR. Suitable statistical inference reveals a multitude of relevant risk spillover channels and determines companies' systemic importance in the U.S. financial system. Our approach can be used to monitor companies' systemic importance allowing for a transparent macroprudential regulation.

C1035: **Measuring degree of dependence**
Presenter: **Priyantha Wijayatunga**, Umea University, Sweden
Dependence is a fundamental concept in applications of probability and statistics; for example, in modeling tasks of various losses and risks in banks. However, most of the measures that are defined for measuring the dependence between two random variables are capable of doing so only in restricted contexts; Pearson's correlation coefficient can only measure the linear dependence accurately and mutual information measure, which is very popular in statistical machine learning literature, does not possess the properties of a metric. Here, we define a new dependence measure using Hellinger distance which is a metric between the joint probability density (pdf) of the two random variables and their pdf when their independence is assumed. First we show how it can be defined in the context of two binary random variables; a requirement here is a normalization within the unit interval. Then we show how the concept can be extended to the cases of two multinary variables. We compare our measure with popular measures such as correlation coefficients in the contexts where they are measuring the dependence accurately. Finally, we discuss directions for future research.

C490: **Robust estimation of regime switching models**
Presenter: **Luigi Grossi**, University of Verona, Italy
Co-authors: Fany Nan

Threshold Autoregressive (TAR) models are quite popular in the nonlinear time-series literature. In the class of non-linear models, studies addressed to robustifying this kind of models are very few, although the problem is very challenging particularly when it is not clear whether aberrant observations must be considered as outliers or as generated by a real non-linear process. A previous work extended the generalized M estimator method to SETAR models. Its simulation results show that the GM estimation is preferable to the LS estimation in the presence of additive outliers. As GM estimators have proved to be consistent with a very small loss of efficiency, at least under normal assumptions, the extension to threshold models, which are piecewise linear, looks quite straightforward. Despite this observation, a cautionary note has been written to point out some drawbacks of proposed GM estimator. In particular, it is argued and shown, by means of a simulation study, that the GM estimator can deliver inconsistent estimates of the threshold even under regularity conditions. According to this contribution, the inconsistency of the estimates could be particularly severe when strongly descending weighting functions are used. A previous work demonstrates the consistency of GM estimators of autoregressive parameters in each regime of SETAR models when the threshold is unknown. The consistency of parameters is guaranteed when the objective function is a convex non-negative function. A possible function holding these properties is the Huber rho-function. However, the authors conclude that the problem of finding a threshold estimator with desirable finite-sample properties is still an open issue. Although, a theoretical proof has been provided, there is not a well structured Monte Carlo study to assess the extent of the distortion of the GM-SETAR estimators. The aim is to fill this gap by presenting an extensive Monte Carlo study comparing LS and GM estimators under particular conditions. Moreover, an application of robust nonlinear estimation is proposed to the series of electricity prices following the results of the simulation experiment. It is well known that among the stylized facts which empirically characterize electricity prices, the presence of sudden spikes is one of the most regularly observed and less explored feature.

C042: **Systemic risk in Europe**
Presenter: **Michael Rockinger**, UNIL-HEC Lausanne, Switzerland
Co-authors: Eric Jondeau, Robert Engle

Systemic risk may be defined as the propensity of a financial institution to be undercapitalized when the financial system as a whole is undercapitalized. Systemic risk is related to the market capitalization of the firm, its financial leverage, and the sensitivity of its equity return to market shocks. We investigate European financial institutions and describe an econometric approach designed to measure systemic risk for non-U.S. institutions. We expand a recent approach to the case with several factors explaining the dynamics of financial firms returns with asynchronicity of time zones. We apply this methodology to the 196 largest European financial firms and estimate their systemic risk over the 2000-2012 period. We find that banks and insurance companies bear approximately 80% and 20% of the systemic risk in Europe, respectively. Over the period of our study, the countries with the highest levels of systemic risk are the U.K. and France, and the firms with the highest levels of systemic risk are Deutsche Bank and Barclays.

CS71 Room Bloomsbury MODELLING OF SKEWNESS AND FAT TAILS Chair: Mark Steel

C221: **On the computation of multivariate scenario sets for skewed distributions**
Presenter: **Emanuele Giorgi**, Lancaster University, United Kingdom
Co-authors: Alexander McNeil

The problem of computing multivariate scenarios sets based on the notion of half-space depth (HD) is examined. Our interest is motivated by the potential use of such sets in the 'stress testing' of insurance companies and banks whose solvency is dependent on a vector of financial 'risk factors'. We use the notion of half-space depth to define convex scenario sets that generalize the concept of quantiles to higher dimensions. In the case of elliptical distributions these sets coincide with the regions encompassed by the contours of the density function. In the context of multivariate skewed distributions, this equivalence does not hold in general. We consider two parametric families that account for skewness and heavy tails: the generalized hyperbolic and the skew-t distributions. By making use of a canonical form representation, where skewness is completely absorbed by one component, we show that the HD contours of these distributions are 'near-elliptical' and, in the case of the skew-Cauchy distribution, we prove that the HD contours are exactly elliptical. Our results suggest that these distributions are particularly attractive candidates for modeling multivariate stress scenarios in practice. We also investigate the application of these ideas to half-spaces defined by expectiles rather than quantiles.

C244: **Bayesian modelling of skewness and kurtosis with two-piece scale and shape transformations**
Presenter: **Francisco Javier Rubio**, University of Warwick, United Kingdom
Co-authors: Mark F J Steel

The double two-piece transformation defined on the family of uni-modal symmetric continuous distributions containing a shape parameter will be introduced. The distributions obtained with this transformation contain five interpretable parameters that control the mode, as well as the scale and shape in each direction. Four-parameter subfamilies of this class of transformations are also discussed. We also present an interpretable scale and location invariant benchmark prior as well as conditions for the existence of the corresponding posterior distribution. Finally, the use of this sort of models is illustrated with a real data example.

C427: **Modelling of non-normal features in film returns**
Presenter: **Cinzia Franceschini**, Urbino University, Italy
Co-authors: Nicola Loperfido

Movie grosses are known to be very skewed and leptokurtic. From the practical point of view, it means that a few movies have high returns while many movies have little returns. Both skewness and kurtosis have been modeled by means of the Pareto distribution, Levy-stable distribution and the skew-t distribution. An alternative model based on compound distributions is proposed, to account for the heterogeneity in genres'appeals and release strategies. It is shown that this model might explain high levels of skewness and kurtosis, both empirically and theoretically. Theoretical results are illustrated with movie grosses in USA and Canada.

C494: **Bayesian inference for the multivariate skew-normal and skew-t distributions**
Presenter: **Brunero Liseo**, Sapienza Universita di Roma, Italy
Co-authors: Antonio Parisi

Frequentist and likelihood methods of inference based on the multivariate skew-normal model encounter several technical difficulties. In spite of the popularity of this class of densities, there are no broadly satisfactory solutions for estimation and testing problems. A general population Monte Carlo algorithm is proposed which: ì) exploits the latent structure stochastic representation of skew-normal random variables to provide a full Bayesian analysis of the model; ii) accounts for the presence of constraints in the parameter space. The proposed approach can be defined as weakly informative, since the prior distribution approximates the actual reference prior for the shape parameter vector. Results are compared with the existing classical solutions and the practical implementation of the algorithm is illustrated via a simulation study and a real data example. A generalization to the matrix variate regression model with skew-normal errors and extensions to the multivariate skew-t distribution and copula are also presented.

CS35 Room Court PROBABILISTIC FORECASTING: DENSITY FORECASTS AND DISAGREEMENT Chair: Svetlana Makarova

C341: **Forecasting exchange rate densities using model averaging**
Presenter: **Emi Mise**, University of Leicester, United Kingdom
Co-authors: Anthony Garratt
A methodology is applied to produce density forecasts for exchange rates using a number of well known exchange rate fundamentals panel models. The density combination of each model utilizes a set of weights computed using a linear mixture of experts framework to produce potentially non-Gaussian exchange rate forecast densities. In our application, to quarterly data for ten currencies (including the Euro) for the period 1990q1-2008q4, we show that the combined forecast densities, taking into account model and break date uncertainty, produce well calibrated densities at both short and long horizons. They outperform densities combined using equal weights, as well as those produced using best single models and random walk benchmark models. Calculations of the probability of the exchange rate rising or falling using the combined model show a good correspondence with known events and potentially provide a useful measure for uncertainty of whether the exchange rate is likely to rise of fall.

C345: **Measuring uncertainty of a combined forecast and a new test for forecaster heterogeneity**
Presenter: **Xuguang Sheng**, American University, United States
Co-authors: Kajal Lahiri, Huaming Peng
Using a standard factor decomposition of a panel of forecasts, we have shown that the uncertainty of a consensus forecast can be expressed as the disagreement among forecasters plus the volatility of common shock. The uncertainty of the average forecast from the standpoint of a policy maker is not the variance of the average forecast but rather the average of the variances of the individual forecasts that incorporates idiosyncratic risks. Using a new statistic to test for the heterogeneity of idiosyncratic errors under the joint limits with both T and N approaching infinity simultaneously, we show that some previously used uncertainty measures significantly underestimate the conceptually correct benchmark forecast uncertainty.

C480: **Forecasting disaggregates: Small open economies and foreign transmission channels**
Presenter: **Leif Anders Thorsrud**, Norwegian Business School BI, Norway
Co-authors: Hilde Bjornland, Francesco Ravazzolo
The aim is to forecast aggregated and disaggregated GDP employing a Bayesian Dynamic Factor Model and predictive density combinations. Across eight different developed economies it is shown that including foreign factors in the predictive regressions improve forecasting performance, both at the aggregate and at the sectoral level. By tracking the performance over time, the effects of the financial crisis are also insulated, and the comprehensive effect this event has had on business cycle synchronization across countries and sectors is documented. The results relate to the theoretical and empirical literature trying to explain the causes of business cycle synchronization as economies and sectors that are more open to trade, and less specialized in production, tend to be more affected by international business cycles. Lastly, by allowing for sectoral heterogeneity, forecasting aggregates by combining disaggregates yields lower forecasting errors than forecasting aggregates directly.

C637: **Two-dimensional fan charts and uncertainties**
Presenter: **Wojciech Charemza**, University of Leicester, United Kingdom
Co-authors: Carlos Diaz Vela, Svetlana Makarova
The aim is to propose two new approaches to constructing bivariate probabilistic forecasts for inflation and output (usually presented in a form of fan charts) where inflationary and output uncertainties are treated as mutually dependent random variables. The first approach consists of modelling these uncertainties by using bivariate weighted skew normal distribution, which parameters are interpretable in the monetary policy context. Statistical properties of the bivariate weighted skew normal distribution are derived and its simulation discussed. The second approach consists of specifying marginal distributions for inflation and output uncertainties and modelling the dependence using copulas. In both cases the parameters are estimated by the simulated minimum distance method, with policy-relevant restrictions imposed in order to minimize the number of estimated parameters. Empirical results for UK and some other countries are obtained by using bivariate forecasts errors for inflation and output data. Computed probabilities and fan charts are presented and compared with analogous univariate estimates.

CS80 Room 349 JAPAN STATISTICAL SOCIETY: HIGH-FREQUENCY DATA ANALYSIS IN FINANCIAL MARKETS Chair: Yasuhiro Omori

C261: **Adaptive Bayes type estimation for stochastic differential equations based on high-frequency data**
Presenter: **Masayuki Uchida**, Osaka University, Japan
Ergodic multi-dimensional diffusion processes defined by stochastic differential equations are treated. The observed data are high-frequent on an increasing time interval. We consider the adaptive Bayes type estimation of both drift and volatility parameters for the discretely observed ergodic diffusion processes. Adaptive Bayes type estimators are derived in the following steps. (i) Obtain an initial Bayes type estimator of volatility parameter using a simple utility function. (ii) Obtain a Bayes type estimator of drift parameter using the quasi-likelihood function and the estimator of volatility parameter derived in step-(i). (iii) Obtain a Bayes type estimator of volatility parameter using the quasi-likelihood function and the estimator of drift parameter derived in step-(ii). (iv) Repeat steps-(ii) and (iii) till the number of times based on a condition. Asymptotic normality and convergence of moments for the adaptive Bayes type estimators are discussed. In order to obtain the asymptotic properties of the adaptive Bayes type estimators, we apply the Ibragimov-Has'minskii-Kutoyants program with the polynomial type large deviation inequality for the statistical random field. Furthermore, we give an example and simulation results of the adaptive Bayes type estimators.

C588: **State space method for quadratic estimator of integrated variance in the presence of market microstructure noise**
Presenter: **Daisuke Nagakura**, Keio University, Japan
Co-authors: Toshiaki Watanabe
Recently, several authors have considered a class of integrated variance estimators, called the *quadratic estimator* (QE), that takes a quadratic form in observed returns. The QE includes several existing IV estimators, such as the realized variance, realized kernels, etc. Although, even in the presence of market microstructure noise (MMN) in observed prices, some special cases of the QE are consistent and hence do not have the bias due to MMN asymptotically, they still have the bias in finite samples. A state space method is developed so that one can filter out the bias in the QE due to the MMN. A slightly restrictive assumption on the price process is assumed. However, under the assumption, the state space method can always reduce the MSE of a given QE in finite samples.

C694: **Broken symmetry before the financial bubble burst: Evidence from NYSE limit order book**
Presenter: **Teruko Takada**, Osaka City University, Japan
Co-authors: Takahiro Kitajima
Phase transitions in general are known to frequently involve an abrupt change in a symmetry property of the system. Assuming that a financial bubble is one of a phase transition in financial markets, we analyze the change in the degree of asymmetry between buyers and sellers' price behavior in relation with the phases of financial markets. We extensively investigate the New York Stock Exchange OpenBook data, limit-order book for most NYSE-traded securities, from 2003 to 2012. Using robust and efficient nonparametric density estimation method appropriate for data

distributed as fat-tailed, we computed the market capitalization weighted NYSE sellers' price index and NYSE buyers' price index from the NYSE OpenBook data. We found persistent high level asymmetry during the upward trend phase, and subsequent symmetrization starting significantly prior to the Lehman Shock. This is the first quantification and visualization of broken symmetry for the case of financial markets.

C698: **Volatility forecast comparison with biased proxy and related test statistic**
Presenter: **Shuichi Nagata**, Doshisha University, Japan
Co-authors: Kosuke Oya
Various loss functions are employed to evaluate competing volatility forecasting models. The loss function for an evaluation requires true volatility, which is unobservable. A previously introduced class of loss functions guarantees the consistency of the ranking (asymptotically) when the unbiased volatility proxy is used instead of true volatility. However, while realized variance (RV) is commonly used as the proxy in practice, it is natural to consider that RV does not satisfy the unbiasedness condition due to market microstructure noise. We show that such a bias in the volatility proxy can cause misspecified rankings of competing models. We also introduce a new notion for the robustness of loss functions to incorporate the effect of the biased volatility proxy and propose a proper method to evaluate the competing forecasting models if the volatility proxy is biased. We conduct a series of Monte Carlo simulations to access the performance of our method and confirm that the proposed method behaves well. We also report the finite sample properties of DMW test with imperfect (noisy and biased) proxy in various settings.

CS84 Room Chancellor's MULTIVARIATE VOLATILITY MODELS Chair: Jean-Michel Zakoian

C157: **Asymptotic theory of DCC models**
Presenter: **Jean-David Fermanian**, Ensae-Crest, France
The existence of strictly and second order stationary solutions of usual Dynamic Conditional Correlation models is studied. We specify under which conditions some moments of these solutions exist.

C520: **Bayesian inference on CoVaR**
Presenter: **Ghislaine Gayraud**, University of Technology of Compiegne, France
Co-authors: Lea Petrella, Mauro Bernardi
Recent financial disasters emphasised the need to investigate the consequence associated with the tail co-movements among institutions; episodes of contagion are frequently observed and increase the probability of large losses affecting market participants' risk capital. Commonly used risk management tools fail to account for potential spillover effects among institutions because they provide individual risk assessment. The aim is to contribute to analyse the interdependence effects of extreme events providing an estimation tool for evaluating the CoVaR defined as the Value-at-Risk of an institution conditioned on another institution being under distress. In particular, the approach relies on Bayesian quantile regression framework. A Markov chain Monte Carlo algorithm is proposed exploiting the Asymmetric Laplace distribution and its representation as a location-scale mixture of Normals. Moreover, since risk measures are usually evaluated on time series data and returns typically change over time, the CoVaR model is extended to account for the dynamics of the tail behaviour. Application on U.S. companies belonging to different sectors of the Standard and Poor's Composite Index is considered to evaluate the marginal contribution to the overall systemic risk of each individual institution.

C604: **Estimating MGARCH models equation-by-equation**
Presenter: **Christian Francq**, CREST and University Lille III, France
Co-authors: Jean-Michel Zakoian
A new approach is proposed to estimate a large class of multivariate GARCH models. Under the assumption that the conditional variance matrix is decomposed into conditional standard deviations of individual returns and correlations, the method is based on estimating the volatility parameters by quasi-maximum likelihood in a first step, and estimating the correlations based on volatility-standardized returns in a second step. In the case of constant conditional correlation, the strong consistency and asymptotic normality of the two-step estimator are established under mild assumptions. In particular, no moment assumption is made on the observed process. Time-varying conditional correlations are also considered. An application to exchange rates series and stock returns indices illustrate the interest of the approach.

C689: **Spatial risk measures and max-stable processes**
Presenter: **Erwan Koch**, ISFA and CREST, France
It is of prime interest for authorities and insurers to detect the most risky areas. The aim is to study the dependence of the risk with respect to the spatial region. To this purpose the notion of spatial risk measure is introduced. Some examples of such risk measures are proposed and their properties are studied. Since the risk of natural disasters is under concern, these examples are based on max-stable processes.

CS85 Room Jessel NONLINEAR TIME SERIES II Chair: Frederique Bec

C392: **Bootstrap fractional integration tests in heteroskedastic ARFIMA models**
Presenter: **Giuseppe Cavaliere**, University of Bologna, Italy
Co-authors: Morten Orregard Nielsen, A.M. Robert Taylor
The aim is to propose bootstrap implementations of the asymptotic Wald, likelihood ratio and Lagrange multiplier tests for the order of integration of a fractionally integrated time series. The main purpose in doing so is to develop tests which are robust to both conditional and unconditional heteroskedasticity of a quite general and unknown form in the shocks. It is shown that neither the asymptotic tests nor the analogues of these which obtain from using a standard i.i.d. bootstrap admit pivotal asymptotic null distributions in the presence of heteroskedasticity, but that the corresponding tests based on the wild bootstrap principle do. An heteroskedasticity-robust Wald test, based around a sandwich estimator of the variance, is also shown to deliver asymptotically pivotal inference under the null, and it is shown that it can be successfully bootstrapped using either i.i.d. resampling or the wild bootstrap. The dependence of the asymptotic size and local power of the asymptotic tests are quantified on the degree of heteroskedasticity present. An extensive Monte Carlo simulation study demonstrates that signficant improvements in finite sample behaviour can be obtained by the bootstrap vis-a-vis the corresponding asymptotic tests in both heteroskedastic and homoskedastic environments. The results also suggest that a bootstrap algorithm based on model estimates obtained under the null hypothesis is preferable to one which uses unrestricted model estimates.

C314: **The co-integrated vector autoregression with errors-in-variables**
Presenter: **Heino Bohn Nielsen**, University of Copenhagen, Denmark
The co-integrated vector autoregression is extended to allow variables to be observed with classical measurement errors (ME). For estimation, the model is parameterized as a time invariant state-space form, and an accelerated expectation-maximization algorithm is derived. A simulation study shows that, (i) The finite-sample properties of the ML estimates and reduced rank test statistics are excellent. (ii) Neglected measurement errors will generally distort unit root inference due to a moving average component in the residuals. (iii) The moving average component may-in

principle-be approximated by a long autoregression, but a pure autoregression cannot identify the autoregressive structure of the latent process, and the adjustment coefficients are estimated with a substantial asymptotic bias. An application to the zero-coupon yield-curve is given.

C241: **On the rebound of stock returns after bear markets: An empirical analysis from five OECD countries**
Presenter: **Songlin Zeng**, University of Cergy-Pontoise and ESSEC Business School, France
Co-authors: Frederique Bec
An empirical study of the shape of recoveries in financial markets from a bounce-back augmented Markov Switching model is proposed. It relies on models first applied to the business cycle analysis. These models are estimated for monthly stock market returns data of five developed countries for the post-1970 period. Focusing on a potential bounce-back effect in financial markets, its presence and shape are formally tested. Our results show that i) the bounce-back effect is statistically significant and large in all countries, but Germany where evidence is less clear-cut and ii) the negative permanent impact of bear markets on the stock price index is notably reduced when the rebound is explicitly taken into account.

C236: **The role of inventory investment in the business cycle recovery**
Presenter: **Melika Ben Salem**, University Paris-Est Marne-la Vallee, France
Co-authors: Frederique Bec
The inventory investment behavior is often considered as exacerbating aggregate fluctuations. To check this widely held belief, the possibly nonlinear joint dynamics of inventory investment and real GDP growth rate are analyzed. More precisely, the goal is to evaluate the role of inventory investment in the real GDP growth rate bounce-back detected at the end of recessions, early in the recovery phase. Our contribution is twofold. First, the framework is extended to a multivariate setup within which we expect the bounce-back non-linearity to be fully captured by the inventory investment equation, and hence, the conditional multivariate dynamics of the real GDP growth rate to be linear but not exogenous. Simple Granger-causality tests can answer this question. Second, the multivariate framework developed here allows for the exploration of the predominant motive for holding inventories, which in turn is a key feature of their destabilizing nature. This is achieved by inspection of Generalized Impulse Response Functions to aggregate supply and demand shocks. First results using quarterly French and American postwar data provide evidence that (i) the null hypothesis of no bounce-back effect is strongly rejected and (ii) GDP growth rate bounce-back originates in investment inventory bounce-back.

CS63 Room Torrington MACRO AND HETEROGENEOUS AGENTS Chair: Xavier Mateos-Planas

C1152: **An equilibrium Roy model of housing**
Presenter: **Jonathan Halket**, University of Essex, United Kingdom
Co-authors: Lars Nesheim, Florian Oswald
Systematic co-variation in rent-to-price ratios and homeownership across location and housing attributes is documented. Homeowners tend to own homes that are relatively more expensive on the owner-occupied market. We build an equilibrium Roy model with heterogeneous households and households to document the nature and extent of friction in the rental market.

C1233: **Job uncertainty and deep recessions**
Presenter: **Vincent Sterk**, University College London, United Kingdom
Co-authors: Morten Ravn
A heterogeneous agents model is studied that combines frictions in the labor market with incomplete asset markets and nominal rigidities in price setting. Workers experience job terminations that can either send them into short term unemployment or into longer term unemployment. An increase in job uncertainty depresses aggregate demand which is transmitted to the supply side and produces a significant drop in job finding rates because firms cut back on vacancy postings. The amplification mechanism is small when asset markets are complete, prices are flexible or unemployment is predominantly short term. Applied to the Great Recession, the model can account for the sharp rise in the level of unemployment and for much of the shift and movement along the Beveridge curve observed during this recession. Job uncertainty also emerges as a plausible candidate for the shock that sent the US economy into a liquidity trap and we show that the zero lower bound on nominal interest rates can amplify the recession very significantly in this environment.

C1251: **Formal versus informal default in consumer credit**
Presenter: **Xavier Mateos-Planas**, Queen Mary University of London, United Kingdom
Informal default (or delinquency) in consumer credit is studied as a process involving negotiations over unpaid debts. We consider an economy with uninsurable individual risk where, as an alternative to informal default, households can also declare formal bankruptcy. Negotiated settlements, which can be reached quickly or with some delay, involve limited recovery, are often followed by further defaults, and may end in bankruptcy as households reduce their assets in the process. When calibrated to aggregate measures of default, debt and wealth, the model yields differences in the financial positions of formal and informal defaulters which are quantitatively consistent with observed data. Informal defaulters have higher debts, assets, and net worth, but lower incomes. The opportunity to bargain - in contrast with informal default via prescriptive collection procedures - provides substantial insurance opportunities. It considerably mitigates the known adverse consequences of formal bankruptcy. Bargaining also enhances the welfare gains following from a tighter asset exemption. Attempts at limiting debt collection outside bankruptcy lower generally, but not uniformly, welfare.

CP01 Room Macmillan POSTER SESSION I Chair: Ayse Ulgen

C802: **Combination of tests for cointegration in cross-correlated panels**
Presenter: **Verena Werkmann**, Goethe University Frankfurt, Germany
A previous combination of "combinations of p-values" (CCP) is applied for testing for cointegration in cross-correlated panels and extended to a combination of three tests. The p-value combination procedures are combined based on a union of rejections rule in order to obtain a test displaying good power properties under uncertainty about the number of false null hypotheses in the panel. This idea is based on the fact that some tests exhibit higher power when many hypotheses are false whereas others are more powerful when the panel features only one or few false hypotheses. An extensive Monte Carlo study shows that the CCP test improves upon the power compared to applying a less powerful test individually if the tests combined are chosen such that their rejections overlap as little as possible. Furthermore, the use of the combinations studied is illustrated by an application to international interest rate linkage. Cross-sectional dependencies in both the simulation and the empirical study are accounted for by using a sieve bootstrap.

C928: **Coupling the risks: An integrated risk measure of economic capital**
Presenter: **Radovan Parrak**, Charles University, Czech Republic
A redefined approach is developed to model integrated economic capital and stress testing, built upon a stochastic generator of economic states-of-world. Unlike the rest of the academic literature focused on coupling different risk types directly, we model the dynamics of underlying risk drivers. Additionally, our framework allows us to build a tractable link between macroeconomic developments and their consecutive impact on

credit risk, being currently a lively topic among practitioners. We show that our economic state-of-world generator is able to deliver a plausible set of scenarios; and use these, in turn, for the computation of an enterprise-wide economic capital of a hypothetical bank. Additionally, we show that a subset of these scenarios is readily available to define a meaningful stress test, in line with the observed mutual dependencies among the risk drivers. To our best knowledge, there is not a fully integrated risk model of this type in the academic literature yet.

C968: **Comparison of two methods for calculating business efficiency**
Presenter: **Felipe Rosa-Gonzalez**, Universidad de La Laguna, Spain
Co-authors: Enrique Gonzalez-Davila, Antonio Arbelo-Alvarez

Any tool that helps companies to improve their results is very welcome. Even more if the economic framework is like today. In a stormy frame, like the present, one of these tools can be a measure of a company's Efficiency. This allows comparison with others operating in the same sector and, as a result of this comparison, it will be possible to draw conclusions about the operation of the company, and this will allow for improvement. For this reason, the measure of Efficiency will serve to help business leaders in their decision making. There are two kinds of methods when measuring Efficiency: the Nonparametric and Parametric. By applying these last methods, it is possible at least to calculate the Efficiency using two ways: the so-called Distribution Free Approach, and another which includes the Stochastic Frontier Approach. For this purpose, a population is generated based on a set of real data and then different samples are taken from it. By means of the two methodologies, population and sample efficiencies are calculated and later compared. The obtained results are then compared in order to study possible different behaviours depending on the used method.

C978: **Cheat and let cheat: An improvement of the tax system induced by a collective cheating**
Presenter: **Benjamin Rausch**, ULB, Belgium

In a world where the government only has redistributional purposes, it is shown that the tax system can be enhanced thanks to a collective cheating that is implicitly supported by the government itself. Indeed, if the different people working in the same sector play collectively, the government's role will be reduced and the costs related to the controls and to the corruption will also be reduced, hence the tax system will become more efficient. This is demonstrated thanks to a simple 2 people example. Finally, this model will be put in parallel with the medieval guilds.

C1121: **A dynamic generalization of orthogonal GARCH models**
Presenter: **Radek Hendrych**, Charles University in Prague, Czech Republic

Analysis of conditional covariances is indisputably one of the most important parts of multivariate financial time series modelling. Time-varying covariances are crucial inputs for many tasks of financial, portfolio and risk management. In particular, they are worth of interest not only from a theoretical perspective, but also from a practical point of view. The aim is to introduce a dynamic generalization of orthogonal GARCH models (OGARCH). The proposed innovation is based on an adaptation of the key principle of the OGARCH: An original time-invariant orthogonal transformation of multivariate time series is generalized into the more realistic time-varying form. Namely, this suggested adjustment is delivered by the LDL decomposition of the conditional covariance matrix and associated state space modelling. The performance of this technique is examined in a simulation study. Moreover, its empirical capabilities are tested in a financial application.

C1124: **Measuring spillovers in fractionally cointegrated time-series: The case of realized volatility**
Presenter: **Tomas Krehlik**, Academy of Sciences of the CR, Czech Republic
Co-authors: Jozef Barunik

Since the economic downturn in 2008 many researchers have concentrated on ways of measuring systemic risk, and spillovers have become a standard measure. Time-series used for spillovers measurement are often cointegrated as well as contain long-memory, and the current methods for measuring spillovers do not explicitly account for these features. Time-series of realised volatility of most liquid assets is investigated by methods that take into account these stylized facts, i.e. long-memory and cointegration. Specifically, fractionally cointegrated vector autoregression (FCVAR) is used to compute the volatility spillover estimates and those are compared to benchmark estimates that are computed using vector autoregression as is standard in the literature. Preliminary results show a significant change in spillovers when long memory is modelled. Moreover, taking advantage of the cointegration, a novel short-term/long-term spillover decomposition is proposed, that greatly enhances our understanding of volatility cross-market dynamics both in turbulent and calm times.

Saturday 14.12.2013 16:40 - 18:45 Parallel Session F – ERCIM

ES13 Room Jessel MODEL ASSESSMENT Chair: M. Dolores Jimenez-Gamero

E052: **Testing skew-normality against scale mixing of skew-normals**
Presenter: **Nicola Loperfido**, University of Urbino, Italy
A method is proposed for testing the hypothesis that data come from a multivariate skew-normal distribution against the alternative hypothesis that data come from a scale mixture of multivariate skew-normal distributions. The method is based on the singular value decomposition of the third multivariate cumulant and is illustrated with well-known datasets.

E109: **A model-specification test for GARCH(1,1) processes**
Presenter: **Anne Leucht**, University of Mannheim, Germany
Co-authors: Jens-Peter Kreiss, Michael H. Neumann
There is already an overwhelming amount of model specification tests in the econometric literature. These methods typically rely on the assumption that the information variables, as well as the response variables, are observable. However, this condition is violated in the case of GARCH models, where unobserved quantities enter the information variable. We establish a consistent model-specification test for GARCH(1,1) models based on an L_2-type test statistic. The latter can be approximated by a degenerate V-statistic under the null hypothesis and its asymptotics are then derived invoking recent results on V-statistics under ergodicity. Since the test statistic and its limit distribution depend on unknown parameters in a complicated way, critical values cannot be derived directly. We present a bootstrap-based testing procedure that overcomes these difficulties. The approach presented can be carried over to many other test problems.

E153: **Modeling stationary data by classes of generalized Ornstein-Uhlenbeck processes**
Presenter: **Alejandra Cabana**, Universitat Autonoma de Barcelona, Spain
Co-authors: Enrique M. Cabana, Argimiro Arratia
Consider the complex Ornstein Uhlenbeck operator defined by $\mathrm{OU}_\kappa y(t) = \int_{-\infty}^{t} \mathrm{e}^{-\kappa(ts)}$, $\kappa \in \mathbf{C}, \Re(\kappa) > 0$ provided that the integral makes sense. In particular, when y is replaced by a standard Wiener process w on R the integral becomes a Wiener integral and $x(t) = \mathrm{OU}_\kappa w(t)$, is an Ornstein Uhlembeck process. When y is replaced by a fractional Brownian motion b^H with $H > 1/2$, we call the resulting process a *Fractional Ornstein-Uhlenbeck process (FOU) with parameter H*. New stationary processes can be constructed by iterating the application of OU to the Wiener process or the fractional Brownian motion. We denote by $\mathrm{OU}(p)$ and $\mathrm{FOU}(p)$ the families of processes obtained by applying successively p OU operators OU_{κ_j}, $j = 1,2,\ldots,p$ to w or b^H. These processes can be used as models for stationary Gaussian continuous parameter processes, and their discretized version, for stationary time series. We show that the series $x(0), x(1), \ldots, x(n)$ obtained from $x = \prod_{j=1}^{p} O\mathcal{U}_{\kappa_j}(\sigma w)$ satisfies an $\mathrm{ARMA}(p, p-1)$ model. We present an empirical comparison of the abilities of OU, FOU and ARIMA models to fit stationary data. In particular, we show examples of processes with long dependence where the fitting of the empirical autocorrelations is improved by using OU or FOU process rather than ARIMA models, and with fewer parameters in the former case.

E487: **Tests of homogeneity of variance functions in nonparametric regression**
Presenter: **Juan-Carlos Pardo-Fernandez**, Universidade de Vigo, Spain
Co-authors: Maria Dolores Jimenez-Gamero, Anouar El Ghouch
When comparing several populations it is interesting not only comparing their means, but also other distributional features such as the variances. For example, that would be the case if one is interested in verifying the homogeneity of several production processes. A conditional setting is considered where along with the variable of interest a covariate is observed, and several statistics are presented to test for the equality of the conditional variances functions in k fully nonparametric regression models. The test statistics are based on the comparison of the probability laws of two samples of residuals, one constructed under the null hypothesis and the other one constructed under the alternative. The comparison of the laws can be performed in terms of the cumulative distribution functions or in terms of the characteristic functions. Both cases are presented and the behaviour of the corresponding test statistics is analysed in detail, both from a theoretical point of view and by means of simulations.

E766: **Adjustment of overdispersion models using loglinear models**
Presenter: **Nirian Martin**, Carlos III University of Madrid, Spain
Co-authors: Juana Maria Alonso-Revenga
A statistical procedure is presented for considering the adjustment of Overdispersion Models which are required when the underlying sampling structure is given by subpopulations such as those following a Dirichlet-multinomial distribution. A well-known example illustrates the followed methodology. A simulation study shows the performance of our proposed method in comparison with existing tools.

ES14 Room Deller MODEL SELECTION IN STATISTICAL LEARNING Chair: Gilles Celeux

E405: **Enhancing the selection of a number of clusters in model-based clustering with external qualitative variables**
Presenter: **Jean-Patrick Baudry**, UPMC, France
Co-authors: Margarida Cardoso, Gilles Celeux, Maria Jose Amorim, Ana Sousa Ferreira
Usual criteria to select a number of clusters in model-based clustering, such as BIC or ICL for example, sometimes lead to an unclear, uncertain, choice. A criterion is proposed that takes into account a classification of the data which is known a priori and may shed new light on the data, help to drive the selection of a number of clusters and make it clearer, without involving this "external" classification in the design of the clustering itself. The variables used to build the clustering and the (qualitative) variables corresponding to the "external" classification have to be chosen carefully and with respect to the modeling purpose. This is an illustration of how the modeling purpose is directly involved in what a "good" number of clusters should be and to what extent the latter should be thought of as dependent on the context: the search of a "good" number of clusters is not blindly looked for but this search is illuminated here by the available "external" classification.

E461: **Model selection with untractable likelihood**
Presenter: **Christine Keribin**, Universite Paris Sud, France
Co-authors: Vincent Brault
Penalised likelihood criteria such as AIC or BIC are popular methods used to deal with model selection and require to compute the maximised likelihood. Unfortunately, this maximised likelihood can be untractable, as it is the case for the latent block model (LBM). LBM is a mixture model that allows us to perform the simultaneous clustering of rows and columns of large data matrices, also known as coclustering. Due to the complex dependency structure of the observations conditionally to the row and column labels, approximations must be defined to perform the E step of the EM algorithm, leading to a lower bound of the maximised likelihood. For the same reason, the usual asymptotic approximation used to derive BIC is itself questionable. On the other hand, the integrated completed likelihood criterion (ICL) is exactly computed for LBM, but requires to

investigate the influence of hyperparameters. This influence is investigated, links between the criteria are discussed and numerical experiments on both simulated and real data sets highlight the proposed model selection procedure.

E553: **Optimal cross-validation in density estimation**
Presenter: **Alain Celisse**, Lille University, France
The main contribution is to provide a new theoretical understanding of cross-validation (CV) in the context of density estimation. The performance of CV is analyzed in two contexts: (i) risk estimation and (ii) model selection. The main focus is given to one CV algorithm, called leave-p-out (Lpo), where p denotes the cardinality of the test set. Closed-form expressions are settled for the Lpo estimator of the risk of projection estimators, which makes V-fold cross-validation completely useless. From a theoretical point of view, these closed-form expressions allow us to study the CV performances in terms of risk estimation. For instance, the optimality of leave-one-out (Loo), that is Lpo with $p = 1$, is proved among CV procedures. Two model selection frameworks are also considered: Estimation, as opposed to Identification. Unlike risk estimation, Loo is proved to be suboptimal as a model selection procedure for Estimation. With finite sample size n, optimality is achieved for p large enough (with $p/n = o(1)$) to balance overfitting. When the purpose is the identification of the best estimator, the optimal choice of p/n is related to the rate of convergence of the best estimator. When the latter reaches a parametric rate for instance, model consistency is proved for p/n converging to 1 as n increases. All these theoretical results are illustrated by simulation experiments.

E624: **On MDL and frequentist statistics**
Presenter: **Tim van Erven**, University Paris-Sud and INRIA, France
To select the right model is to find the right trade-off between the model's fit on the data and the complexity of the model. The minimum description length (MDL) principle provides a way to formalize this trade-off using ideas from information theory. Suppose it takes L(M) bits to describe model M, together with its parameters, and it takes L(D|M) bits to describe the (random) deviations of the data D from the model M. Then MDL selects the model that minimizes L(M) + L(D|M), which may be interpreted as choosing the model that leads to the simplest or shortest description of the data. The question of how well MDL works from a frequentist perspective is considered. A partial answer was previously given by showing that MDL works very well, except that you have to multiply L(M) by a factor of 2. That is, you have to describe the model twice. Although this is fine if you just want to construct a frequentist estimator, it goes completely against the MDL philosophy of finding the shortest possible description.The effect of the factor of 2 is explained, and whether it is really necessary is discussed.

E743: **Relevant statistics for Bayesian model choice**
Presenter: **Jean-Michel Marin**, University Montpellier, France
Co-authors: Natesh Pillai, Christian Robert, Judith Rousseau
The choice of the summary statistics used in Bayesian inference and in particular in ABC algorithms has bearings on the validation of the resulting inference. Those statistics are nonetheless customarily used in ABC algorithms without consistency checks. We derive necessary and sufficient conditions on summary statistics for the corresponding Bayes factor to be convergent, namely to asymptotically select the true model. Those conditions, which amount to the expectations of the summary statistics to asymptotically differ under both models, can be exploited in ABC settings to infer whether or not a choice of summary statistics is appropriate, via a Monte Carlo validation.

ES16 Room Senate NONPARAMETRIC AND ROBUST STATISTICS Chair: Andreas Christmann

E402: **Qualitatively robust estimation for stochastic processes**
Presenter: **Robert Hable**, University of Bayreuth, Germany
Co-authors: Katharina Strohriegl
In the case of i.i.d. random variables, the well-known Hampel-Theorem offers a comfortable way for proving qualitative robustness: if the estimator can be represented by a statistical functional, then qualitative robustness follows from continuity of the statistical functional. It is shown that a similar result is true for a large class of stochastic processes. If the empirical measure corresponding to the stochastic process converges, then continuity of the statistical functional is again sufficient for qualitative robustness. In view of Varadarajan's Theorem, these processes are called Varadarajan processes, and it is shown that many processes have this property, e.g., stationary ergodic processes, α-mixing processes, weakly dependent stationary processes. In particular, all processes which fulfill a weak law of large numbers have the Varadarajan property. It is important here that only the "ideal" stochastic process has to be Varadarajan but not the "contaminated" process. Therefore, the results also cover violations of assumptions such as stationarity. The fact that continuity of the statistical functional turns out to be sufficient not only in the i.i.d. case has a nice practical consequence: if a method is known to be qualitatively robust for i.i.d. data, then it is also qualitatively robust for many non-i.i.d. data.

E566: **Robust estimators for additive models with missing responses using local polynomials**
Presenter: **Graciela Boente**, Universidad de Buenos Aires and CONICET, Argentina
Co-authors: Alejandra Martinez, Matias Salibian-Barrera
Additive models are widely used to avoid the difficulty of estimating regression functions of several covariates without assuming a parametric model. Different estimation procedures for these methods have been proposed in the literature, including several local polynomials methods combined with marginal integration. It is easy to see that most of these estimators can be affected by a small proportion of atypical observations, and hence the interest is obtaining robust estimators for additive models that can also accommodate missing responses in the data. A robust approach is considered that combines marginal integration with robust local polynomial kernel M-estimators. This approach circumvents the curse of dimensionality.

E629: **Robust and efficient residual scale estimators**
Presenter: **Stefan van Aelst**, Ghent University, Belgium
Co-authors: Gert Willems, Ruben Zamar
Robustness and efficiency of residual scale estimators in regression is important for robust inference. A class of robust generalized M-scale estimators for the regression model is introduced. The influence function and gross-error sensitivity is derived and also their maxbias behavior is investigated. In particular, overall minimax bias estimates for the general class and also for well-known subclasses are obtained. Moreover, it is shown how generalized M-scale estimators can be obtained with maximal efficiency subject to a lower bound on the global and local robustness of the estimators.

E1047: **Density estimation in infinite dimensional exponential families**
Presenter: **Bharath Sriperumbudur**, University of Cambrige, United Kingdom
Co-authors: Kenji Fukumizu, Arthur Gretton, Revant Kumar, Aapo Hyvarinen
The problem of estimating densities in an infinite dimensional exponential family indexed by functions in a reproducing kernel Hilbert space is considered. Since standard techniques like maximum likelihood estimation (MLE) or pseudo MLE (based on the method of sieves) do not yield practically useful estimators, we propose an estimator based on the score matching method, which involves solving a simple linear system. We show that the proposed estimator is consistent, and provides convergence rates under smoothness assumptions (precisely, under the assumption

that the true parameter or function that generates the data generating distribution lies in the range space of a certain covariance operator). We also empirically demonstrate that the proposed method outperforms the standard non-parametric kernel density estimator.

E1055: **The Kullback-Leibler divergence and model uncertainty**
Presenter: **Stephan Morgenthaler**, EPFL, Switzerland
The Kullback-Leibler Divergence (KLD) is an information measure and a useful guide for statisticians. We will first recall some properties of this information measure that are useful for applied statisticians and then apply it to the problem of identifying a weak signal hidden in random noise. Two examples will be considered. In the first case, the signals are expected values and only a few of the variables have a positive signal. In the second case the signals are correlations or partial correlations with only a few couples carrying a positive signal.

ES17 Room Russell SPATIAL CLUSTERING Chair: Peter Congdon

E085: **Penalized fast subset scanning**
Presenter: **Skyler Speakman**, Carnegie Mellon University, United States
Co-authors: Sriram Somanchi, Edward McFowland, Daniel Neill
The Penalized Fast Subset Scan (PFSS) is a new and general framework for scalable and accurate pattern detection. The recently proposed Fast Subset Scan (FSS) enables exact and efficient identification of the most anomalous subsets of the data, as measured by a likelihood ratio scan statistic. However, PFSS allows us to incorporate prior information about each data element's probability of inclusion, which was not possible in the FSS framework. As a concrete example, we generalize the Fast Localized Scan approach to incorporate 'soft' constraints on spatial proximity, enabling more accurate detection of irregularly-shaped spatial clusters. We develop a distance-based penalty function in the PFSS framework which rewards spatial compactness and penalizes spatially dispersed clusters. This approach was evaluated on the disease surveillance task using two years of Emergency Department data from 97 Allegheny County zip codes, and increased power to detect disease outbreaks while maintaining efficient computation.

E155: **Spatial cluster detection and interpretation**
Presenter: **Alan Murray**, Arizona State University, United States
Co-authors: Tony Grubesic, Ran Wei
Discovering significant clusters in data has been recognized as an important yet challenging endeavor. Clustering techniques are a fundamental group of exploratory methods designed to identify areas exhibiting elevated levels of disease, risk, danger, etc. Given the intent of cluster detection, spatial structure plays an important role in geographic data and must be taken into account appropriately if meaningful clusters are to be found. The aim is to discuss contiguity and the ways in which it is central to local clusters that may be of interest for planners, managers and policy makers. While spatial contiguity is widely considered an important condition of a geographic cluster, most detection approaches employ a priori artificial structure, leading to misleading significance, unintended spatial biases and hindering meaningful discovery and interpretation. The basis for significance is reviewed, and methods for maximizing likelihood are detailed. An approach is presented for addressing spatial contiguity explicitly in this context. A case study using crime events within a major urban region is presented, with empirical results used to illustrate capabilities for identifying significant and meaningful clusters.

E214: **Identifying clusters in Bayesian disease mapping**
Presenter: **Duncan Lee**, University of Glasgow, United Kingdom
Co-authors: Craig Anderson, Nema Dean
The aim of disease mapping is to identify clusters of areal units which have a high-risk of a particular disease, so that the appropriate remedial action can be taken. Conditional autoregressive (CAR) models are generally used to model the spatial correlation present in areal disease data, but they have shortcomings when attempting to identify the spatial extent of high-risk clusters. We propose a two-stage method for identifying such high-risk clusters, and are motivated by a study of respiratory disease in the Greater Glasgow and Clyde health board. The first stage is a spatially-adjusted clustering technique, which is applied to data prior to the study period and produces multiple potential cluster structures for the disease data. The second stage uses the real data to choose between the set of potential cluster structures using Bayesian hierarchical models.

E280: **Dealing with missing spatial locations in social networks for epidemiological purposes**
Presenter: **Renato Assuncao**, Universidade Federal de Minas Gerais, Brazil
Co-authors: Erica Rodrigues, Gisele Pappa, Renato Miranda, Wagner Meira
Social networks, like Twitter and Facebook, are valuable sources to monitor real-time events such as earthquakes and epidemics. For this type of surveillance the user's location is an essential piece of information, but a substantial number of users choose not to disclose their geographical information. However, characteristics of the users' behavior, such as the friends they associate with and the types of messages published may hint on their spatial location. We present a method to infer the spatial location of Twitter users. Unlike the approaches proposed so far, we incorporate two sources of information to learn the geographical position: the text posted by users and their friendship network. We propose a probabilistic approach that jointly models the geographical labels and the Twitter texts of the users organized in the form of a graph representing the friendship network. We use the Markov random field probability model to represent the network and learning is carried out through a Markov chain Monte Carlo simulation technique to approximate the posterior probability distribution of the missing geographical labels. We demonstrate the utility of this model in a large dataset of Twitter users, where the ground truth is the location given by GPS devices. The method is evaluated and compared to two baseline algorithms that employ either of these two types of information, The accuracy rates achieved are significantly better than those of the baseline methods.

E227: **Local join counts for spatial clustering in disease risk: Simulations and a real data application**
Presenter: **Peter Congdon**, QMUL, United Kingdom
Models for spatial variation in disease risk often consider posterior exceedance probabilities of elevated disease risk in each area. Such "hotspot" detection does not express aspects of broader local clustering: for example, two contrasting configurations are (a) when both area i and surrounding areas j have elevated risk (i.e. area i belongs to a high risk cluster), and (b) when area i has dissimilar risk from neighbour areas (area i is a high risk outlier). This talk considers local join-count statistics in conjunction with Bayesian models of area disease risk to detect different forms of disease clustering over neighbouring areas. The technique is illustrated with a simulated spatial pattern where relative risk contrasts have a known form, and where posterior probabilities of high risk cluster status are obtained that accurately reproduce features of the simulated pattern. A real data application is to suicide deaths in 922 small areas in the North West of England. The univariate join-count approach is extended to an analysis of bivariate spatial clustering in suicide deaths and hospital admissions for intentional self-harm in these areas. The bivariate analysis illustrates the impact of the spatial prior (whether correlated between outcomes or not) on inferences about spatial clusters.

ES29 Room Montague MODELING AND FORECASTING IN POWER MARKETS Chair: Carolina Garcia-Martos

E200: **Weather variables scenarios and simulation for electricity markets risk management**
Presenter: **Virginie Dordonnat**, EDF research and development, France

Electricite de France (EDF) is the major competitor in the French electricity market. The company produces and sells electricity to other market participants as well as to end-use customers and therefore manages uncertainties related to both electricity production and demand. We concentrate here on the physical hazards involved in electricity generation, mostly the weather-related ones (e.g. consumption, wind farms production, inflows in hydroelectric power plants). Physical risk management means simulating electricity generation hazards from tomorrow up to 5 years. In order to generate adequate simulations we need to deal with all these uncertainties, including any predictive information and anticipate future changes (such as climate change), but also consider dependence between hazards. A first method is to use historical data: working with the same period of data is a natural way to account for the implicit dependence between hazards. We present the second approach that implies statistical modelling and focus on why it is useful for risk management compared to historical data and how we consider dependence in our models (correlation, weather regimes).

E483: **Demand response impact on price development in a smart grid context**
Presenter: **Emil Larsen**, Technical University of Denmark, Denmark
Co-authors: Pierre Pinson

Demand response is set to flourish in Europe, encouraged by a target of 80% of consumers having hourly-rate electricity meters by 2020. It is proposed how demand response can be integrated into existing electricity markets. How accurate price and demand response forecasts have to be useful is also discussed. The implications of high uncertainty in forecasting the demand response and prices can be grid instability, resulting in higher primary reserve use. More accurate forecasts can deliver savings to customers and grid operators, while primary and secondary reserve capacity payments can even be considered for the most flexible customers. A bottom up approach is employed to model how distributed energy resource (DER) controllers work. The demand response and its forecast are then considered by a unit commitment model, which can be used to consider price evolution in the future. Methods and results from the EcoGrid EU demonstration, which is a smart grid demonstration with 2000 houses on the Danish island of Bornholm, is drawn upon to support the theoretical work.

E515: **Nonlinear power price forecasting with wind effects**
Presenter: **Angelica Gianfreda**, London Business School, United Kingdom
Co-authors: Derek Bunn, Dipeng Chen

The aim is to present a nonlinear regime switching method for short term power price forecasting to reflect the effects of increased wind generation on price risk. It is shown that the stochastic nature of wind generation poses a new set of distributional properties for the power price risks. Alternative regime switching methods are compared as well as linear methods, and it is controlled for the price formation fundamentals of demand and fuels. The approach is applied to the main reference market for Europe, the EEX, with extensive out-of-sample testing.

E914: **Quantile forecasting of wind power using variability indices**
Presenter: **Georgios Anastasiades**, University of Oxford, United Kingdom
Co-authors: Patrick McSharry

Wind power forecasting techniques have received substantial attention recently due to the increasing penetration of wind energy in national power systems. While the initial focus has been on point forecasts, the need to quantify forecast uncertainty and communicate the risk of extreme ramp events has led to an interest in producing probabilistic forecasts. Using four years of wind power data from three wind farms in Denmark, we develop quantile regression models to generate short-term probabilistic forecasts from 15 minutes up to six hours ahead. More specifically, we investigate the potential of using various variability indices as explanatory variables in order to include the influence of changing weather regimes. These indices are extracted from the same wind power series and optimized specifically for each quantile. The forecasting performance of this approach is compared with that of appropriate benchmark models. Our results demonstrate that variability indices can increase the overall skill of the forecasts and that the level of improvement depends on the specific quantile.

E749: **Optimal risk management in the energy context**
Presenter: **Carolina Garcia-Martos**, Universidad Politecnica de Madrid, Spain
Co-authors: Francisco Javier Nogales, Eduardo Caro

The main objective is the development of quantitative tools for the adequate risk management in the current context of the energy sector. Particularly, we propose a tool for the optimal management energy portfolios including electricity, gas, oil and CO2 allowances contracts. We firstly consider a multivariate model for the aforementioned commodity prices, and calculate point forecasts and forecasting intervals for different forecasting horizons by using bootstrap methods. This will allow us to improve the estimation of the variance-covariance matrix, which is a crucial task for the correct management of the profitability of the portfolio. In a second stage we focus on the development of optimization strategies to hedge against the risk implied by prediction errors of the commodities. The use of conditional VaR or the Expected Shortfall is considered and these measures are preferred in the energy context given that here small changes in the inputs of the optimization problem (the forecasts) could cause a great deviation in the optimal decision. Both the forecasting step, for different horizons, including the evaluation of the coverages of the intervals, as well as the risks associated to each portfolio proposed, are validated, both for simulated data including all the main features that real data present as well as for these real datasets. The results could improve the economic efficiency by means of the improvement of the competitiveness between the companies of a strategic sector such as energy.

ES32 Room Holden BIAS REDUCTION IN STATISTICS OF EXTREMES Chair: Ivette Gomes

E758: **Computational study of a bootstrap algorithm for the adaptive estimation of the extreme value index**
Presenter: **Frederico Caeiro**, Universidade Nova de Lisboa, Portugal
Co-authors: M. Ivette Gomes

The estimation of a non-positive extreme value index, the shape parameter of the extreme value distribution, is dealt with under a semi-parametric framework. We consider a recent class of estimators of such a parameter. Apart from the usual integer parameter k, related with the number of top order statistics involved in the estimation, these estimators depend on an extra real parameter q, which makes them highly flexible and possibly second-order unbiased for a large variety of models. We are interested in the study a bootstrap algorithm for the adaptive estimation of the extreme value index.

E297: **Estimation of the second order parameter for heavy-tailed distributions**
Presenter: **Stephane Girard**, Inria, France
Co-authors: Laurent Gardes, El Hadji Deme

The extreme-value index is an important parameter in extreme-value theory since it controls the first order behavior of the distribution tail. Numerous estimators of this parameter have been proposed especially in the case of heavy-tailed distributions, which is the situation considered here.

Most of these estimators depend on the largest observations of the underlying sample. Their bias is controlled by the second order parameter. In order to reduce the bias of extreme-value index estimators or to select the best number of observations to use, the knowledge of the second order parameter is essential. We propose a simple approach to estimate the second order parameter leading to both existing and new estimators. We establish a general result that can be used to easily prove the asymptotic normality of a large number of estimators proposed in the literature or to compare different estimators within a given family. Some illustrations on simulations are also provided.

E514: **Bias corrected estimators for estimating high order quantiles**
Presenter: **Laura Cavalcante**, Universidade do Porto, Portugal
Co-authors: Margarida Brito, Ana Freitas
The problem of reducing the bias of the estimators has received considerable interest in the recent scientific studies. In order to improve the performance of the geometric-type tail index estimator, two geometric-type bias corrected estimators are proposed and their asymptotic behaviour is studied. These estimators are used to estimate high order quantiles. The proposed estimators are applied to the analysis of real data sets and their performance is illustrated using simulated data.

E621: **Adapting extreme value statistics to financial time series: Bias correction and serial dependence**
Presenter: **Chen Zhou**, Erasmus University Rotterdam, Netherlands
Co-authors: Laurens de Haan, Cecile Mercadier
The aim is to study bias correction method in extreme value analysis when observations exhibit weakly serial dependence, namely the $\beta-$mixing series. Firstly, for estimating the extreme value index, an asymptotically unbiased estimator is proposed and its asymptotic normality under the $\beta-$mixing condition is proved. The bias correction procedure and the dependence structure have an interactive impact on the asymptotic variance of the estimator. Secondly, an unbiased estimator of high quantiles is constructed. Thirdly, simulation is used to illustrate the finite sample performance of the estimators. Finally, the method is applied to real financial data for estimating the Value-at-Risk.

E275: **A reduced-bias mean-of-order-p Value-at-Risk estimator**
Presenter: **Ivette Gomes**, University of Lisbon, Portugal
Reduced-bias (RB) estimators of the Value-at-Risk at a level q (VaR_q), a value, high enough, so that the chance of an exceedance of that value is equal to q, small, are crucial in the most diverse fields of research. Dealing only with heavy tails, we try improving the performance of the existent RB VaR-estimators. High quantiles depend on the extreme value index (EVI), and recently, new classes of RB mean-of-order-p (RBMOP) EVI-estimators have been introduced. Often, either implicitly or explicitly, they also depend on the extremal index (EI), and the use of RB EI-estimators is also advised. The use of the RBMOP EVI-estimators in quantile-estimation enables us to introduce new and interesting classes of RBMOP VaR-estimators. The asymptotic distributional properties of the proposed classes of estimators are derived and the estimators are compared with alternative ones, not only asymptotically, but also for finite samples through Monte Carlo simulation techniques. An application to the log-exchange rates of the Euro against the Sterling Pound is also provided.

ES33 Room Bloomsbury STATISTICS IN FUNCTIONAL AND HILBERT SPACES Chair: Gil Gonzalez-Rodriguez

E046: **An Fmax-test for one-way ANOVA for functional data**
Presenter: **Jin-Ting Zhang**, National University of Singapore, Singapore
Co-authors: Ming-Yen Cheng, Chi-Jen Tseng, Hau-Tieng Wu
A new global test, namely Fmax-test, for the one-way ANOVA problem for functional data is discussed. The Fmax-test statistic is the maximum value of the usual pointwise F-test statistic over a given interval. We discuss the asymptotic random expression of the test statistic and study its asymptotic power. We show that the Fmax-test is $\sqrt{n}-$consistent. Via some simulation studies, we show that in terms of size-controlling and power, the Fmax-test outperforms some existing testing procedures when the functional data are highly and moderately correlated. We illustrate the Fmax-test using a real data example.

E090: **Random projections and functional data analysis**
Presenter: **Juan A Cuesta-Albertos**, Universidad de Cantabria, Spain
A review of some applications of the random projection method in functional data analysis is presented. This procedure consists in, instead of testing a given null hypothesis in a functional space, testing (the transformation of) this hypothesis on a one-dimensional randomly chosen projection, thus taking advantage of the multiple procedures which are available in the one-dimensional case. We will see how this idea can be applied to some problems like two-sample goodness of fit tests, normality tests and multiway ANOVA models including covariates. We will also analyze a test for the linear model in functional spaces.

E277: **Functional graphical models**
Presenter: **Gareth James**, University of Southern California, United States
Recently there has been a great deal of interest in constructing graphical models, or networks, from high-dimensional data. A graphical model is used to depict the conditional dependence structure among p different random variables. Such a network consists of p nodes, one for each variable, and a number of edges connecting a subset of the nodes. Two variables connected by an edge are conditionally correlated with each other. The aim is to construct a functional graphical model depicting the conditional dependence structure among p random functions. Functional data of this sort can arise in a number of contexts. For example, rather than observing only a static set of p gene expression levels at a single point in time, it is now becoming common to observe multiple expression levels over time. A convex penalized criterion is proposed which has connections to both the graphical lasso and the group lasso. Minimizing the functional graphical lasso criterion using an ADMM algorithm efficiently produces a sparse estimate for the network structure. It will be demonstrated that this functional graphical lasso approach has a number of advantages over other possible approaches.

E369: **On extensions of the band depth to multivariate functional data**
Presenter: **Sara Lopez Pintado**, Columbia University, United States
A new notion of depth for multivariate functional data that is an extension of the univariate functional band depth is proposed. The proposed simplicial band depth provides a simple and natural criterion to measure the centrality of a trajectory within a sample of curves. Based on this depth, a sample of multivariate curves can be ordered from the center outward and order statistics can be defined. Properties of the simplicial band depth, such as invariance and consistency, can be established. A simulation study shows the robustness of this new definition of depth and the advantages of using a multivariate depth versus the marginal depths for detecting outliers. Real data examples from growth curves and signature data are used to illustrate the performance and usefulness of the proposed simplicial band depth.

E409: **An application of functional singular component analysis in human nutrition**
Presenter: **Christian Ritz**, University of Copenhagen, Denmark
In dietary studies intervention effects may be recorded by means of a number of different outcomes, e.g., blood parameters and visual analogue scores. Outcomes are typically measured repeatedly over time on the same subjects (5-20 times), 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 and they suffice for the primary evaluation of the intervention effect, but they do not fully allow for a multivariate appreciation of the results in terms of understanding to which degree the different outcomes reflect the same trends in the data. Therefore, it is often also of interest to understand how the different outcome profiles over time are mutually correlated. A recently introduced approach is introduced, based on 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. Some implementation details are also provided. Results from the analysis of a dietary study are presented.

ES35 Room Gordon REGULARIZATION METHODS AND ALGORITHMS Chair: Patrick J. F. Groenen

E814: **Regularized generalized canonical correlation analysis for multiblock and multigroup data analysis**
Presenter: **Arthur Tenenhaus**, Supelec, France
Co-authors: Michel Tenenhaus

On the one hand, multiblock data analysis concerns the analysis of several sets of variables (blocks) observed on the same set of individuals. The main aim is to investigate the relationships between blocks. On the other hand, multigroup data analysis concerns the analysis of one set of variables observed on a set of individuals taking into account a group- structure at the level of the individuals. The main aim is to investigate the relationships among variables within the various groups. Through a single optimization problem, an overview of methods for the analysis of data structured in blocks of variables or groups of individuals is presented. More specifically, regularized generalized canonical correlation analysis (RGCCA), which is a unifying approach for multiblock data analysis, is extended so that it can also be a unifying tool for multigroup data analysis. The versatility and usefulness of our approach are illustrated on two real data sets.

E834: **Flexible multiclass support vector machines: An approach using iterative majorization and Huber hinge errors**
Presenter: **Gertjan van den Burg**, Erasmus University Rotterdam, Netherlands
Co-authors: Patrick Groenen

A flexible multiclass support vector machine (SVM) is proposed which can be used for classification problems where the number of classes $K \geq 2$. Traditional extensions of the binary SVM to multiclass problems such as the one-vs-all or one-vs-one approach suffer from unclassifiable regions. This problem is avoided in the proposed method by constructing the class boundaries in a $K-1$ dimensional simplex space. Nonlinear classification boundaries can be constructed by using either kernels or spline transformations in the method. An Iterative Majorization algorithm is derived to minimize the constructed flexible loss function. The performance of the method is measured through comparisons with existing multiclass classification methods on several datasets. From this we find that in most cases the performance of the proposed method is similar to that of existing techniques, but in some cases classification accuracy is higher.

E836: **Weighted sparse PCA**
Presenter: **Katrijn van Deun**, KU Leuven, Belgium
Co-authors: Iven van Mechelen, Johan Westerhuis, Age K. Smilde

Sparse PCA has been shown to be a fruitful method for the analysis of high-dimensional data, including biclustering and network component analysis as special cases. However, a limiting feature of current approaches to sparse PCA is that they do not allow us to weight the values of a particular variable for a particular observation. Such weighting may be useful, for example, to downweight unreliable values or to account for missing values. Here, we propose a weighted sparse PCA method to offer the flexibility to weight particular values. The method has been developed within a regularized weighted least-squares framework that imposes sparsity on the variable loadings and/or component scores through the lasso (ordinary, adaptive, or randomized). Our method also allows for components to be non-orthogonal. To illustrate the importance of weighting and to evaluate the performance of the method, we illustrate it with simulated data stemming from a challenge for network component analysis.

E1176: **Genetic risk scores using ridge regression**
Presenter: **Ronald de Vlaming**, Erasmus School of Economics, Netherlands
Co-authors: Patrick Groenen

Genome-wide association studies (GWASs) are used to explore the genetic architecture of outcomes such as diseases and behaviour. In GWASs the data often consist of $M = 2,500,000$ predictors called SNPs, observed for a limited number of respondents N, e.g. $N \leq 10,000$. In case the outcome variable is quantitative, classical multiple regression would be a natural method for predicting the outcome. However, ordinary least squares (OLS) is troubled by overfitting and multicollinearity. In fact, as $N \ll M$, the OLS estimator does not have a solution. The standard has become to estimate each SNP effect within a set of SNPs in a separate regression. The estimate of the outcome – the Genetic Risk Score – is the weighted linear sum using these regression weights. Instead, we propose to return to one regression using all SNPs as predictors, adding a penalty to overcome the problem of multicollinearity. We choose Ridge Regression (RR), which has a penalty term consisting of the sum of squared weights. We implement RR in a computationally efficient way suitable for GWASs. We compare the out-of-sample predictive performance of the RR and classical GWAS estimates.

E958: **Sparse CCA using sparse multivariate alternating regression**
Presenter: **Ines Wilms**, KU Leuven, Belgium
Co-authors: Christophe Croux

Canonical correlation analysis (CCA) describes the associations between two sets of variables by maximizing the correlation between linear combinations of the variables in each data set. However, in high-dimensional settings where the number of variables exceeds the sample size or when the variables are highly correlated, traditional CCA is no longer appropriate. A method for sparse CCA is proposed which simultaneously performs model estimation and variable selection. As such, sparse estimation produces linear combinations of only a subset of variables from each data set, thereby increasing the interpretability of the canonical variates. We covert the CCA problem into a multivariate regression framework, and combine a multivariate alternating regression approach together with a lasso penalty to obtain such sparse canonical vectors.

ES95 Room Athlone DISCRETE-VALUED TIME SERIES Chair: Robert Jung

E238: **Multivariate models for integer-valued time series**
Presenter: **Maria Eduarda Silva**, University of Porto and CIDMA, Portugal

Time series of counts are available in a wide variety of fields and the need to analyse such data adequately led to a multiplicity of approaches and a diversification of models that explicitly account for the discreteness of the data. One approach is to replace the multiplication in the conventional ARMA models by an appropriate random operator, denominated thinning operator, producing the so called INARMA models. In the context of univariate time series of counts, the class of INARMA models has been widely studied in the literature. However, for multivariate time series of counts several difficulties arise and the literature is not so elaborate. We address the problem of modelling multivariate time series of counts. Models based on thinning operations for bivariate time series of counts are introduced and their statistical and probabilistic properties discussed. The models and methods are illustrated in simulated and real data sets.

E243: **Hierarchical Bayesian spatio-temporal Conway-Maxwell Poisson models with dynamic dispersion**
Presenter: **Scott Holan**, University of Missouri, United States
Co-authors: Guohui Wu, Christopher Wikle
Modeling spatio-temporal count processes is often a challenging endeavor. That is, in many real-world applications the complexity and high-dimensionality of the data and/or process do not allow for routine model specification. For example, spatio-temporal count data often exhibit temporally-varying over/underdispersion within the spatial domain. In order to accommodate such structure, while quantifying different sources of uncertainty, we propose a Bayesian spatio-temporal Conway-Maxwell Poisson (CMP) model with dynamic dispersion. Motivated by the problem of predicting migratory bird settling patterns, we propose a threshold vector-autoregressive model for the CMP intensity parameter that allows for regime switching based on climate conditions. Additionally, to reduce the inherent high-dimensionality of the underlying process, we consider nonlinear dimension reduction through kernel principal component analysis. Finally, we demonstrate the effectiveness of our approach through out-of-sample one-year-ahead prediction of waterfowl migratory patterns across the United States and Canada. The proposed approach is of independent interest and illustrates the potential benefits of dynamic dispersion in terms of superior forecasting.

E251: **Periodic self-excited threshold integer-valued autoregressive processes with periodic structure**
Presenter: **Isabel Pereira**, University of Aveiro, Portugal
Co-authors: Raquel Nicolette, Manuel Scotto
Periodically correlated processes play an important role in the analysis of a variety of data sets. Periodic integer-valued autoregressive models of order one, with period t, driven by a sequence of independent Poisson distributed random variables, are considered. The periodic INAR(1) is extended to the threshold case with two regimes. Basic probabilistic and statistical properties of these models are discussed. Moreover, parameter estimation is also addressed and the performance of several estimation methods is compared by a simulation study.

E685: **Forecasting periodic discrete-valued time series**
Presenter: **David Matteson**, Cornell University, United States
A method for forecasting that combines discrete-valued time series models with a dynamic latent factor structure is introduced. The factor structure models the observed non-stationary patterns in periodic data and greatly reduces the number of model parameters. The factor model is combined with stationary discrete-valued time series models to capture the remaining serial dependence in the intensity process. We apply Bayesian estimation methods to forecast and conduct inference for applications in staffing and manufacturing.

E650: **Modelling count conditional distributions: methodology and appcations**
Presenter: **Robert Jung**, University of Hohenheim, Germany
The purpose is using different modelling approaches to specification, estimation and diagnostic checking for time series models of (low) counts. The methods used are especially well suited to modelling stock-type data, rather than count incidence data. Rarely, if ever, is there just a single approach to modelling that might be adopted. The use of two classes of models and their associated conditional distributions is considered in detail: integer autoregressive models (with regression effects); and those from the autoregressive conditional mean family. Both are used in conjunction with three novel data sets that differ according to dispersion, serial correlation and conditional mean variation due to exogenous components. Parameter estimation is generally done by maximum likelihood methods and a rigorous attempt is made to examine the adequacy of fitted models. A range of different models transpire to be preferred across the set of applications. We conclude that model specification search in this field needs to entertain different types of models for good subsequent application of forecasting or prediction which is considered to be a principal aim of modelling.

ES44 Room Bedford SEMI AND NON PARAMETRIC BAYESIAN METHODS FOR COMPLEX MODELS Chair: Brunero Liseo

E464: **Bayesian nonparametric modeling of stratified survival data**
Presenter: **Michele Guindani**, MD Anderson Cancer Center, United States
Co-authors: Bernardo Nipoti, Alejandro Jara
Stratified Cox proportional models are routinely used in survival analysis to estimate the underlying hazard functions when they are supposed to vary among a set of groups (strata), e.g. as a way to account for nonproportional hazards in the all population. Proportional hazard models are discussed where the baseline hazard is modeled non-parametrically as a random hazard mixture. In particular, it focuses on the dependence structure that arises, in the multivariate setting, from different choices of mixing measures. It is shown that hazard mixtures with a stable mixing measure feature appealing properties for modeling the dependence structure of the data. The model allows for exact posterior estimation for the parameters of interest. The methodological framework is illustrated by means of a simulation study as well as an application to a set of clinical data.

E535: **Cluster analysis of curved-shaped data with species-sampling mixture models**
Presenter: **Alessandra Guglielmi**, Politecnico di Milano, Italy
Co-authors: Raffaele Argiento, Andrea Cremaschi
An interesting problem is clustering of data whose support is "curved" (in an Euclidean space). To deal with it, two Bayesian nonparametric mixture models are introduced, both exploiting the "geometry" of the dataset. The first combines two ingredients: species sampling mixtures of parametric densities on one hand, and a deterministic clustering procedure (DBSCAN) on the other. The model is called b-DBSCAN: in short, two observations share the same cluster if the distance between the densities corresponding to their latent parameters is smaller than a threshold. The second model is given hierarchically, with data in each cluster parametrically distributed around a curve (principal curve), and the prior cluster assignment given on the latent variables at the second level of hierarchy according to a species sampling model. More specifically, conditioning on the "cluster parameter", data in each cluster are i.i.d. according to a parametric distribution centered around a principal curve, which here is a penalized spline (other choices could be investigated). Cluster estimates from these two mixture models are compared to those from competitor cluster algorithms for a simulated bivariate dataset from two clusters, one being banana-shaped.

E537: **Approximate Bayesian inference for discretely observed continuous-time multi-state models**
Presenter: **Andrea Tancredi**, Sapienza University of Rome, Italy
Inference for continuous time multi-state models presents considerable computational difficulties when the process is only observed at discrete time points with no additional information about the state transitions. In fact, for a general multi-state Markov model, the evaluation of the likelihood function is possible only via intensive numerical approximations. Moreover, in real applications, transitions between states may depend on the time since entry into the current state and semi-Markov models, where the likelihood function is not available in closed form, should be fitted to the data. Approximate Bayesian Computation (ABC) methods, which make use only of comparisons between simulated and observed summary statistics, represent a solution to intractable likelihood problems and provide alternative algorithms when the likelihood calculation is computationally too costly. The aim is to investigate the potentiality of ABC techniques for multi-state models by means of a real data example.

E688: **Frequentist properties of empirical Bayes approaches**
Presenter: **Judith Rousseau**, CREST-ENSAE and University Paris Dauphine, France
Co-authors: Sophie Donnet, Vincent Rivoirard, Catia Scricciolo
Some generic conditions to derive the posterior concentration rate in empirical Bayes procedure are given. We then apply this Theorem to the case of Dirichlet mixtures of Gaussians in the density estimation problem and to the estimation of a monotone non increasing intensity in the case of Aalen multiplicative intensity Poisson process models. Some simulations are provided to compare empirical Bayes with fully Bayes procedures.

ES45 Room Court CONSTRUCTION OF EXPERIMENTAL DESIGNS Chair: Jesus Lopez-Fidalgo

E144: **Construction of experimental designs for estimating models with flexible regressors**
Presenter: **Alexander Donev**, University of Manchester, UK
Co-authors: Liam Brown, Camilla Sammut-Powell
Statistical models with flexible regressors have useful applications, but their successful estimation depends much on the quality of the available data, which have to allow for simultaneous estimation of the model parameters and the form of the regressors. We propose a method for constructing experimental designs for such situations. We also show that this approach can also be used to construct designs suitably optimized for model selection.

E171: **Approximate Bayesian computation design**
Presenter: **Werner Mueller**, Johannes Kepler University Linz, Austria
Co-authors: Markus Hainy, Henry Wynn
A new technique of generating optimal designs by means of simulation is proposed. The method combines ideas from approximate Bayesian computation and optimal design of experiments and allows great flexibility in the employed criteria and models. We illustrate the idea by simple expository examples.

E172: **Optimal design for smooth supersaturated models**
Presenter: **Henry Wynn**, London School of Economics, United Kingdom
Smooth supersaturated models (SSM) are polynomial models which have a higher polynomial degree than the standard polynomial models, but in which the extra degrees of freedom are used to increase smoothness. This can be done by having dummy design points or, equivalently, by having extra parameters. The theory of algebraic statistics is used to guide the choice of extra monomial terms. In the limit, as the degree goes to infinity, the models tend to spline type (thin plate spline) models, known to have optimal smoothness properties. We seek optimal experimental designs for these models and hence provide a computational framework for obtaining approximate solutions to the intractable problem of optimal designs for spline-type models over any domain. The relationship between the support of the optimal designs and the optimal placement of knots provides additional interest.

E423: **Sequential augmentation of designs**
Presenter: **Hugo Maruri**, QMUL, United Kingdom
Latin hypercube designs have become a standard tool for the design of computer experiments. However, a potential disadvantage is the relative difficulty of adding extra points. An alternative to sequential design is proposed, using properties of low discrepancy sequences. The proposal has the advantage of allowing both an initial design and further extra points as desired. Its performance is evaluated and commented.

E664: **Bayesian D-optimal choice designs for mixtures**
Presenter: **Aiste Ruseckaite**, Erasmus University Rotterdam, Netherlands
Co-authors: Peter Goos, Dennis Fok
Consumer products and services can often be described as mixtures of ingredients. For example, the mixture proportions in a cocktail or the mix of waiting time, in-vehicle and out-of-vehicle travel time in a transportation setting. However, there are no studies on (the optimal design of) choice experiments involving mixtures. The aim is to introduce mixture models in the choice context and present an algorithm to construct Bayesian D-optimal experimental designs when the multinomial logit model is used to represent preferences. Both locally and Bayesian D-optimal designs are considered, to allow the use of prior knowledge. In the former, a design is constructed given a particular parameter vector, while in the latter case uncertainty in the prior knowledge is explicitly incorporated. The performance of the designs is compared to the utility neutral (UN) approach. It is demonstrated that the proposed designs differ in look and in general perform better than the UN designs. The proposed designs are robust in the sense that if the UN design performs better, the difference is not large. Further, it is documented that the optimal design often shows evidence of approximate utility balance, whereas this is not necessarily so for the UN design.

ES100 Room Chancellor's RANDOM FORESTS AND RELATED METHODS: THEORY AND PRACTICE Chair: Jean-Michel Poggi

E207: **Analysis of purely random forests bias**
Presenter: **Robin Genuer**, ISPED Universite Bordeaux Segalen, France
Co-authors: Sylvain Arlot
Random forests (r.f) are a very effective and commonly used statistical method, but their full theoretical analysis is still an open problem. As a first step, simplified models such as purely random forests (p.r.f) have been introduced, in order to shed light on the good performance of r.f. We study the approximation error (the bias) of some p.r.f. models in a regression framework, focusing in particular on the influence of the number of trees in the forest. Under some regularity assumptions on the regression function, we show that the bias of a large enough forest decreases at a faster rate (w.r.t. the size of each tree) than a single tree. As a consequence, large forests attain a better risk rate than single trees. For example, in a simple case with one-dimensional inputs, when the regression function is C^2, the (p.r.) forest estimator reaches the minimax rate of convergence $n^{-4/5}$, whereas the single tree estimator reaches the rate $n^{-2/3}$.

E532: **Variable importances in forests of randomized trees: some theory and application to gene network inference**
Presenter: **Pierre Geurts**, University of Liege, Belgium
One of the most interesting features of Random Forests models is their ability to derive variable importance scores that allow us to rank the input variables according to their relevance for predicting the output. Two importance measures have been proposed in the literature: the mean decrease of impurity (MDI) importance and the permutation importance. These measures have been successfully applied on many problems but they are not well understood from a theoretical point of view. The aim is to first present a theoretical characterization of the MDI importances as measured by an ensemble of totally randomized trees in asymptotic conditions. In particular, it is demonstrated that MDI importances of totally randomized trees exhibit desirable properties for assessing the relevance of a variable: they are equal to zero only for irrelevant variables and they depend only on the relevant variables. We then present the application of the MDI measure in the context of the GENIE3 method for the inference of gene regulatory

networks. In particular, the results in the context of the DREAM4 and DREAM5 network inference challenges are briefly reported, where GENIE3 was acknowledged as the best performer.

E194: **Random forest for feature selection**
Presenter: **Witold Rudnicki**, University of Warsaw, Poland
Co-authors: Miron Kursa, Mariusz Wrzesien, Wieslaw Paja
The aim is to review methods for all-relevant feature selection (ARFS) based on random forests. The ARFS problem will be introduced. A description of the differences between ARFS and more usual minimal optimal feature selection and a motivation for development of this feature selection will be provided. Then a short review of the existing approaches will be presented. A more detailed description of our approach will follow – the description of the algorithm will be given, as well as a detailed discussion of the results obtained for a synthetic data set, when the important attributes are known a priori. A few applications to the real-world data (from the UCI database) will be presented with discussion of possible methods for controlling the level of false positive discoveries. Finally, a brief discussion of possible developments and applications will be given.

E132: **Imputation methods in mixed datasets: Random forests versus PCA**
Presenter: **Julie Josse**, Agrocampus, France
Co-authors: Francois Husson, Vincent Audigier
Missing values are ubiquitous in the statistical practice. They are problematic, since most statistical methods cannot be applied directly on an incomplete data set. A common solution to deal with missing values consists of using single imputation methods. We compare two new imputation methods for mixed data: a method which predicts the missing entries using a principal components analysis method dedicated to mixed data and a random forests-based method. Both methods take into account the relationship between the continuous and the categorical variables for the imputation. The methods are evaluated from a theoretical point of view as well as on simulated and real data sets. The random forest method provides good results when the relationships between variables are non-linear, but poorest results for the categorical variables. We then point out the major problem of the underestimation of the variability of the single imputation methods and discuss multiple imputation solutions.

E258: **Random ferns: A random forest variant for special operations**
Presenter: **Miron Kursa**, University of Warsaw, Poland
Random ferns is a simple and deeply stochastic machine learning algorithm created for specific high-throughput computer vision applications. A general-purpose implementation of this algorithm, rFerns, is presented. This algorithm has been designed to be a drop-in replacement of the random forest method. Next, the properties of this method are discussed, especially its relevance to the analysis of a large amount of high dimensional, noisy data and theoretical aspects of its practically optimisation-less training. Finally, two important applications, all-relevant gene selection and real-time audio analysis, will be presented. In both applications random ferns achieve accuracy indistinguishable from random forest while significantly reducing their computational cost.

ES76 Room Torrington MATHEMATICAL ASPECTS OF COPULAS I Chair: Fabrizio Durante

E195: **Dependence in multivariate Levy processes revisited**
Presenter: **Oliver Grothe**, University of Cologne, Germany
Multivariate Levy processes are tools suitable for modeling risk when jumps and sudden large values may occur. Since these processes are entirely specified by their distribution over one fixed time horizon, the most suggesting way of modeling their dependence is to choose a time horizon and to combine the marginal distributions of the process by a copula function. However, although this approach seems simple, it has a number of serious drawbacks. The copula has to be chosen such that the resulting multivariate distribution is infinitely divisible, which may be very complicated to ensure. Furthermore, the distribution of the process on other time horizons may become intractable if it is not stable under convolution. For these reasons, the concept of Levy copulas, which does not suffer from the mentioned drawbacks, has been introduced. Unfortunately, the number of parametric Levy copulas presented in the literature is rather small, which limits their use in risk management applications. We propose an alternative way of modeling dependence in multivariate Levy processes which combines the advantages of both mentioned approaches. It is based on the idea of Levy copulas, but uses common copula functions to ensure a rich variety of models. Furthermore, any marginal Levy processes and any copula may be combined to a proper multivariate Levy process. We provide estimation and simulation algorithms in detail.

E265: **Very singular copulas**
Presenter: **Wolfgang Trutschnig**, ECSC, Spain
The main objective is to show how surprisingly singular (from the purely analytical point of view) two-dimensional copulas (doubly stochastic measures) and their conditional distribution functions may be. After recalling the construction of copulas via Iterated Function Systems (IFS) we consider IFSs induced by so-called modifiable transformation matrices T and use tools from Symbolic Dynamical Systems and Ergodic Theory to construct mutually singular copulas $A_T^\star$ with identical (possibly fractal or full) support that are at the same time singular with respect to the Lebesgue measure λ_2 on $[0,1]^2$. The established results are then utilized for a simple proof of the existence of families of singular copulas $A_T^\star$ with full support for which all conditional distribution functions $y \mapsto F_x^{A_T^\star}(y)$ are continuous, strictly increasing and have derivative zero *lambda*-almost everywhere, i.e. all conditional distribution functions are strictly singular in the classical sense.

E318: **Nonlinear modelling: Copulas versus econometrics**
Presenter: **Piotr Jaworski**, University of Warsaw, Poland
There are two main approaches in the multivariate modelling. One is symmetric. It is based on a copula describing the interdependencies and distributions of individual variables. The second is asymmetric. The dependent variable is given as a function of explanatory variables and an independent noise factor. The aim of the study is to investigate the link between these two methods.

E638: **A multivariate technique based on conditional copula for the imputation of complex dependent data**
Presenter: **Marta Di Lascio**, Free University of Bozen-Bolzano, Italy
Co-authors: Simone Giannerini
Imputation of missing data is a crucial task in statistical practice and the problem poses several challenges, like data sets with variables of different type, non-Gaussian data or large data sets. An approach based on conditional copulas that allows us to avoid most of these issues is proposed. The basic idea is to derive the conditional densities for each incomplete variable given the complete ones through the corresponding conditional copulas and, then, impute missing values by drawing observations from them. The proposal can be applied to multivariate missing with generic (monotone or non-monotone) dependence pattern. Also, an R software package, called CoImp, that implements the method is provided. The CoImp aims to impute missing data by preserving their possible complex dependence structure. It is appealing since it does not restrict the conditional distributions to being Gaussian and allows us to model multivariate distributions with different margins through a flexible semi-parametric approach. The advantages of the CoImp over classical imputation techniques are shown in a simulation study; also, applications to real data sets are presented.

E223: **Asymptotics and multiplier bootstrap of the sequential empirical copula process under strong mixing**
Presenter: **Ivan Kojadinovic**, University of Pau, France
Co-authors: Axel Buecher
Two key ingredients to carry out inference on the copula of multivariate observations are the empirical copula process and an appropriate resampling scheme for the latter. Among the existing techniques used for i.i.d. observations, a multiplier bootstrap frequently appears to lead to inference procedures with the best finite-sample properties. An extension of this technique to strictly stationary strongly mixing observations has been recently proposed by adapting a *dependent* multiplier bootstrap to the empirical copula process. The main contribution is a generalization of the multiplier resampling scheme along two directions. First, the resampling scheme is now genuinely sequential, thereby allowing to transpose to the strongly mixing setting all of the existing multiplier tests on the unknown copula, including nonparametric tests for change-point detection. Second, the resampling scheme is now fully automatic as a data-adaptive procedure is proposed which can be used to estimate the bandwidth (block length) parameter.

ES81 Room Woburn CURE MODELS AND COMPETING RISKS IN SURVIVAL ANALYSIS Chair: Ingrid Van Keilegom

E107: **Flexible Bayesian cure survival models**
Presenter: **Vincent Bremhorst**, Universite Catholique De Louvain, Belgium
Co-authors: Philippe Lambert
A common hypothesis in the analysis of survival data is that any observed unit will experience the monitored event if it is observed for a sufficient long time. Alternatively, one can explicitly acknowledge that an unknown and unidentified proportion of the patient population under study is cured and will never experience the event of interest. The promotion time model, which is motivated using biological mechanisms in the development of cancer, is one of the survival models taking this feature into account. The promotion time model assumes that the failure time of each subject is generated by the minimum of N latent event times which are independent with a common distribution $F(t) = 1 - S(t)$ independent of N. We propose an extension which allows the covariates to influence simultaneously the probability of being cured and the latent distribution $F(t)$. We estimate the latent distribution $F(t)$ using a flexible Cox proportional hazard model where the logarithm of the baseline hazard function is specified using Bayesian P-splines. The identification issues of the related model are also investigated. A simulation study evaluating the accuracy of our methodology is presented. The proposed model is illustrated on the data from the phase III Melanoma e1684 clinical trial.

E306: **Resampling-based inference with the Aalen-Johansen estimator of cumulative incidence functions of competing risks**
Presenter: **Markus Pauly**, Heinrich-Heine University Duesseldorf, Germany
Time-simultaneous inference for time-to-event data is often hampered by the unknown limiting covariance structure. In practice such situations are often attacked by Lin's wild bootstrap technique: There, for fixed data, standard normal multipliers are introduced into an appropriate version of the test statistic that originates from a martingale representation. Finally, the unknown finite sample distribution of the test statistic is approximated via Monte Carlo by repeatedly generating standard normal variates. Theoretical properties of resampling techniques (including Lin's wild bootstrap) for nonparametric inference on the cumulative event probability of a competing risk are analyzed.

E591: **Regression modelling with missing causes of death: Assessing the sensitivity of inferences to missingness assumptions**
Presenter: **Aurelien Latouche**, Conservatoire National des Arts et Metiers, France
Co-authors: Margarita Moreno
There are various regression strategies for modeling the cause-specific hazards and the cumulative incidence function when some causes of death are missing. However, most approaches found in the literature rely on the assumption that the causes of death are missing at random (MAR), that is, that the probability of missingness is independent of the cause of death when conditioning on the observed data. Unfortunately, this assumption can never be verified from the data available. This issue, common to all missing data problems, has led to an increasing awareness of the need to perform sensitivity analyses to assess the robustness of inferences to departures from the MAR assumption. However, no standard method exists for setting up such an analysis, each specific scientific context requires different considerations and this is still an active area of research. The aim is to first present regression models with missing causes of death and propose a flexible procedure to perform sensitivity analyses following an initial MAR analysis. The methodology relies on the pattern-mixture model factorization of the full data likelihood and allows the analyst to formulate assumptions about the missing data distribution in an explicit manner. The study illustrates the practical value of the approach and underlines the need for sensitivity analyses when analyzing competing risks data with missing causes of death.

E671: **Vertical modeling: Analysis of competing risks data with a cured proportion**
Presenter: **Mioara Alina Nicolaie**, Catholic University of Louvain, Belgium
Co-authors: Catherine Legrand
The aim is to extend vertical modeling approach for the analysis of survival data with competing risks to incorporate a cured fraction in the population, that is a proportion of the population for which none of the competing events occur. Two main scenarios are considered: 1. cure is observed and 2. cure is latent. In essence, the proposed method combines a model for the heterogeneous surviving fraction corresponding to cured individuals with the vertical modeling approach corresponding to susceptible individuals to the competing events. The most appealing fact of the method is that it appears as a natural extension to competing risks of the standard mixture cure model in ordinary survival; thus, causes of failure are assigned only in case a failure occurs. This contrasts with the existing mixture cure model for competing risks of Larson and Dinse, which conditions at the onset on the future causes of failure presumably attained. Regression parameter estimates of the model are obtained by means of an EM-algorithm technique. The performance of the estimators is analyzed in a simulation study. The method is illustrated using a real data set on melanoma cancer.

E803: **The SIMEX method for correcting the bias in a survival cure model with mismeasured covariates**
Presenter: **Ingrid van Keilegom**, Universite catholique de Louvain, Belgium
Co-authors: Aurelie Bertrand, Catherine Legrand
In traditional survival analysis, all subjects in the population are assumed to be susceptible to the event of interest: every subject has either already experienced the event or will experience it in the future. In many situations, however, a fraction of individuals will never experience the event: they are considered to be cured. The promotion time cure model is one of the survival models taking this feature into account. We consider the case where some of the explanatory variables in the model are supposed to be subject to measurement error. This measurement error should be taken into account in the estimation of the model, to avoid biased estimators of the model. We propose an adaptation of the SIMEX method to the promotion time cure model. We show that the proposed estimator is asymptotically normally distributed. We also show via simulations that the suggested method performs well in practice by comparing it with a method already proposed which considers a corrected score approach for this problem. We compare the two approaches in the simulation as well as in more practical settings, whose features (e.g. cure rate, censoring rate, maximal censoring time) are closer to what is likely to be observed in real data.

ES109 Room 349 CONTRIBUTIONS IN BIOSTATISTICS AND BIOINFORMATICS Chair: Giovanni Montana

E1053: **Biological marker selection for detecting gene-gene interaction in genome-wide association studies**
Presenter: **Mathieu Emily**, Agrocampus Ouest - IRMAR, France
Co-authors: Chloe Friguet
Genome-wide association studies (GWAS) aim at detecting correlation between a phenotypic trait and a set of hundreds of thousands biological markers, called single nucleotide polymorphism or SNPs. As usual in genomics, data suffer from heterogeneity that generate dependency between SNPs. To account for spatial dependency due to linkage disequilibrium, the first step in GWAS, called tagging, consists in selecting the most informative markers to analyse a smaller set of markers. However usual tagging techniques are not designed to account for interaction. We propose a novel method to select a set of tag-SNPs that optimally represent the total set of all pairs of SNPs. The correlation between two pairs of SNPs is measured by the normalized mutual information. To demonstrate its feasibility, we apply our method to a set of simulated datasets obtained from a reference panel of individuals. Furthermore, the comparison of our method with existing tagging strategies proved that, on the one hand our method is powerful and that, on the other hand, it significantly decreases the proportion of false discoveries.

E1033: **Bayesian inference for stochastic SIR epidemic model**
Presenter: **Tewfik Kernane**, University of Sciences and Technology Houari Boumediene, Algeria
Co-authors: Hamid Elmaroufy, Sidali Becheket, Abdellah Ouddadj
In stochastic SIR epidemic models, there are two types of parameters, the contact rate and the reciprocal average infectious period. There are several methods to estimate from a discretization of the observation time used in this area as inference by maximum likelihood or martingale techniques. The problem lies in the fact that the transition probabilities are not valid in explicit form. By approximating the epidemic process in discrete time by a diffusion process we treated this estimation problem using the data augmentation approach. This leads to increase the low frequency data by the introduction of latent data between each pair of observations, simulated using the standard method of the Euler approximation of stochastic differential equations. This involves estimating missing values, in addition to the model parameters, where the missing data and parameters are treated as random variables. This task is performed by application of MCMC algorithm to update alternately missing data and parameter settings.

E1030: **ANOVA for ordinally scaled factors**
Presenter: **Jan Gertheiss**, University of Goettingen, Germany
In its simplest case ANOVA can be seen as a generalization of the $t-$test for comparing the means of a continuous variable in more than two groups defined by the levels of a discrete covariate, a so-called factor. Testing is then typically done by using the standard $F-$test. Here, we consider the special but frequent case of factor levels that are ordered. We propose an alternative test using mixed models methodology. The new test often outperforms the standard $F-$test when factor levels are ordered. We illustrate the proposed testing procedure in simulation studies and real data applications. Although focus is on one-factorial ANOVA here, we also discuss extensions to multi-factorial ANOVA and studies with repeated measurements.

E1006: **A study of upper limit of human longevity using the extreme value theory in Japan**
Presenter: **Nobutane Hanayama**, Shobi University, Japan
Co-authors: Kenta Murotani, Yoshiaki Shikata
Theories of ageing are roughly divided into two major groups. One is consisting of damage theories and the other is consisting of program theories. According to damage theories we age because our systems break down over time. Meanwhile, in the program theories, it is considered that we age because there is an inbuilt mechanism that tells us to die. If the damage theories are true, we can survive any longer by avoiding damaging our organism. If the program theories are true, on the other hand, we cannot survive longer than the upper limit of longevity with any effort. For examining whether the damage theories or the programming theories is true, data for deaths given by age and birth-period in Japan is analyzed using the extreme value theory and the extinct cohort method. From the results of fitting the binomial regression model with probabilities calculated from the generalized Pareto distribution to the data, the upper limit of human longevity is estimated from 107 to 128 years for male or from 119 to 159 years. This result supports the program theories for human longevity. It is also shown, however, that the limit has not been increasing for both male and female.

E1118: **Health care expenditures and time to death: Focus on a cardio vascular cohort**
Presenter: **Luca Grassetti**, University of Udine, Italy
Co-authors: Laura Rizzi
Population ageing and health care expenditures (HCE) increase are central issues in nowadays health economics. While there are evidences that population ageing (or time to death - TTD) causes costs growth, some literature is trying to investigate the true direction of this relationship both at micro and macro level. We try to disentangle this issue exploring the correlation between TTD and HCE considering a specific micro level dataset: 9,898 persons entering the cohort of cardiovascular patients after the first acute episode in the period 2002-2007. The entire study is based on administrative secondary data source. The two main interest variables are quite peculiar. Remaining lifetime is a three components mixture of sudden deaths (17.14%), survivors (60.62%) and deaths within the five years observational period (22.24%). HCE must be carefully studied because expenditure patterns are quite different when comparing acute and chronic patients. Specific models are used to study the relationship between TTD and HCE considering the effect of some "control variables" such as age, gender, presence of comorbidities, exceptions and so on. Moreover, both variables must be tested for endogeneity and geographical heteroskedasticity has to be deepened as probably connected with the highly concentrated distribution of health care service suppliers.

Saturday 14.12.2013 16:40 - 18:45 Parallel Session F – CFE

CSI02 Room Beveridge MODEL SELECTION AND INFERENCE Chair: Martin Wagner

C217: **Inference post-model selection**
Presenter: **Benedikt Poetscher**, University of Vienna, Austria
Model selection has an important impact on subsequent inference. Ignoring the model selection step in general leads to invalid inference. We review the effect of model selection on the distributional properties of post-model-selection estimators and discuss ways for performing valid inference post-model selection. Related results applying to shrinkage-type estimators (e.g., penalized least-squares estimators) will also be discussed.

C357: **Conditional predictive inference post model selection**
Presenter: **Hannes Leeb**, University of Vienna, Austria
A finite-sample analysis of predictive inference procedures after model selection in regression with random design is given. The analysis is focused on a statistically challenging scenario where the number of potentially important explanatory variables can be infinite, where no regularity conditions are imposed on unknown parameters, where the number of explanatory variables in a "good" model can be of the same order as sample size and where the number of candidate models can be of larger order than sample size. The performance of inference procedures is evaluated conditional on the training sample. Under weak conditions on only the number of candidate models and on their complexity, and uniformly over all data-generating processes under consideration, we show that a certain prediction interval is approximately valid and short with high probability in finite samples, in the sense that its actual coverage probability is close to the nominal one and in the sense that its length is close to the length of an infeasible interval that is constructed by actually knowing the "best" candidate model. Similar results are shown to hold for predictive inference procedures other than prediction intervals like, for example, tests of whether a future response will lie above or below a given threshold.

C932: **Empirical economic model discovery and theory evaluation**
Presenter: **David Hendry**, Oxford, United Kingdom
Economies are too high dimensional and wide sense non-stationary for all features of models to be derived by either prior reasoning or data modelling alone. Selecting a viable representation intrinsically involves empirical discovery jointly with theory evaluation. Automatic methods can formulate very general initial specifications with many candidate variables, long lag lengths, and non-linearities, while allowing for outliers and location shifts at every observation, then select congruent parsimonious-encompassing models. Theory-relevant variables are retained without selection while selecting other candidate variables. Under the null that the latter are irrelevant, by orthogonalizing with respect to the theory variables, estimator distributions of the theory-model's parameters are unaffected by selection, even for more variables than observations and for endogenous variables. Under the alternative, when the initial model nests the local data generating process, an improved outcome results from selection, allowing rigorous evaluation of any postulated models to ascertain their validity.

CS02 Room B34 MODELLING AND FORECASTING STOCK MARKET VOLATILITY Chair: Walter Distaso

C340: **Testing for zero jump intensity**
Presenter: **Valentina Corradi**, University of Warwick, United Kingdom
Co-authors: Mervyn Silvapulle, Norman Swanson
Financial returns and exchange rates are often modeled as jump diffusions or stochastic volatility jump diffusions. It is common practice to jointly estimate the parameters of the continuous time components, i.e. the parameters characterizing drift and variance, and the parameters of jump component, i.e. the intensity and the parameters characterizing the jump size density. However, if the intensity is zero, the parameters characterizing the jump size density are not identified. Available tests for jumps are performed over a finite time span: if we fail to reject the null for a given day, this does not imply that there are no jumps. On the opposite, if we sequentially perform the test over each day, the probability of falsely discover jumps, when there are none, will eventually approach one. This is because of the well known sequential testing size distortion problem. The objective is to suggest direct tests for the null of zero intensity versus the alternative of positive intensity. The test is performed using an increasing time span of observations. To the best of our knowledge this is the first test for (no) jump based on directly testing the intensity parameters. We shall consider two distinct cases. First, we introduce a model free test for zero intensity, which requires no parametric assumptions on the specification of the continuous time part or on the density of the jump size. The test is based on the sample third moment. Second, we consider the case in which we can obtain the closed form of the likelihood of a stochastic volatility jump diffusion. We introduce a score test for the null of at most one non-zero intensity parameter. This is accomplished via the Union-Intersection principle.

C397: **Nonparametric identification and estimation of Euler equations**
Presenter: **Sorawoot Srisuma**, University of Cambridge, United Kingdom
Co-authors: Juan Carlos Escanciano, Stefan Hoderlein, Arthur Lewbel, Oliver Linton
Nonparametric identification and estimation of consumption based asset pricing Euler equations are considered. This entails estimation of pricing kernels or equivalently marginal utility functions up to scale. The standard way of writing these Euler pricing equations yields Fredholm integral equations of the first kind, resulting in the ill-posed inverse problem. It is shown that these equations can be written in a form that resembles Fredholm integral equations of the second kind, having well posed rather than ill-posed inverses. Durables, habits, or both are allowed to affect utility. The usual method of solving Fredholm equations is extended to allow for the presence of habits. Using these results, it is shown with few low level assumptions that marginal utility functions and pricing kernels are locally nonparametrically identified, and conditions for finite set and point identification of these functions are given. Consistent nonparametric estimators for these functions are provided and associated limiting distributions, and an empirical application to US consumption and asset pricing data.

C370: **Perfect fit in asset pricing models with irrelevant risk factors**
Presenter: **Cesare Robotti**, Imperial College London, United Kingdom
Co-authors: Nikolay Gospodinov, Raymond Kan
The aim is to study some seemingly anomalous results that arise in estimation and evaluation of possibly misspecified and unidentified linear asset pricing models by maximum likelihood or optimal one-step GMM. Strikingly, when irrelevant factors are present, the models exhibit perfect fit, measured by the squared correlation between the model's fitted expected returns and the actual returns. Furthermore, it is shown that when the identification condition fails, the tests for correct model specification have asymptotic power that is equal to the nominal size. In other words, applied researchers will erroneously conclude (with high probability) that the model is correctly specified even when the degree of misspecification is arbitrarily large. In deriving this result, it is shown that the specification test based on the objective function can be equivalently rewritten as a rank test. A similar result is obtained for the likelihood ratio test of correct specification of the beta pricing model. The practical relevance of the results is demonstrated using simulations and an empirical application.

C592: **Seasonal ARFIMA model**
Presenter: **Liudas Giraitis**, Queen Mary University of London, United Kingdom
Co-authors: Karim Abadir, Walter Distaso
The objective is to introduce a new approach for seasonal modeling of stationary time series that allows long memory and does not suffer from the drawbacks of GARMA model. The suggested SARFIMA model has tractable analytical expressions for the spectral density and autocovariance functions and enables practical multi-pole modeling. The periodocity of the seasonal model is directly related to the frequency ω of the pole in the spectrum. It establishes the n-consistency of the estimator of ω and derives its asymptotic distribution. It also contains asymptotic results on estimation of the memory parameter of SARFIMA model. The results are applied in inference of regression model with seasonal SARFIMA errors.

C636: **Inference with few heterogenous clusters**
Presenter: **Rustam Ibragimov**, Imperial College London, United Kingdom
Co-authors: Ulrich Mueller
Consider inference with a small number of potentially heterogeneous clusters or groups of observations in a financial or economic model that exhibits dependence and heterogeneity of a largely unknown type. Suppose estimating the model on each cluster yields q asymptotically unbiased, independent Gaussian estimators with potentially heterogeneous variances. Following an earlier work by the authors, one can then conduct asymptotically valid inference with a standard t-test based on the q cluster estimators, since at conventional significance levels, the small sample t-test remains valid under variance heterogeneity. Two contributions are made. First, the aim is to establish a new corresponding small sample result for the two-sample t-test under variance heterogeneity. One can therefore apply the robust large sample t- statistic based inference also for comparisons of parameters between two populations, such as treatment and control groups or pre- and post-structural break (e.g., pre-crisis and crisis period) data. Second, a test for the appropriate level of clustering is developed, with the null hypothesis that clustered standard errors from a fine partition are correct, against the alternative that only q clusters provide asymptotically independent information. Applications of the new robust large sample inference approaches include, among others, robust analysis of treatment effects, structural breaks and the effects of crises on financial and economic markets, as well as the robust analysis of volatility and autocorrelation structure of financial and economic time series.

CS03 Room B33 QUANTITATIVE BEHAVIORAL FINANCE Chair: Jorgen-Vitting Andersen

C569: **Risk and return of mutual funds and stocks: A network analysis on funds holdings**
Presenter: **Giulia Rotundo**, Sapienza University of Rome, Italy
Co-authors: Anna Maria D'Arcangelis
The increasing interconnection of markets, their sensitivity to contagions and the recent overall fragility of the financial system raise some questions on risk due to the overlap and interlacement of investments. The aim is to explore the relationship among the equity exposure and both performance and risk of mutual funds investing in stocks of European companies. At the same time, the analysis represents an attempt to shed light on the relationship between stocks' position within the network and their future market performance and risks. It starts analyzing whether there is any evidence that funds' managers take common decision on trading, both for benchmark constraints and for style management purposes. It explores the network structure of both stocks and funds through the construction of a bipartite network and the projection of the available information on both the space of mutual funds and stocks traded in the stock exchange. Methods typical of complex networks are then applied. Cluster analysis also evidences club effects and the most relevant stocks on markets. The results of the research become the grist for new insights in the picture of portfolio management techniques and on the role of benchmarks in the mutual fund world.

C539: **Communication and pricing**
Presenter: **Jorgen Vitting Andersen**, Sorbonne, France
Co-authors: Serge Galam, Giulia Rotundo
A model on opinion formation of market participants is introduced. Through communication the market participants form their view of the price level of a given market. A feedback mechanism is introduced via the dependence of the opinion formation on the market performance as well as through a dependence of the pricing on the change in opinion. The model is able to account for the so-called 'stylized facts' of financial markets, and gives a possible explanation for the origin behind regime changes in financial markets.

C712: **Trading behaviour, portfolio performance and psychological attributes of individual investors**
Presenter: **Svetlana Gherzi**, University of Warwick, United Kingdom
Co-authors: Neil Stewart
Recent literature has turned to behavioural explanations to account for empirical deviations from models of modern finance. One observed phenomenon is the extraordinary high volume of equity trading, especially in spite of the normatively prescribed buy-and-hold strategy and the no trades theorems. Empirical evidence also suggests that some investors systematically beat the market while others are systematic losers even though according to the Efficient Market Hypothesis risk-adjusted performance should be random. The most prominent explanation to account for the observed phenomena has been overconfidence. We focus on the individual differences of UK based investors. Controlling for demographics, including age, wealth, income, marital status and number of dependents we find that these relationships are more complex than previously suggested. Individual investors' psychological attributes such as propensity to engage in avoidant behavior, objective and subjective financial expertise as well as cognitive and affective processes account for significant variation in the observed trading behavior, differences in risk-adjusted portfolio performance and risk preferences.

C734: **The role of emotions in financial decision making**
Presenter: **Magda Roszczynska-Kurasinska**, University of Warsaw, Poland
Co-authors: Katarzyna Samson, Andrzej Nowak, Jorgen Andersen, Mikolaj Biesaga
It is quite undisputable that human decisions do not depend on strict mathematical expectations. The process of decision making, even in a pure economic context of financial markets, beside the objective data, is affected by many psychological phenomena like memory , expectations, beliefs, cognitive biases (e.g., overconfidence, confirmation error, reverse thinking) and emotions. The presented research shows high susceptibility of decision makers to emotions which are the integral component of human nature. The experimental research in finance and psychology has made some attempt to explain the economically incomprehensible investment decisions with emotions like fear, euphoria, regret and depression but results from different research often appear to be contradictory. We explore the character of the relation by introduction of uniform methodology and categorization of emotions across experimental studies. Moreover, we investigate the duration and the strength of the impact on behavior showing patterns of interaction between emotions and decision in time. The experimental research (laboratory experiments with humans) is complemented with analysis of real financial and communication data. We extract emotions from communication and show how they relate to the short and long term behavior of investors.

C1027: **News, jumps and volatility: Evidence from energy markets**
Presenter: **Svetlana Borovkova**, Vrije Universiteit Amsterdam, Netherlands

The impact of news on energy prices (crude oil and natural gas) is investigated. The news sentiment is obtained from Thomson Reuters News Analytics which assesses whether a news item conveys positive, negative or neutral outlook on the commodity price. We propose a Local News Sentiment Level model to construct a running series of news sentiment. We employ event studies and Granger causality tests to assess the effect of news on the returns, price jumps and volatility. We find that the arrival of news in non-trading periods causes overnight returns, that news sentiment is caused by volatility and that news sentiment is more sensitive to negative than to positive price jumps. We find strong evidence that news sentiment causes jumps and conclude that market participants trade as some function of aggregated news. We apply several volatility models augmented with news sentiment and conduct an out-of-sample volatility forecasting study. The first class of models is the GARCH and the second class is the high-frequency volatility models (HEAVY). We adapt both models to account for asymmetric volatility, leverage and time to maturity effects. We find that including news sentiment in volatility models results in superior volatility forecasts.

CS93 Room B35 QUANTITATIVE METHODS FOR CREDIT RISK MANAGEMENT Chair: Raffaella Calabrese

C216: **SME performance through the credit crunch**
Presenter: **Jonathan Ansell**, The University of Edinburgh, United Kingdom
Co-authors: Paul Orton, Meng Ma, Galina Andreeva

The financial crisis that developed in 2007 has raised many issues for banks, regulators and government. Banks have been urged to be prudent in lending, but also urged to lend more to Small and Medium size Enterprises (SMEs) through governmental initiatives. The aim is to report on current research which explores a large part of UK SMEs' records during the 'credit crunch', over the period 2007 to 2010. The data consists of performance measures and other aspects of SMEs such as demographics and financial information. Logistic models have been used to explore the relationship between performance and the other attributes. As cross-sectional models, though, do not allow time series effect, random effect logit panel models are used to explore SMEs. The macroeconomic variables considered are indicators of directions of the economy, financial markets and economic general conditions. Newly established and matured SMEs are contrasted in the analysis to see if there has been a differential effect. Measures of model prediction accuracy are given for comparison across the models employed.

C334: **Estimating time-to-profit to assess creditworthiness of potential borrowers in revolving credit**
Presenter: **Galina Andreeva**, University of Edinburgh, United Kingdom
Co-authors: Luis Sanchez-Barrios, Jake Ansell

The aim is to present the first time-to-profit model designed to aid the decision as to whom grant credit in a revolving consumer credit product. Time-to-profit is the time in months that it takes a customer to break-even for the first time. This study builds on previous research conducted by the authors, in which return measures were used for the first time to assess creditworthiness of potential borrowers. As a concept, time-to-profit can be used not only to design application profit scorecards (i.e. decide whom to accept for credit), but also for investment planning activities of credit programmes. Discrete survival methods are used for both purposes in application to 35,530 customers of a Colombian lending institution. Sociodemographic, product and credit behaviour attributes are used as predictors. Results of time-to-profit scorecards are compared against more traditional models that estimate profit and return as continuous measures, and the resulting change in portfolio profit/return is investigated. In addition, the value of segmentation on the initial borrowing/contract duration is explored. Overall the research results have major implications for the growth of credit programmes that rely on profits generated by current customers to grant credit to new customers and hence to expand their scope.

C382: **Binary generalized extreme value additive modelling**
Presenter: **Giampiero Marra**, University College London, United Kingdom
Co-authors: Raffaella Calabrese, Silvia Angela Osmetti

Logistic regression is the commonly used model for bankruptcy prediction of small and medium enterprises, for instance. However, the assumptions of symmetric link function and linear or pre-specified covariate-response relationships may not be realistic, especially in scoring applications. To deal with these issues a binary generalized extreme value additive model is introduced. The approach uses the quantile function of the generalized extreme value distribution as link function as well as smooth functions of continuous predictors to flexibly model their effects. The framework is implemented in the bgeva R package which has a bgeva() function that works in a similar way to the glm() and gam()-like functions in R. The main ideas behind the methodology will be discussed and the bgeva package illustrated using Italian data on small and medium enterprises.

C665: **Improving credit card pricing by combing default and transactor/revolver regression models**
Presenter: **Mee Chi So**, University of Southampton, United Kingdom
Co-authors: Lyn Thomas, Hsin-Vonn Seow

The traditional approach in consumer lending is to develop a scorecard which ranks borrowers into Goods and Bads according to their risk of defaulting. To maximise the profitability of credit card customers, a second classification between Revolvers and Transactors becomes important. Transactors are credit card users who pay off all their balance every month. Thus they will not default but the bank does not make much profit from them. Revolvers are users who do not pay off their balance every month and so although they are possible defaulters, the bank makes extra income from them because of the interest charged. It is shown how Transactor/Revolver scorecards could be used in the accept/reject decision and to help decide on the optimal interest rate to charge. The resulting model is shown to be more accurate than the profitability model that ignores such estimates, on a dataset provided by a financial institution.

C167: **Stress testing retail loan portfolios using discrete survival models with macroeconomic variables**
Presenter: **Tony Bellotti**, Imperial College London, United Kingdom
Co-authors: Jonathan Crook

Stress tests of portfolios of retail loans are performed by specifying values of macroeconomic variables that reflect adverse economic conditions, such as the FSA anchor scenario, and quantify how those economic conditions affect risk on the portfolio. Therefore, there is a need for statistical models of retail credit risk that allow the inclusion of macroeconomic variables so that their effect as risk factors can be estimated. We discuss a method using discrete survival models and consider how to apply them for stress testing. A Monte Carlo simulation-based approach to stress testing is also discussed, as an alternative to scenario approaches to stress testing. The methods are illustrated using UK credit card portfolios.

CS09 Room B36 APPLICATIONS OF REALIZED VOLATILITY Chair: Massimiliano Caporin

C473: **Horizon effect in the term structure of long-run risk-return trade-offs**
Presenter: **Cedric Okou**, University of Quebec at Montreal, Canada
Co-authors: Eric Jacquier
The aim is to investigate the horizon effect in the long-run predictive relationship between market excess return and historical market variance. To this end, the asymptotic multivariate distribution of the term structure of risk-return trade-offs is derived, accounting for short- and long-memory in the market variance dynamics. A rescaled Wald statistic is used to test whether the term structure of risk-return trade-offs is flat, that is, the risk-return slope coefficients are equal across horizons. When the regression model includes an intercept, the premise of a flat term structure of risk-return relationships is rejected. In contrast, it fails to obtain statistical evidence against the equality of slope coefficients from constrained risk-return regressions estimated at different horizons. Then, a smoothed cross-horizon estimate is proposed for the trade-off intensity at the market level. The findings underscore the importance of economically motivated restrictions to improve the estimation of the Intertemporal Capital Asset Pricing Model.

C542: **From volatility to liquidity: Simple estimation from high and low prices**
Presenter: **Farshid Abdi**, University of St Gallen, Switzerland
Co-authors: Angelo Ranaldo
Using readily available data on daily high and lows prices, a simple estimation method of the efficient price volatility and bid-ask spread is developed. The model relies on general assumptions and it provides a closed-form solution for an unbiased estimator of efficient volatility. Moreover, it provides a better treatment of the volume effect caused by trading discontinuity and non-trading time. Using a comprehensive data set of high-frequency FX rates, it is shown that the liquidity estimator proposed is highly correlated with the actual bid-ask spread and other measures of market liquidity.

C531: **Forecasting volatility and covariance with jumps and co-jumps**
Presenter: **Aleksey Kolokolov**, Tor Vergata, Italy
Co-authors: Roberto Reno, Massimiliano Caporin
Understanding the dynamics of volatility and correlation is crucial for several financial applications. In particular, forecasting covariances between asset prices is primary important for asset allocation and risk management, as well as for derivatives pricing and risk hedging. In recent years, the literature on the estimation and forecasting of both volatility and correlation has benefited from the growing availability of high-frequence data. The aim is to contribute to the field by evaluating the impact of jumps and co-jumps of financial assets prices on volatilities and correlations among them. The contribution follows several directions. First, in multivariate stochastic volatility jump-diffusion settings, a new methodology of disentangling the covariation between the continuous parts of price processes from the co-jumps is developed. Moreover, a new high-frequency measure of co-jumps was developed to be used for forecasting. The second contribution deals with utilizing idiosyncratic jumps for forecasting realized volatilities and covariances. By means of a HAR-type model augmented with jumps empirical evidences on the relevance of jumps and co-jumps in volatility forecasting are provided. Finally, a fixed effect panel regression of covariations between assets is estimated, in order to study the effect of co-jumps on covariations.

C847: **The microstructure of the European sovereign bond**
Presenter: **Davide Tomio**, Copenhagen Business School, Italy
Co-authors: Marti Sunbrahmanyam, Loriana Pelizzon, Jun Uno
Market microstructure and liquidity in the Italian sovereign bond market, the largest in the Euro-zone, are studied using a unique new dataset, recently obtained from the Mercato dei Titoli di Stato (MTS), which provides tick-by-tick trade and quote data from individual broker-dealers. Our data cover the sovereign bonds of most European Union countries, for the period June 1, 2011 to November 15, 2012, which includes the Euro-zone crisis period. This database is unique for any market, in that it allows us to track individual orders and their revisions during the trading day. We document the strong non-linear relationship between changes in Italian sovereign risk and liquidity in the secondary bond market. We pinpoint which subset of bonds was the least affected by the worsening of the crisis, in terms of liquidity, and to what extent it was resilient to the deterioration of Italy's creditworthiness. We document that, under conditions of stress, a fraction of market makers withdraws from the market, and hence frequent quote revisions do not necessarily translate into higher liquidity. We also examine how liquidity improved after intervention by the European Central Bank (ECB), through its Long-Term Refinancing Operations (LTRO) and Outright Monetary Transactions (OMT) programs, starting in December 2011. Thus, we are able to assess the efficacy of the intervention by studying the changing interaction between the liquidity measures and credit default swap (CDS) spreads, and examine whether the intervention was successful in ameliorating credit risk and illiquidity.

C935: **Lassoing the HAR model: A model selection perspective on realized volatility dynamics**
Presenter: **Francesco Audrino**, University of St Gallen, Switzerland
Co-authors: Simon Knaus
Realized volatility computed from high-frequency data is an important measure for many applications in finance. However, its dynamics are not well understood to date. Recent notable advances that perform well include the heterogeneous autoregressive (HAR) model which is economically interpretable and but still easy to estimate. It also features good out-of-sample performance and has been extremely well received by the research community. We present a data driven approach based on the absolute shrinkage and selection operator (lasso) which should identify the aforementioned model. We prove that the lasso indeed recovers the HAR model asymptotically if it is the true model, and we present Monte Carlo evidence in finite sample. The HAR model is not recovered by the lasso on real data. This, together with an empirical out-of-sample analysis that shows equal performance of the HAR model and the lasso approach, leads to the conclusion that the HAR model may not be the true model but it captures a linear footprint of the volatility dynamics.

CS28 Room B30 COMPUTATIONAL DECISION THEORY Chair: Richard Hahn

C976: **False discovery rate regression**
Presenter: **James Scott**, University of Texas at Austin, United States
Many approaches for multiple testing begin with the assumption that all tests in a given study should be combined into a global false-discovery-rate analysis. But this may be inappropriate for many of today's large-scale screening problems, where auxiliary information about each test is often available, and where a combined analysis can lead to poorly calibrated error rates within different subsets of the experiment. To address this issue, we introduce an approach for false-discovery-rate regression (FDRR) that uses this auxiliary information to improve power while maintaining control over the global error rate. The method can be motivated by a hierarchical Bayesian model in which covariates are allowed to influence the local false discovery rate (or equivalently, the posterior probability that a given observation is a signal) via a logistic regression.

C761: **Numerical computation of minimax-regret treatment rules**
Presenter: **Matthew Masten**, Duke University, United States
Statistical treatment rules map data into treatment choices. Optimal treatment rules maximize social welfare. Although some finite sample results exist, it is generally difficult to prove that a particular treatment rule is optimal. Numerical results on minimax-regret treatment rules when there are many treatments are developed. The finite sample performance of several treatment rules is compared. The empirical success rule is found to perform poorly in unbalanced designs, and that when prior information about treatments is symmetric, balanced designs are preferred to unbalanced designs. Finally, the computation of optimal finite sample rules by applying methods from computational game theory is discussed.

C774: **Importance of finite sample inference for statistical mechanisms**
Presenter: **Aleksey Tetenov**, Collegio Carlo Alberto, Italy
In a game between self-interested proponents of innovations (e.g., new pharmaceuticals, new government programs) and a regulator, classical one-sided hypothesis tests judging the evidence of the proposal's effectiveness could be seen as a minimax strategy for the regulator. If the regulator sets the statistical test size equal to the ratio of the proponents' cost of collecting data to their benefit from approval of a null innovation, then proponents of null innovations will be deterred from attempting costly trials to gain approval by chance. To implement this strategy effectively, the regulator needs to know exact finite sample inference properties of the procedure. If the regulator uses asymptotic tests that do not control size in finite samples and the proponent could ascertain that (by simulation, for example), then the procedure will not have the desired deterrent effect.

C768: **Modular priors for partially identified models**
Presenter: **Ioanna Manolopoulou**, University College london, United Kingdom
Co-authors: Richard Hahn
The motivation is the challenges of drawing inferences from presence-only data. For example, when trying to determine what habitat sea-turtles "prefer" we only have data on where turtles were observed, not data about where the turtles actually are. Therefore, if we find that our sample contains very few turtles living in regions with tall sea grass, we cannot conclude that these areas are unpopular with the turtles, merely that we are unlikely to observe them there. Similar issues arise in forensic accounting: attempts to determine which companies are apt to misreport their official earnings based on a history of which firms were censured by the SEC are confounded by the fact that we only observe which firms got caught cheating, not which firms cheat (many of whom do not get caught). This sort of confounding is insurmountable from a point-estimation perspective, but the data are not entirely uninformative either. Parameterizing observation models is considered in a way that isolates which aspects of the model are informed by the data and which aspects are not. This approach allows us to construct priors which are informative with respect to the unidentified parts of the model without simultaneously (and unintentionally) biasing posterior estimates of the identified parameters; these priors do not "fight against" the data. In addition, their modularity allows for convenient sensitivity analysis in order to examine the extent to which our ultimate conclusions are driven by prior assumptions as opposed to our data.

C757: **DSS: Decoupled shrinkage and selection in linear models**
Presenter: **Richard Hahn**, University of Chicago, United States
Co-authors: Carlos Carvalho
A new variable selection approach is proposed from a fully Bayesian decision theory viewpoint. The method draws an explicit distinction between actions and inferences, effectively dealing with the trade-off associated with the competing goals of predictive generalization and interpretability. By decoupling posterior learning from model reporting, our approach creates a flexible framework where continuous shrinkage priors can be used but "sparse solutions" can be obtained. The method generalizes straightforwardly to the GLM setting.

CS56 Room B20 BANKING AND FINANCIAL MARKETS Chair: Arvid Raknerud

C199: **The pass-through of ban funding costs to loan and deposit rates: UK evidence**
Presenter: **Garry Young**, Bank of England, United Kingdom
Co-authors: Rashmi Harimohan, Michael McLeay
The period since 2008 has seen substantial heterogeneity in UK banks' cost of raising funds from wholesale markets, with implications for the transmission of policy rates to retail rates. Error correction models are estimated using a new panel dataset to determine the speed and extent of pass-through of changes in funding costs. Different specifications are tested to allow for varying degrees of heterogeneity in pass-through between banks. The results suggest that there is faster, and more complete, pass-through from changes in whole-market funding costs than changes in idiosyncratic funding costs. There is also differential pass-through between banks and across products.

C240: **On the influnce of anticipated changes to short-term market rates on banks' retail market rates**
Presenter: **Paul Mizen**, University of Nottingham, United Kingdom
Co-authors: Anindya Banerjee, Victor Bystrov
It is argued that banks anticipate short-term market rates when setting interest rates on loans and deposits. In order to include anticipated rates in an empirical model, we use two methods to forecast market rates - a level, slope, curvature model and a principal components model - before including them in a model of retail rate adjustment for four retail rates in four major euro area economies. Using both aggregate data and data from individual French banks, we find a significant role for forecasts of market rates in determining retail rates; alternative specifications with futures information yield comparable results.

C466: **Stressing bank profitability for interest rate risk**
Presenter: **Rochelle Edge**, Federal Reserve Board, United States
Co-authors: Valentin Bolotnyy, Luca Guerrieri
Bank stress testing is one of the key reforms to U.S. capital regulation in response to the shortcomings of the pre-crisis regime. The form of stress testing mandated by the Dodd-Frank Act for the Federal Reserve is undertaken around a small number of scenarios. This involves specifying macroeconomic and financial scenarios that represent stressful conditions over the horizon of the stress test and projecting, for each BHC in the stress test, its losses, income, and capital given that scenario. The focus is income. The aim is to consider the ability of empirical models to translate developments in interest rates into bank net interest margins (NIMs) by evaluating the pseudo-out-of-sample forecasting performance of NIMs using a range of forecast methods. The key findings are that (i) a forecast combination method based on different maturities of Treasury yields is superior to other methods and is the only method that outperforms a no change forecast; (ii) RMSEs even with the best model are nonetheless large even relative to the in-sample variation of NIMs; and (iii) conditional forecasts using models based on public data have difficulty distinguishing the implications of different interest rate configurations such as those included in recent Federal Reserve stress tests.

C651: **Cost of funds, credit risk and bank loan interest rates in the crisis**
Presenter: **Patrick Sevestre**, Banque de France and Universite de Paris I, France
Co-authors: Sanvi Avouyi-Dovi, Guillaume Horny
The aim is twofold. First, a thorough description of the evolution of debtor interest rates is provided for several categories of bank loans to non-

financial corporations before and during the recent crisis. Second, the changes in banks' valuation of risk induced by the crisis are assessed. It is shown in particular that the strong decline in the ECB refinancing rate that followed the collapse of Lehman Brothers allowed banks to increase the spread between interest rates they charged to firms and the ECB rate. Then, it is shown that this increase in the spread did not affect all types of loans, nor all firms, equally. In particular, low risk, large and well-established firms almost fully benefited from this decrease while smaller and younger firms did not and may even have seen the cost of their loans slightly increase. This discrepancy seems to have remained at least until the fourth quarter of 2010, the last period of observation currently available.

C660: **On what determines default rates on bank loans**
Presenter: **Arvid Raknerud**, Statistics Norway, Norway
Co-authors: Bjorn Helge Vatne
To assess banks' degree of losses or defaults on loans to households and businesses -henceforth termed default rates - under different macro economic conditions is an important part of stress testing. This is examined more systematically by means of an econometric model. First, the aim is to examine the relation between default rates and loan rates within a behavioral model with monopolistic competition, where banks are price setters in the loan markets (Cournot competitors) and the probability of default is endogenously determined for each bank as a function of the loan rate. Second, it is quantified the relation between, on the one side, default rates at the industry level and, on the other side, key variables at the macro and industry level. Third, findings from the theory model are utilized to address empirically how and why banks differ with respect to default rates.

CS87 Room B18 RECENT DEVELOPMENT IN SEMIPARAMETRIC AND NONPARAMETRIC ECONOMETRICS Chair: Patrick Saart

C078: **Regularization with instrumental variables for spatial lag panel with spatial fixed effects**
Presenter: **Clifford Lam**, London School of Economics and Political Science, United Kingdom
Co-authors: Pedro CL Souza
Regularized estimation of the spatial weight matrices in a large spatial lag model with spatial fixed effects is considered. We introduce the LASSO and adaptive LASSO with instrumental variables to perform sparse estimation of these matrices. This theoretically removes some restrictive technical assumptions compared to just using LASSO alone, in particular the decaying variance assumption for the noise. In practice this leads to improved performance of the resulting estimators compared with LASSO. We name the method IVLASSO. Theoretical properties are thoroughly studied when the size of the panel N can be larger than the sample size T , including the sign consistency for the adaptive LASSO estimators with instrumental variables presented and proved. Rate of estimation error is spelt out for the estimation of componentwise spatial fixed effect. In the special case where the covariates are exogenous, they themselves can be used as the instruments with LASSO/adaptive LASSO, so that no external instrumental variables are needed. Some simulation results will be presented along with a real data analysis.

C084: **Fixed-b estimation of fractional cointegration**
Presenter: **Fabrizio Iacone**, University of York, United Kingdom
Co-authors: Javier Hualde
In a model of fractional cointegration, we consider a time-domain, first stage, semiparametric estimator of the cointegrating parameter based on weighted co-variances. This estimator, which we refer to as fixed-b, is very easy to compute, only requiring widely available routines for the estimation of long run variances and covariances, and approximates a narrow band estimator with fixed bandwidth. We derive its limiting distribution also finding that it always achieves the fastest convergence rate in the class of first stage, semiparametric estimators. Finally, we compare the fixed-b with other first stage alternatives in a Monte Carlo experiment, finding that it always performs well in terms of asymptotic bias and variance. Recommendations for choices of kernel and bandwidth are also given.

C548: **Semi-parametric analysis of shape-invariant Engel curves with control function approach**
Presenter: **Nam Hyun Kim**, University of Canterbury, New Zealand
An extended generalized partially linear single-index (EGPLSI) model provides flexibility of a partially linear model and a single-index model. Furthermore, it also allows for the analysis of the shape-invariant specification. Nonetheless, the model's practicality in the empirical studies has been hampered by lack of appropriate estimation procedure and method to deal with endogeneity. The aim is to establish an alternative control function approach to address the endogeneity issue in the estimation of the EGPLSI model. It is also shown that all attractive features of the EGPLSI model discussed in the literature are still available under the proposed estimation procedure. Economic literature suggests that semiparametric technique is an important tool for an empirical analysis of Engel curves, which often involves endogeneity of the total expenditure. It is shown that the newly developed method is applicable and able to address the endogeneity issue involved in semiparametric analysis of the empirical Engel curves.

C797: **Construction of structural functions in conditional independence models**
Presenter: **Sami Stouli**, University College London, United Kingdom
Nonparametric identification of structural functions is considered in models imposing an independence restriction between a continuous endogenous variable and a scalar disturbance conditional on a control variable. The proposed construction complements existing approaches by formulating a set of alternative conditions, sufficient to achieve full knowledge of structural features of the model. Structural effects of a policy or treatment are allowed to vary across subpopulations that can be located on the joint distribution of unobservables of the model. Estimation of a semiparametric nonseparable triangular model with an endogenous regressor illustrates the methodology. An empirical application to the estimation of gasoline demand functions in the United States is given. Finite-sample simulations show the estimator performs well in practice.

C549: **Nonparametric hypothesis testings with generated covariates**
Presenter: **Patrick Saart**, University of Canterbury, New Zealand
A wide range of economic and statistical applications requires estimation of an unknown function when one or more of the covariates are not directly observed, but themselves must be estimated (parametrically, semiparametrically, nonparametrically or (computationally) algorithmically) in a preliminary step. There are already some recent studies in the literature which examine the influence of generated covariates on the asymptotic properties of some nonparametric estimators of such functions in the independent data. Nonetheless, the influence of the generated covariates on other statistical inferences, particularly hypothesis testing, has not received much attention to date. The aim is to propose an optimal test procedure for a parametric model of an unknown function against a nonparametric alternative when one of the variables is generated. A set of general asymptotic results is derived, which are then used to establish the asymptotic behavior of test for testing the marginal density and the regression function. The usefulness of the results is also illustrated by applying them to the hypothesis testing in various models which are well-known in the literature.

CS59 Room B29 TIME SERIES AND OPTION DATA Chair: Jeroen V.K. Rombouts

C387: **Sparse change-point models**
Presenter: **Arnaud Dufays**, CREST, France
Co-authors: Jeroen Rombouts
Change-Point (CP) specifications constitute flexible models that capture structural changes by allowing for abrupt switches in the model parameters. Nevertheless, up to day, CP models generally produce poorer forecasts than the no structural break counterpart. This feature doubtfully arises due to the large parameter set of the CP model coming from the assumption that all the model parameters evolve when it faces a structural break. Bearing on the Normal-Gamma prior, the assumption is relaxed by shrinking irrelevant parameters toward zero. Empirical exercises demonstrate that it allows us to detect which parameters vary from one regime to another. Some well known CP applications are revisited and it is found that many breaks are only due to a change in one parameter in the mean or in the variance.

C554: **The price of risk**
Presenter: **Francesco Violante**, CREATES Aarhus University, Denmark
Co-authors: Jeroen Rombouts, Lars Stentoft
The variation over time of the magnitude of price movements of financial assets (variance risk) represents a source of uncertainty that agents are subject to. Consistently, risk adverse agents should require a compensation for bearing the randomness of future variance, i.e., a variance risk premium. Despite the number of empirical studies, there is no clear consensus about sign and magnitude of the premium and its linkages with the economy. These results are criticized and a class of flexible structural time series models is proposed, which, using signal extraction techniques, estimate more precisely the market price of risk. The inclusion of interactions is advocated, non-linearities and discontinuities - essential to replicate complex dynamics - and linkages with macro-finance and business-cycle variables, thus providing a new tool to investigate changes in agents' behavior over time, e.g., the transition between periods of economic expansion and recession.

C1213: **Predicting crash risk using options trading information**
Presenter: **Ruslan Tuneshev**, University of Durham, United Kingdom
A novel stock market tail loss measure is constructed using intraday trading volume information from S&P500 index option quotes and trades from 2005 to 2011. We employ a trade classification algorithm to incorporate the presence of buying pressure in the construction of the tail loss measure aiming to improving the predictability of extreme negative returns. Our crash risk evaluation is performed along three dimensions: (i) the probability of a negative price movement, (ii) the expected magnitude of a downward price jump, and (iii) the premium that investors claim for taking the risk. The results show that the tail loss measure is particularly important in explaining the magnitude of realized stock market declines in short- and medium-term horizons. Moreover, it outperforms other option-based measures in predicting the extreme downward stock market movements in the short-run. Overall, our findings support the hypotheses that options markets lead stock markets and that trading volume direction provides crucial information regarding future price fluctuations.

C990: **Separating the integrated volatility into CIR latent factors**
Presenter: **Rachidi Kotchoni**, Universite de Cergy-Pontoise, France
The integrated volatility is modeled as a mixture of CIR latent factors. The latent factors are identified by imposing an ordering on the mean reversion rates of the underlying spot processes. A state-space representation implied by the CIR mixture model is used to estimate the parameters of interest and to filter the integrated volatility from realized measures. This decomposition turns out to have important implications for the prediction of the risk premium. The relevance of this approach is confirmed by an empirical application based on the Alcoa index listed in the Dow Jones Industrials.

C563: **Asymmetric mixtures models, jumps and option pricing**
Presenter: **Jeroen Rombouts**, HEC Montreal, Canada
Asymmetric heteroskedastic normal mixture models are used to fit return data and to price options. The models can be estimated straightforwardly by maximum likelihood, have high statistical fit when used on S&P 500 index return data, and allow for substantial negative skewness and time varying higher order moments of the risk neutral distribution. When forecasting out-of-sample a large set of index options between 2006 and 2011, substantial improvements are found compared to several benchmark models in terms of dollar losses and the ability to explain the smirk in implied volatilities.

CS70 Room G15 TOPICS IN FINANCIAL ECONOMETRICS Chair: Leopold Soegner

C253: **Variance reduction for infinite dimensional problems**
Presenter: **Klaus Poetzelberger**, WU Vienna, Austria
A variance reduction method for the Monte Carlo estimation of certain infinite-dimensional problems is presented. Applications in financial mathematics are the estimation of marginal distributions, conditional expectations of diffusion processes or boundary crossing probabilities. To estimate such a functional of a continuous-time stochastic process, n paths of discrete-time approximations are generated and corresponding averages are used as estimates. This Monte Carlo method is biased. Typically, the generation of each time-discrete path involves the generation of m univariate random variables and the bias is $O(1/m^p)$. The mean-squared error is then of order $O(1/m^{2p}+1/n) = O(1/N^q)$, for constants p and q and $N = mn$, the number of used univariate random variables. Naive applications of MC often have a MSE of order only $1/\sqrt{N}$. The method of control variables allows for an efficient reduction of the MSE. In infinite-dimensional and biased problems, the approximating functional is itself approximated by a functional of a discrete-time path of smaller complexity. Although the expectation of the control variable has to be estimated, the combination of expectation and approximation allows an improvement of the convergence rate. Iterating the control variables leads even to a MSE which is $O(1/N)$, the approximation rate of finite-dimensional problems.

C270: **Analysis of multivariate financial time series via Bayesian factor stochastic volatility models**
Presenter: **Gregor Kastner**, WU Vienna University of Economics and Business, Austria
Co-authors: Sylvia Fruehwirth-Schnatter, Hedibert Freitas Lopes
In recent years, multivariate factor stochastic volatility (SV) models have been increasingly used to analyze high-dimensional financial and economic time series because they can pick up the joint volatility dynamics by a small number of latent time-varying factors. The main advantage of such a model is its parsimony; all variances and covariances of a time series vector are governed by a low-dimensional common factor with the components following independent SV models. For problems of this kind, MCMC is a very efficient estimation method; nevertheless, it is associated with a considerable computational burden when the number of assets is moderate to large. To overcome this, we sample the latent states 'all without a loop' (AWOL), consider various reparameterizations such as (partial) non-centering, and apply an ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation at various levels. Moreover, we use modern supercomputing architectures for parallel

implementation. Our algorithm is designed in a way such that existing software crafted for efficient Bayesian univariate SV estimation can easily be incorporated. Finally, to show the effectiveness of our approach, we apply the model to a vector of daily returns.

C287: **Parameter estimation and inference with spatial lags and cointegration**
Presenter: **Jan Mutl**, European Business School, Germany
Co-authors: Leopold Soegner

Dynamic panel data models are studied where the long run outcome for a particular cross-section is affected by a weighted average of the outcomes in the other cross-sections. We show that imposing such a structure implies a model with several cointegrating relationships that, unlike in the standard case, are nonlinear in the coefficients to be estimated. Assuming that the weights are exogenously given, we extend the dynamic ordinary least squares methodology and provide a dynamic two-stage least squares estimator. We derive the large sample properties of our proposed estimator under a set of low-level assumptions and investigate its small sample distribution in a simulation study. Then our methodology is applied to US financial market data, which consist of credit default swap spreads, firm specific and industry data. We construct the economic space using a closeness measure for firms based on input-output matrices. This measure reflects the spreading or contagion of credit risk through the supply chain. Our estimates show that this particular form of spatial correlation of credit default spreads is substantial and highly significant.

C406: **Generalized density forecast combinations with censored likelihood scoring rules**
Presenter: **Justinas Pelenis**, Institute for Advanced Studies Vienna, Austria

Generalized density forecast combination schemes are considered to improve the tail and value-at-risk forecast accuracy. The quantile and tail behavior is of interest from the risk management perspective and developing specific density forecast combination schemes would be useful to address this research question. To focus on the particular density forecasts over a specific region of interest the forecast combinations are based on alternative censored likelihood scoring rules instead of the commonly used log predictive scoring rule. The performance of the new density forecast combination schemes is evaluated in a number of simulation and empirical applications.

C354: **Generalized method of moment based parameter estimation of affine term structure models**
Presenter: **Leopold Soegner**, Institute for Advanced Studies, Austria
Co-authors: Jaroslava Hlouskova

Parameter estimation of affine term structure models is investigated by means of the generalized method of moments. Results obtained for m-polynomial processes in mathematical finance literature to derive the moments of the affine latent process driving the term structure are used, and after specifying the properties of the micro-structure noise, the moments of the yields observed. The moments have been derived in such a way to investigate parameter estimation for all affine models with a dimension of the latent process smaller or equal to three. Equipped with these moments parameter estimations by means of the generalized method of moments can be performed. To implement GMM estimation the number of moment restrictions has to be chosen carefully to obtain reliable parameter estimates. In addition, the Wald and the distance difference test to test for redundant moments and to test for different market price of risk specifications are implemented. After a simulation study, the estimation procedure is applied to empirical interest rate data.

CS94 Room G16 UNCERTAINTY AND REAL-TIME TURNING POINTS DETECTION II Chair: Gianluigi Mazzi

C154: **Monthly US business cycle indicators: A new multivariate approach based on a band-pass filter**
Presenter: **Martyna Marczak**, University of Hohenheim, Germany
Co-authors: Victor Gomez

A new multivariate method to construct business cycle indicators is proposed. The method is based on a decomposition into trend-cycle and irregular. To derive the cycle, a multivariate band-pass filter is applied to the estimated trend-cycle. The whole procedure is fully model-based. Using a set of monthly and quarterly US time series, two monthly business cycle indicators are obtained for the US. They are represented by the smoothed cycles of real GDP and the industrial production index. Both indicators are able to reproduce previous recessions very well. Series contributing to the construction of both indicators are allowed to be leading, lagging or coincident relative to the business cycle. Their behavior is assessed by means of the phase angle and the mean phase angle after cycle estimation. The proposed multivariate method can serve as an attractive tool for policy making, in particular due to its good forecasting performance and quite simple setting. The model ensures reliable realtime forecasts even though it does not involve elaborate mechanisms that account for, e.g., changes in volatility.

C188: **Forecasting US recessions: The role of sentiments**
Presenter: **Jonas Nygaard Eriksen**, Aarhus University, Denmark
Co-authors: Charlotte Christiansen, Stig Vinther Moller

Sentiment variables are examined as new predictors for US recessions. We combine sentiment variables with either classical recession predictors or with common factors based on a large panel of macroeconomic and financial variables. Sentiment variables hold vast predictive power for US recessions in excess of both the classical recession predictors and the common factors. The strong importance of the sentiment variables is documented both in-sample and out-of-sample.

C285: **The trilemma between accuracy, timeliness and smoothness in real-time signal extraction**
Presenter: **Marc Wildi**, University of Applied Sciences Zurich, Switzerland
Co-authors: Tucker McElroy

The evaluation of economic data and monitoring of the economy is often concerned with an assessment of mid- and long-term dynamics of time series (trend and/or cycle). Frequently, one is interested in the most recent estimate of a target signal, a so-called real-time estimate. Unfortunately, real-time signal extraction is a difficult prospective estimation problem which involves linear combinations of one- and possibly infinitely many multi-step ahead forecasts of a series. We here address performances of real-time designs by proposing a generic Direct Filter Approach (DFA). We decompose the ordinary MSE into Accuracy, Timeliness and Smoothness error components, and we propose a new two-dimensional tradeoff between these conflicting terms, the so-called ATS-trilemma. With this formalism, we are able to derive a general class of optimization criteria that allow the user to address specific research priorities, in terms of the Accuracy, Timeliness and Smoothness properties of the resulting real-time filter.

C613: **Cycles, syllogisms and semantics: examining the idea of spurious cycles**
Presenter: **Stephen Pollock**, University of Leicester, United Kingdom

The claim that linear filters are liable to induce spurious fluctuations has been repeated many times of late. However, there are good reasons for asserting that this cannot be the case for the filters that, nowadays, are commonly employed by econometricians. If these filters cannot have the effects that have been attributed to them, then one must ask what effects the filters do have that could have led to the aspersions that have been made against them.

C614: **The KOF Barometer, version 2013: A composite leading indicator for the Swiss business cycle**

Presenter: **Jan-Egbert Sturm**, ETH Zurich, Switzerland

Co-authors: Klaus Abberger, Michael Graff, Boriss Siliverstovs

The aim is to present a composite leading indicator for the Swiss business cycle as measured using a growth rate cycle concept. It is the result of a complete overhaul of the KOF Barometer as published on a monthly basis by the KOF Swiss Economic Institute since 1976. By focusing on providing leading information on the business cycle it aims to signal turning points (peaks and troughs) between economic expansions and downturns at an early stage and in real time.

Sunday 15.12.2013 08:45 - 10:25 Parallel Session G – ERCIM

ES08 Room Senate OPTIMAL DESIGN FOR MEDICAL APPLICATIONS Chair: Stefanie Biedermann

E215: **Optimal designs for trials with survival endpoints in discrete time**
Presenter: **Mirjam Moerbeek**, Utrecht University, Netherlands
In a trial with survival endpoints the occurrence and timing of some event, such as death or smoking initiation, is compared across different treatment conditions. Although the survival process is often continuous, the occurrence and timing of the event may be measured discretely using time intervals. In the planning phase of such a trial it is important to carefully select the optimal design. Designs for such trials are characterized by the number of subjects and the number of time intervals. As costs are associated with including subjects in the trial and taking measurements, a cost function is used to select the optimal design. The variance of the treatment effect estimator is used as optimality criterion since smallest variance leads to largest power of the test on treatment effect. The optimal design depends on the underlying survival function, the treatment effect size and costs at the subject and measurement level. A general recommendation with respect to the optimal design can therefore not be given. The PODSE software facilitates in finding the optimal design for any underlying survival function. A study on recovery from anorexia nervosa is used to illustrate the optimal design methodology.

E616: **Optimal designs for survival data**
Presenter: **Maria Konstantinou**, University of Southampton, United Kingdom
Co-authors: Stefanie Biedermann, Alan Kimber
Survival data are observed in many industrial or biomedical 'time to event' experiments when the exact measurement of all the subjects is not observed. Finding efficient designs for such experiments can be problematic since the statistical models involved will usually be nonlinear, making the optimal choice of design parameter dependent. Analytical characterisations of locally D- and c-optimal designs for a wide class of models are provided, thus reducing the numerical effort for design search substantially. Designs are also constructed which are robust to parameter misspecifications when a range of parameter values is provided. Finally, the issue of model robustness is addressed, since in practice the proposed parametric model will only hold approximately.

E468: **Designing experiments for accelerated failure time models**
Presenter: **Jesus Lopez-Fidalgo**, University of Castilla-La Mancha, Spain
Co-authors: Maria Jesus Rivas-Lopez, Rodrigo del Campo
Proportional Hazards (PH) models have been widely used to analyze survival data. The restrictions for this model are satisfied just for a few probability distributions, so in many cases survival data do not verify the assumption of proportional hazards. An alternative to the PH model with more relaxed conditions could be Accelerated Failure Time (AFT) models. AFT models are fairly common used in the field of manufacturing, but they could be useful for analyzing clinical trial data. AFT models focus on the direct effect of the explanatory variables on the survival time allowing an easier interpretation of the effect of the correspondent covariate on the survival time.

E677: **Optimal group sequential tests for survival data**
Presenter: **Lisa Hampson**, Lancaster University, United Kingdom
Co-authors: Christopher Jennison
The benefits of monitoring clinical trials as they progress are typically measured in terms of a reduction in expected sample size. However, when conducting survival trials, recruitment will often be near to completion before an adequate number of events has been accumulated to perform the first interim analysis. We seek group sequential tests (GSTs) for survival data achieving a rapid time to a conclusion rather than a low sample size. We derive optimal GSTs assuming survival times are exponentially distributed, optimizing boundaries for the scenario when observed information sequences are equal to their expected values, noting that this will be accurate asymptotically. We compare the optimal tests with existing designs that attain low sample sizes for monitoring "standard" data types following parametric regression models. It is not obvious that these designs should be efficient for survival data given that the relationship between time and expected information is non-linear. However, we find that tests achieving low sample sizes for monitoring normally distributed data are also efficient for survival data. In particular, flexible error spending designs are highly efficient, often performing within 1% of optimal tests. We explain this trend and verify the small sample efficiency of the optimal designs via simulation.

ES99 Room Gordon MULTIVARIATE SURVIVAL ANALYSIS Chair: Jacobo de Una

E333: **Estimation from cross-sectional samples under bias and dependence**
Presenter: **Micha Mandel**, The Hebrew University of Jerusalem, Israel
Co-authors: Yosef Rinott
A population that can be joined at a known sequence of discrete times is sampled crosssectionally, and the sojourn times of individuals in the sample are observed. It is well known that cross-sectioning leads to length-bias, but less well known and often ignored that it may result also in dependence among the observations. We study conditions under which observed sojourn times are independent and conditions under which treating observations as independent using the product of marginals in spite of dependence results in proper or improper and wrong inference. We provide conditions for consistency, and further asymptotic properties, including normal and non-normal distributional limits of estimators. The theoretical study is supported by simulations.

E538: **Statistical inference based on the nonparametric maximum likelihood estimator under double-truncation**
Presenter: **Takeshi Emura**, National Central University, Taiwan
Co-authors: Konno Yoshihiko
Doubly truncated data consist of samples whose observed values fall between the right- and left- truncation limits. With doubly truncated samples, the distribution function of interest is estimated using the nonparametric maximum likelihood estimator (NPMLE), which is obtained through a self-consistency algorithm. Since the asymptotic distribution of the NPMLE is complicated, the bootstrap method has been suggested for statistical inference. A closed-form estimator for the asymptotic covariance of the NPMLE is proposed, which is shown to be a computationally attractive alternative to bootstrapping. Furthermore, developing various statistical inferences, such as confidence interval, goodness-of-fit test and confidence bands, the usefulness of the proposed method is demonstrated. Simulations are carried out to compare the proposed method with both the bootstrap and jackknife methods. The methods are illustrated using the childhood cancer dataset.

E594: **Computationally simple estimation and improved efficiency for special cases of double truncation**
Presenter: **Rebecca Betensky**, Harvard School of Public Health, United States
Co-authors: David Simon, Matthew Austin
Doubly truncated survival data arise when event times are observed only if they occur within subject specific intervals of times. Existing iterative

estimation procedures for doubly truncated data are computationally intensive. These procedures assume that the event time is independent of the truncation times, in the sample space that conforms to their requisite ordering. This type of independence is referred to as quasi-independence. Two special cases of quasi-independence are identified and considered: complete quasi-independence and complete truncation dependence. For the case of complete quasi-independence, the nonparametric maximum likelihood estimator is derived in closed-form. For the case of complete truncation dependence, a closed-form nonparametric estimator is derived that requires some external information, and a semi-parametric maximum likelihood estimator that achieves improved efficiency relative to the standard nonparametric maximum likelihood estimator, in the absence of external information. The consistency and potentially improved efficiency of the estimators are demonstrated in simulation studies, and their use in application is illustrated to studies of AIDS incubation and Parkinson's disease age of onset.

E692: **Nonparametric estimation of a distribution function from doubly truncated data under dependence**
Presenter: **Carla Moreira**, University of Vigo, Spain
Co-authors: Jacobo de Una-Alvarez, Roel Braekers

Nonparametric estimation of a distribution function which is observed under random double truncation is considered. We introduce a new estimator for situations in which the lifetime and the truncation times may be dependent. The dependence structure between the lifetime and the truncation times is given by a copula function which is known but for a finite number of parameters. Two different iterative algorithms to compute the estimator in practice are introduced, and their performance is explored through an intensive Monte Carlo simulation study. The asymptotic properties of the proposed estimator are discussed. Several applications to medical data are included for illustration proposes.

ES22 Room Athlone STATISTICS IN BIOMEDICINE Chair: M. Brigida Ferraro

E024: **Mixture of extended Plackett-Luce ranking models for epitope mapping in bioassay experiments**
Presenter: **Cristina Mollica**, Sapienza University of Rome, Italy
Co-authors: Luca Tardella

The use of probability models for ranked data is proposed as a useful alternative to a quantitative data analysis to investigate the outcome of bioassay experiments, when the preliminary choice of an appropriate normalization method for the raw numerical responses is difficult or subject to criticism. In the context of multistage ranking models we present an original generalization of the popular *Plackett-Luce model*, accounting for the order of the ranking elicitation process. The usefulness of the novel model is illustrated with its maximum likelihood estimation for data from the Large Fragment Phage Display experiment, aimed at the epitope mapping of a specific human oncoprotein. Necessary steps for a successful likelihood maximization in a general mixture modeling framework are obtained through a hybrid version of the Expectation-Maximization algorithm. The performance of the mixture model using the new distribution as mixture components is compared with those relative to alternative ranking models. A discussion on the interpretation of the identified clusters and a comparison with more standard quantitative approaches are finally provided.

E139: **Hidden Markov of factor analyzers for biclustering of microarray time course data in multiple conditions**
Presenter: **Francesca Martella**, La Sapienza University of Rome, Italy
Co-authors: Antonello Maruotti

A challenging task in time course microarray data is to discover groups of genes that show homogeneous temporal expression patterns when time course experiments are collected in multiple biological conditions. In such a case, an appealing goal would be related to discover local structures composed by sets of genes that show homogeneous expression patterns across subsets of biological conditions which also capture the history of the gene's and condition's dynamic behavior across time. To address this, at each time point one could apply any biclustering method for identifying differentially expressed genes across biological conditions. However, a consideration of each time point in isolation can be inefficient, because it does not use the information contained in the dependence structure of the time course data. Our proposal is an extension of the Hidden Markov of factor analyzers model allowing for simultaneous clustering of genes and biological conditions. The proposed model is rather flexible since it allows genes to move between the (hidden) components during the period of observation to appropriately account for heterogeneity across the gene sequences. To cluster biological conditions we suggest a factorial representation of component-specific means by using a binary row stochastic matrix representing condition membership.

E567: **Integrative approaches for modeling of omics and genetic data in epidemiological studies**
Presenter: **Jeanine Houwing-Duistermaat**, Leiden University Medical Center, Netherlands

Genome wide association studies in large case-control series have revealed many loci for complex genetic diseases such as osteoarthritis and thrombosis. These genetic variants explain only a part of the genetic variation. To unravel the genetic basis of these diseases further, epidemiologists currently consider the use of intermediate phenotypes such as metabolomics, gene expression, and methylation. To improve the power for detecting genetic variants and to obtain more insight in the biological mechanisms of the disease, the intermediate phenotypes and the genotypes are jointly modeled. A two-stage model is used, where the outcome is modeled as a function of the intermediate phenotype and the intermediate phenotype is modeled as a function of the genetic variants. To increase the information even further multiple case families are used. These families are genetically more homogeneous than single cases and are enriched for "rare" genetic variants. Moreover such a selection scheme enables application of the joint likelihood of the outcomes and genotypes which also increase the information on the model parameters. Note that for valid analysis of these families the selection scheme and the correlation between relatives have to be modeled. The methods are illustrated with osteoarthritis and thrombosis datasets.

E811: **Supervised classification for rare variant calling in next generation sequencing pooled experiments**
Presenter: **Mario Guarracino**, ICAR-CNR, Naples, Italy
Co-authors: Maria Brigida Ferraro

The study of rare variants in next generation sequencing (NGS) experiments enables the detection of causative mutations in the human genome. NGS is a relatively new approach for biomedical research, useful for the genetic diagnosis in extremely heterogeneous conditions. Nevertheless, only few publications address the problem when pooled experiments are considered and existing tools are often inaccurate. We describe how data generated by high-throughput NGS experiments are aligned and filtered. We describe de facto standard techniques and their organization in the pre-processing phase. We will show how to detect rare single nucleotide polymorphism by filtering and constructing features employed in the learning phase. Then, we focus on a supervised learning approach in order to obtain new knowledge about genomic variations in human diseases and we compare different computational procedures to identify and classify these variants.

ES25 Room Russell STATISTICAL METHODOLOGY FOR NETWORK DATA Chair: Jason Wyse

E186: **Estimation of posterior graph functions in networks using stochastic block models**
Presenter: **Pierre Latouche**, Sorbonne university, France
Co-authors: Stephane Robin
Networks have been widely used in many scientific fields, and in particular in social sciences, in order to represent interactions between objects of interest. Many random graph models have been proposed to extract knowledge from these structured data sets. For instance, the stochastic block model (SBM) allows the search of groups of vertices sharing homogeneous connection profiles. We consider the W-graph model which is known to generalize many random graph models, but for which very few methods exist to perform inference on real data. First, we recall that the SBM model can be represented as a W-graph with a block-constant graphon function. Using a variational Bayes expectation maximization algorithm, we then approximate the posterior distribution over the model parameters of SBM and we show how this variational approximation can be integrated in order to estimate the posterior distribution of W-graph graph function. In this Bayesian framework, we also derive the occurrence probability of a motif. In practice, this allows us to test if a motif is over-represented in a given network. All the results presented here are tested on simulated data and the French political blogosphere network.

E771: **Random and nonparametric effects in exponential random graph models**
Presenter: **Goeran Kauermann**, LMU Munich, Germany
Co-authors: Stephanie Thiemichen
Exponential Random Graph Models (ERGM) are a common tool in network data analysis. The models describe the distribution of a network graph with an exponential family, where the statistics are counts of edges, k stars or triangles, for instance. Though the models mirror the welcome properties of exponential families, their behavior is quite unstable and the parameters space resulting in suitable network graphs is peculiarly shaped. In practice, this results in either fully connected or fully unconnected networks and adds to complications with respect to numerical methods for fitting ERGMs. To compensate for this deficit exponentially weighted canonical statistics have been proposed in the literature. We present two other contributions to compensate degeneracy. Firstly, we include random nodal effects to compensate for heterogeneity in the nodes of a network. Secondly, we show how smoothing models based on pseudo likelihood methods can be used to derive nonlinear statistics which add to the numerical stability of ERGMs.

E1108: **Clustering bipartite networks using collapsed latent block models**
Presenter: **Jason Wyse**, University College Dublin, Ireland
Co-authors: Nial Friel, Pierre Latouche
The latent block model (LBM) and collapsed LBM are useful models for bi-clustering of categorical, count and continuous valued data matrices. We show how it is possible to model bipartite networks using LBMs. Typically one can think of a bipartite network as an affiliation network with members being affiliated to clubs. The task here is searching for clusterings or groupings of clubs and members simultaneously. Ties between members and clubs need not be restricted to the presence-absence domain as one can use the LBM set-up to model categorical, count and continuous valued affiliations of members to clubs. Use of MCMC for estimation for large networks can be impractical. We thus propose an algorithm based on the ideas of greedy search using the integrated complete likelihood (ICL) criterion. This strategy can be used to select the number of club and member clusters for the LBM and provides a flexible and pragmatic framework for working with large bipartite networks. The approach is tested in a number of scenarios including large and challenging bipartite networks. Scalability considerations related to the approach are discussed.

E1160: **Using semiparametric copulas for assessing overall goodness-of-fit to network data**
Presenter: **Johan Koskinen**, University of Manchester, United Kingdom
The complex dependencies of network data typically mean that there are few useful asymptotic results available for the purpose of model selection. As a consequence a number of heuristically based goodness-of-fit procedures now are common practice in a number of modeling frameworks. Researchers typically investigate the marginal fit of several functions or goodness of fit statistics of the simulated networks. This procedure does not give an overall measure of fit that takes into account the associations between statistics. Combining a collection of measures using a distance measure such as the Mahalanobis distance has been proposed. The distributions of the statistics do however tend to suffer from complicated dependencies, be highly skewed and discrete, and often have a very small range. This suggests that the combined measure is going to be sensitive to the choice of distance metric. We propose to use a type of multivariate empirical distribution function for assessing the fit of data that preserves the rank order and does not rely on any particular distance metric. This involves transforming the statistics using a semiparametric copula in a Markov chain Monte Carlo scheme that does not deal explicitly with the marginal distribution functions.

ES27 Room Jessel STATISTICAL PROCESS CONTROL Chair: Axel Gandy

E393: **Incorporating parameter uncertainty into the setup of EWMA control charts monitoring normal variance**
Presenter: **Sven Knoth**, Helmut Schmidt University Hamburg, Germany
EWMA charts usually monitor normal means in case of unknown parameters. It would be useful to evaluate the estimation effects also for variance control charts. Given a sequence of batches of size n, $\{X_{ij}\}$, $i = 1,2,\ldots$ and $j = 1,2,\ldots,n$, the batch sample variance is calculated and the EWMA smoothing with given λ is applied. The upper scheme stops when EWMA statistic is larger than the threshold c_u. The parameters $0 < \lambda \leq 1$ and $c_u > 0$ are chosen to enable a certain useful detection performance. The most popular performance measure is the so-called Average Run Length (ARL), that is $E_\sigma(L)$ for the true standard deviation σ. If σ_0 has to be estimated by sampling data during a pre-run phase, then this uncertain parameter effects, of course, the behavior of the applied control chart. Typically the ARL is increased. The uncertainty impact usually considers the changed ARL patterns and possible adjustments. A different way of designing the chart is treated: Setup the chart through specifying a certain false alarm probability such as $P_{\sigma_0}(L \leq 1000) \leq \alpha$. This results in a specific c_u. A feasible way to determine this value c_u is described also in case of unknown parameters for a pre-run series of given size (and structure). A two-sided version of the introduced EWMA scheme is analyzed as well.

E852: **Adjusting for estimation error in control charts via bootstrap methods**
Presenter: **Jan Terje Kvaloy**, University of Stavanger, Norway
Co-authors: Axel Gandy
To use control charts in practice, the in-control state usually has to be estimated. The estimation error might then have a detrimental effect on the performance of the control chart. We suggest a method which, based on bootstrapping the data used to estimate the in-control state, can account for the effect of this estimation error. The bootstrap method works for various types of control charts, and can be adapted to make different types of adjustments. We primarily propose to use the method to give a guaranteed conditional in-control performance, but other adjustments like for instance bias adjustments can also be made. With the guaranteed performance the monitoring scheme is tuned to guarantee with a high probability a certain in-control performance, conditional on the estimated in control distribution. Large sample properties of the adjustment have been derived. If time permits, model choice with regard to chart performance, when e.g. regression parameters have to be estimated, will also be briefly discussed.

E873: **A class of Dirichlet process priors for nonparametric multiple change point detection**
Presenter: **Gordon Ross**, University of Bristol, United Kingdom

Many statistical process control applications call for the detection of multiple change points in a sequence of data, which divides the sequence into a collection of segments. In a parametric context where the distribution of the sequence is known except for a finite number of unknown parameters, the number of change points is typically chosen using model selection criteria such as AIC or BIC. However, this does not generalise easily to nonparametric contexts where the sequence distribution is unknown, and choosing the number of change points in this setting remains a difficult task. A class of Dirichlet Process based priors is presented for performing nonparametric change point detection in a fully Bayesian setting which allows for a principled selection of the number of change points. These priors also allow for the pooling of information across segments in order to capture various forms of possible dependence. This includes the simple case where the (unknown) sequence distribution only undergoes shifts in mean and/or variance, up to more complex cases where the whole distribution can change. It is demonstrated that such pooling allows for more effective change point detection than standard approaches which treat each segment independently.

E1262: **Implementing Bayesian monitoring through Monte Carlo methods**
Presenter: **Axel Gandy**, Imperial College London, United Kingdom

When frequentist methods are being used for monitoring, the in control state is often not really known and needs to be estimated. In practice, this problem is often ignored. If one tries to take account of it then difficulties arises with eg calibrating the false alarm rate properly, in particular if multiple items need monitoring. Bayesian models allow a more satisfactory resolution to this issue, allowing averaging of estimation and prediction error. To implement these models in practice, one could work exclusively in a suitable conjugate setup making exact computations easy. This can, however, severely restrict the model choice. One of the reasons for the tremendous success of Bayesian methods is the flexibility afforded by simulation methods such as MCMC and particle filtering/sequential Monte Carlo. The purpose is to discuss how Monte Carlo methods, in particular a recently proposed system called Rolling Markov Chain Monte Carlo (RMCMC), can be used to implement monitoring through Bayesian models.

ES93 Room Torrington CLUSTER-WEIGHTED MODELLING Chair: Salvatore Ingrassia

E274: **Flexible model-based clustering via the cluster-weighted approach**
Presenter: **Antonio Punzo**, University of Catania, Italy

Cluster-Weighted Models (CWMs) are a flexible family of mixture models for fitting the joint density of a pair $(\mathbf{X}, Y)$ composed by a response variable Y and by a vector of covariates $\mathbf{X}$. Statistical properties are investigated from both a theoretical and a practical point of view. Some particular models, based on Gaussian and t distributions as well as on generalized linear models, will be introduced and maximum likelihood parameter estimation is presented. Extension to high-dimensional data modeling is finally outlined. Theoretical results are illustrated using some empirical studies, considering both simulated and real data.

E378: **Mixtures of cluster-weighted models with latent factor analyzer structure**
Presenter: **Sanjeena Dang**, University of Guelph, Canada
Co-authors: Antonio Punzo, Salvatore Ingrassia, Paul McNicholas

Cluster-weighted modelling (CWM) is a flexible statistical framework for modelling local relationships in heterogeneous populations using weighted combinations of local models. Cluster-weighted models are extended to include an underlying latent factor structure resulting in a family of Gaussian parsimonious cluster-weighted factor analyzers (CWFA) and a robust family of parsimonious cluster-weighted t-factor analyzers (CWtFA). In addition to the latent factor structure, CWtFA also contains the common factor analyzer structures. This provides even more parsimony, visualization options, and added flexibility when clustering high-dimensional data. The expectation-maximization framework along with Bayesian information criterion will be used for parameter estimation and model selection. The approach is illustrated on simulated data sets as well as a real data set.

E379: **A family of multiple response parsimonious cluster-weighted models**
Presenter: **Utkarsh Dang**, University of Guelph, Canada
Co-authors: Salvatore Ingrassia, Paul McNicholas

An extension to cluster-weighted modelling that can deal with multivariate responses is proposed and illustrated. These mixture models account for regression relationships in the data and also explicitly model the distribution of the covariates. Maximum likelihood estimation of the parameters is done via an EM framework. A parsimonious family of models is also proposed by imposing constraints on the component covariance matrices. The Bayesian information criterion is used for model selection. Lastly, performance of the model on simulated and real data is discussed.

E683: **Heterogeneity and healthcare structures effectivenes: The proposal of a cluster-weighted multilevel model**
Presenter: **Giorgio Vittadini**, University Milano Bicocca, Italy
Co-authors: Paolo Berta

To compare relative effectiveness of healthcare structure risk adjustment statistical models are utilized. Particularly quite an extensive literature has proposed the use of Multilevel Models (MM). They can be usefully utilized for studying relationships between health outcomes and risk adjustment variables in hierarchical structures, considering simultaneously both individual and aggregate levels of analysis. But, when the heterogeneity of patients within-hospitals is big and, in consequence the confidence intervals width of hospitals effectiveness parameters is large, the comparison between hospitals by means of MM is very difficult. In order to overcome the problem of heterogeneity, Multilevel regression Mixture Models (MRMM) have been proposed. MRMM show that unobserved heterogeneity in the form of latent classes can be mistaken for level 2 heterogeneity in the form of the random effects that are used in conventional two-level regression analysis. For the same goal Cluster Weighted Multilevel Models (CWMM), more general and more flexible than CWMM can be proposed, increasing the possibility of capturing the latent heterogeneity and individuating the subpopulations characterized by homogeneous effectiveness present in different hospitals.

E705: **Robustness and asymptotics properties of trimmed cluster-weighted restricted modeling**
Presenter: **Alfonso Gordaliza**, Universidad de Valladolid, Spain
Co-authors: Luis A. Garcia-Escudero, Francesca Greselin, Salvatore Ingrassia, Agustin Mayo-Iscar

Cluster-weighted modeling (CWM) is a mixture approach to modeling the joint probability of data coming from a heterogeneous population. CWM includes mixtures of distributions and mixtures of regressions as special cases. The addition of trimming and constraints to the Gaussian CWM estimation methodology is shown to be useful in order to provide robustness and to avoid not only the singularities of the objective function, but also the appearance of spurious solutions. We give theoretical results on existence and consistency and we study the robustness properties of the proposed methodology.

ES38 Room Chancellor's BAYESIAN NONPARAMETRIC REGRESSION Chair: George Karabatsos

E1240: **Nonparametric prior specification through mean functionals**
Presenter: **Antonio Lijoi**, University of Pavia, Italy
The aim is to discuss the determination of the base measure of a nonparametric prior, which yields a given distribution for the corresponding mean functional. In particular, we focus on the two-parameter Poisson-Dirichlet process, also known as the Pitman-Yor process. An interesting application is related to hierarchical mixture models typically used for clustering and density estimation. Then, we provide an illustration based on specific examples that allow us to draw a comparison with known results in the Dirichlet case.

E541: **A Bayesian nonparametric regression model with normalized weights**
Presenter: **Sara Wade**, University of Cambridge, United Kingdom
Co-authors: Isadora Antonio Villalobos, Stephen Walker
The aim is to develop an interpretable Bayesian nonparametric regression model which allows inference with combinations of both continuous and discrete covariates. Simple arguments regarding the interpretation of Bayesian nonparametric regression mixtures lead naturally to regression weights based on normalized sums. Difficulty in working with the intractable normalizing constant is overcome thanks to recent advances in MCMC methods and the development of a novel auxiliary variable scheme. The new model and MCMC method are applied to study the dynamics of hippocampal volume in Alzheimer's disease, and the results provide statistical evidence in support of recent theoretical hypotheses.

E748: **Geometric averaging of models**
Presenter: **Stephen Walker**, University of Texas at Austin, United States
Co-authors: Chris Holmes
The usual procedure of model averaging involves the predominant use of averages using summation. The aim is to propose and motivate the alternative idea of using a geometric average of models. Illustrations will be presented.

E540: **Bayesian nonparametric predictive approaches for causal inference**
Presenter: **George Karabatsos**, University of Illinois-Chicago, United States
A Bayesian nonparametric regression model for causal inference is proposed. The model can provide estimates of causal effects, in terms of how a treatment condition, versus a control condition, changes the posterior predictive mean, quantiles, variance, p.d.f., and/or c.d.f. of the outcome variable. Because of the model's flexibility, such estimates can either be based on the conditional average treatment effect, or based on a regression discontinuity design, and they do not require a preliminary matching of units (e.g., subjects). The regression model is an infinite-mixture model that has normal component densities, and has covariate-dependent mixture weights that are modeled by a latent Gaussian process. The regression model predicts a probability density that becomes increasingly unimodal (increasingly multimodal, respectively) as the explanatory power of the covariate(s) increases (decreases, respectively), and the model allows the explanatory power to vary from one value of the covariate(s) to another. Previous research showed that the model tends to have better predictive performance than many other flexible regression models. These Bayesian nonparametric regression approaches are illustrated, through the causal analysis of observational data arising from education.

ES39 Room Holden SCALING PROBLEMS IN STATISTICS Chair: Thomas Kneib

E113: **Exploring technological and biological frontiers for the identification of genetic disease factors**
Presenter: **Saskia Freytag**, Georg-August University Goettingen, Germany
The fundamental pursuit of genetic epidemiology is to determine and model genetic factors that explain the existence of chronic diseases in humans. In recent years, technologies and analysis methods, specifically developed for this pursuit, have undergone a rapid change of scale. Sequencing genetic variations used to be possible for small regions of the human genome only. Now it is routinely done on a genome-wide scale for large case-control studies. We show that while sequencing of more genetic variants increases genome coverage, it does not necessarily increase prediction ability. Analysis methods evolved with the need to analyse ever larger numbers of genetic variants and overcome the associated multiple testing problem. One development is to test sets of genetic variants, like genes or pathways, for their association with a disease. Even though this approach is often biologically insightful by suggesting perturbed mechanisms, it also bears unique challenges. In particular, we demonstrate that different sizes of genes or pathways need to be corrected for through appropriate methodological adjustments in order to prevent inflation of type I error.

E310: **Plant invasions as a scaling problem in ecology**
Presenter: **Janina Radny**, University of Goettingen, Germany
Co-authors: Katrin Meyer
Spatial and temporal scaling is an important issue in ecology. Certain ecological processes are spatially restricted to local scales, such as competitive interactions between plant individuals. However, a spatially restricted process can influence patterns on a much larger scale such as the abundance of species or genetic diversity. This leads to two general scaling problems in ecology: applying the appropriate scale to a process and extrapolating from small-scale processes to a larger-scale patterns. As a case study, an ecological simulation model will be presented, designed to estimate quantitative and qualitative differences of invasion success of functionally different types of exotic plants. While invasions occur at large scales, invasive success of a species is crucially influenced by the small-scale establishment success of single individuals. Earlier modelling approaches have mostly focused on one scale only, predicting either local establishment or large-scale spread of populations. The proposed model is operating on two nested spatial and temporal scales, calculating long-term movement in space as outcome of short-term community dynamics in subsets of the overall region. Model presentation will emphasize the upscaling methods that were used to link local processes to large-scale invasion patterns.

E459: **Hospital efficiency under prospective reimbursement schemes: an empirical assessment for the case of Germany**
Presenter: **Helmut Herwartz**, Georg-August-University Goettingen, Germany
The introduction of prospective hospital reimbursement based on diagnosis related groups (DRG) has been a conspicuous attempt to decelerate the steady increase of hospital expenditures in the German health sector. The effect of the financial reform on hospital efficiency is subjected to empirical testing by means of two complementary testing approaches. On the one hand, a two-stage procedure based on non-parametric efficiency measurement is applied. On the other hand, a stochastic frontier model is employed that allows a one-step estimation of both production frontier parameters and inefficiency effects. To identify efficiency gains as a consequence of changes in the hospital incentive structure it accounts for technological progress, spatial dependence and hospital heterogeneity. The results of both approaches do not reveal any increase in overall efficiency after the DRG reform. In contrast, a significant decline in overall hospital efficiency over time is observed.

E682: **Scaling problems in model choice and variable selection**
Presenter: **Benjamin Saefken**, Georg-August University Goettingen, Germany
Exponential families allow us to analyze data on different scales like binary data, count data or strictly positive data. The mixed model framework extends generalized linear models to generalized mixed models. The possible dependent variables can be for example nonlinear effects, continuous or discrete spatial effects or cluster specific effects. These models offer a broad range of different modeling possibilities. Therefore model choice

and variable selection is an essential part of the fitting process. If the model selection is based on choosing the model with the lowest estimated prediction error, scaling problems arise. While for continuous data from a Gaussian distribution and count data from a Poisson distribution unbiased estimates for the expected prediction error exist, there are no equivalent estimates for binary data from a Bernoulli distribution. As inference in generalized mixed models is based on refitting Gaussian models, it seems appropriate to apply the estimates of the expected prediction error for Gaussian models to other exponential family distributions. The impact of the resulting criteria on the model choice behaviour has to be analyzed carefully since small changes can have large impact on the model choice.

ES50 Room Montague COMPUTATIONAL APPROACHES IN FINANCIAL ECONOMICS Chair: Dietmar Maringer

E404: **Updating views by learning from the others: Dynamically combining asset allocation strategies**
Presenter: **Bjoern Fastrich**, University of Giessen, Germany
The well-known difficulties in obtaining satisfactory results with Markowitz' intuitive portfolio theory have led to an innumerable amount of proposed advancements by researchers and practitioners. As different as these approaches are, they typically appear to exhibit a satisfactory out-of-sample performance; however, at the same time, studies show that the equally weighted portfolio still cannot be dominated. The starting point of the study is therefore not an(other) entirely new idea, which is based on a new strategy we claim performs well, but instead the acknowledgement that the strategies proposed in earlier studies have specific advantages, which, though not consistently apparent, might prevail in specific situations of dynamic markets. A strategy is therefore proposed that "learns from" a population of already existing strategies and dynamically combines their respective characteristics, resulting in a strategy that is expected to perform best in light of the predicted market situation. The success of the approach is shown by carrying out an empirical backtest study for investor clienteles with mean-variance, mean-conditional value-at-risk, and maximum Omega utility functions. The improvements of the flexible approach stay significant, even when transaction costs are taken into account and when the competing strategies are improved by employing robust input parameter estimates.

E475: **Sparse and robust portfolio selection by penalized q-entropy minimization**
Presenter: **Sandra Paterlini**, European Business School Germany, Germany
Co-authors: Davide Ferrari, Margherita Giuzio
In asset allocation, two important problems are the sensitivity of the allocation weights to estimation error and the high dimensionality of the set of candidate assets. A new approach is proposed to construct optimal portfolios by minimizing a generalized description length criterion, where the information about the model and the data are coded by a generalized entropy measure. Model selection and estimation are performed in a single step, since the criterion explicitly incorporates information from the model by considering a prior distribution on the model parameters to enhance sparsity. The resulting portfolios are doubly robust, as they can tolerate deviations from both the assumed data model and the prior distribution for the model parameters. Empirical results on simulated and real-world data support the validity of the approach in comparison to state-of-art benchmark.

E201: **Transition variable selection for regime-switching recurrent reinforcement learning**
Presenter: **Jin Zhang**, University of Basel, Switzerland
Co-authors: Dietmar Maringer
Non-linear time series models, such as regime-switching (RS), have become increasingly popular in economics. Regime-switching recurrent reinforcement learning (RSRRL), a combined technique of statistical modeling and machine learning, has been suggested by researchers as a strategy for enhancing trading profits by modeling the nonlinear dynamics of stock returns. We addressed the transition variable selection issue in RSRRL. Four indicators (volume, relative strength index, implied volatility and conditional volatility) were considered as possible options for transition variable selection in RSRRL. We were interested in answering two questions: does the conventional linearity against smooth transition autoregressive (STAR) test assist in selecting the transition variable; and which indicator is more appropriate for improving profitability? Optimization heuristics was chosen to estimate the transition rate and the threshold value for RSRRL traders, as conventional numerical approaches tend to report local optima while estimating STAR parameters. Of the four indicators, it was found that the RSRRL with the volume indicator produced higher Sharpe ratios than others, although the RSI was the most suitable indicator according to the linearity against the STAR test.

E815: **Estimating time series models with heuristic methods: The case of economic parity conditions**
Presenter: **Sebastian Deininger**, University of Basel, Switzerland
Co-authors: Dietmar Maringer
Belonging to the class of multivariate time series methods, Vector Error Correction (VEC) models turned out to serve a suitable instrument for financial economics modelling where current changes in variables are explained by past, lagged changes. Allowing for more lags is often highly desirable as it helps model delayed reactions, long memory, or seasonalities. At the same time, however, this quickly increases the number of parameters to estimate and might lead to undesirable overfitting. Blanking out less important parameters seems a good compromise, yet at the price of eluding solvability with traditional methods. The aim is to investigate how Differential Evolution (DE) can tackle the parameter selection and estimation problem simultaneously. DE iteratively generates candidate solutions by crossing over an existing solution with a linear combination of another two. Evolutionary principles then decide whether the new replaces an existing candidate. This approach is applied on data for the US, the Euro Area, and Switzerland to investigate the concepts of the uncovered interest rate parity and the expectation hypothesis of the term structure. Under the Bayesian Information Criterion, the results indicate that for the considered currencies and economic regions, only some, but not all of these parities hold. Other criteria sometimes produce conflicting results.

ES64 Room Bedford INDEPENDENT COMPONENT ANALYSIS Chair: Davy Paindaveine

E031: **Independent component analysis via nonparametric maximum likelihood estimation**
Presenter: **Richard Samworth**, University of Cambridge, United Kingdom
Co-authors: Ming Yuan
Independent Component Analysis (ICA) models are very popular semiparametric models in which we observe independent copies of a random vector $X = AS$, where A is a non-singular matrix and S has independent components. We propose a new way of estimating the unmixing matrix $W = A^{-1}$ and the marginal distributions of the components of S using nonparametric maximum likelihood. Specifically, we study the projection of the empirical distribution onto the subset of ICA distributions having log-concave marginals. We show that, from the point of view of estimating the unmixing matrix, it makes no difference whether or not the log-concavity is correctly specified. The approach is further justified by both theoretical results and a simulation study.

E059: **On asymptotic properties of the scatter matrix based estimates for complex valued independent component analysis**
Presenter: **Pauliina Ilmonen**, Universite libre de Bruxelles, Belgium
In the complex valued independent component (IC) model it is assumed that $X = (x_1, ..., x_n)$ is an i.i.d. random sample from a p-variate distribution and that $x_i = \Omega z_i + \mu,\ \ i = 1, ..., n$, where μ is a location vector, Ω is a full-rank $p \times p$ complex valued mixing matrix, and z_i arises from a complex valued distribution with mutually independent and standardized components. The goal in the independent component analysis (ICA) is to find an

estimate for any Γ such that Γx_i has independent components. We consider the standard complex valued IC model, and the asymptotic properties of the complex valued unmixing matrix estimates that are based on simultaneous use of two scatter matrix functionals.

E235: **Independent component analysis based on fourth moments - FOBI and JADE**
Presenter: **Hannu Oja**, University of Turku, Finland
Co-authors: Jari Miettinen, Klaus Nordhausen, Sara Taskinen

Perhaps the simplest independent component functional is the FOBI functional (fourth order blind identification), that is obtained by simultaneously diagonalizing the covariance matrix and the matrix of fourth moments. Unfortunately FOBI can only separate components with distinct kurtosis values. In JADE (joint approximate diagonalization of eigen matrices) this problem is circumvented by jointly diagonalizing p(p+1)/2 fourth order cumulant matrices. JADE is then computationally heavy as the number of the cumulant matrices grows fast with the number of independent components. We show that FOBI and JADE estimates are affine equivariant and that, under general assumptions, their limiting distributions are multivariate normal. We compare their limiting efficiencies and propose a new fastJADE procedure that provides an efficient and fast tool for practical data analysis.

E676: **On infinite-dimensional independent component analysis**
Presenter: **Harold Gutch**, Max Planck Institute for Dynamics and Self-Organization, Germany
Co-authors: Fabian Theis

The original Independent Component Analysis (ICA) problem of blindly separating a mixture of a finite number of real-valued statistically independent one-dimensional sources has been extended in a number of ways in recent years. These include dropping the assumption that all sources are one-dimensional and some generalizations to the case where the sources are not real-valued; however up to recently all approaches assumed a finite number of sources. We present a generalization of ICA to the case of infinitely many sources. While it appears that the usual definition of independence of infinitely many random variables is too weak to allow ICA, it turns out that this is not the case, and the decomposition in this setting indeed is unique, up to the usual indeterminancies of ICA, permutation and rescaling. We discuss algorithmic aspects on how to model and generate infinite-dimensional toy data and possible applications of the model.

ES66 Room Bloomsbury TIME SERIES ANALYSIS Chair: Alessandra Luati

E315: **Nowcasting raw industrial production series: the role of survey data**
Presenter: **Barbara Guardabascio**, ISTAT - Italian Statistic Institute, Italy
Co-authors: Marco Ventura, Alessandro Girardi

A forecasting model is developed for the (raw) industrial production index, the most commonly used monthly hard indicator to describe the cyclical phase of the economy. Moreover, the use of raw series as a target and (almost) unrevised predictors makes it possible to evaluate forecast performances in a real-time context. In particular, the proposed modelling strategy consists of three steps. First, we use an ARDL model with a few pre-selected hard indicators. Second, we construct a PLS model by projecting the residuals from the first step on a large set of several hundred soft indicators. Third, the forecasting equation at several horizons is built by augmenting the ARDL model by the PLS factors from the second step. Our modelling approach also takes advantage of the jagged edge nature of both qualitative and quantitative series given that our predictors are disclosed one month ahead of the target series. Using monthly Italian data, we document that our approach outperforms a number of alternative benchmark models in terms of RMSFE, especially at the very short forecast horizons. The use of qualitative data turns out to be useful to improve forecasts of quantitative aggregates, as the gains in terms of forecast ability are statistically significant.

E735: **On spatial models for multivariate time series analysis**
Presenter: **Maria Lucia Parrella**, University of Salerno, Italy
Co-authors: Qiwei Yao, Baojun Dou

Many datasets are derived by observations taken at time t from p locations over space, where the space can be physical or economic in nature. It is common practice in spatial econometrics to consider a spatial weight matrix, summarizing the distances among the locations, and to assume it to be known. It can simply record the geographical distances between the locations or the distances reflecting the correlation or association between the variables at the different locations. The spatial models proposed in the literature can be seen as multivariate linear regression models. They generally consider 'fixed effect' coefficients, which are constant over time but adaptive over space. On the other side, the slope coefficient is considered equal for all the locations. This is, in our opinion, a paradox for spatial models. In fact, using a 'fixed' slope forces each series to have the same reaction to 'neighbour data', in contrast with the same philosophy of spatial analysis. We propose an adaptive spatial autoregressive model which considers a different slope coefficient for each location. We suggest a method to estimate the coefficients and investigate the performance on simulated datasets. Finally, we discuss some relations among these models and the classic VAR models.

E626: **GMM inference for costationary systems and efficient covariance estimation for locally stationary processes**
Presenter: **Alessandro Cardinali**, Bristol, United Kingdom

A novel estimator for the time-varying covariance of locally stationary time series is proposed. This new approach is based on costationary combinations, that is, time-varying deterministic linear combinations of locally stationary time series that are second-order stationary. The aim is to first review the theory of costationarity and formalize a GMM estimator for the coefficients vectors. Then the concept of costationary factors is introduced in order to identify representative costationary combinations, since these are typically multiple. This new framework is then used to derive an efficient covariance estimator. It is shown that the new approach has smaller variance than the available estimators exclusively based on the evolutionary cross-periodogram, and is therefore appealing in a large number of applications. The theoretical findings are confirmed through a simulation experiment. This shows that the approach improves substantially over competitors for finite sample sizes which are of common use. Then, a new analysis is presented of the FTSE and SP500 log return series. DEM/USD and GBP/USD exchange rate return series are also analysed. The new estimator is shown to compare favorably with existing approaches and is capable to highlight certain economic shocks in a clearer manner.

E1211: **Large precision matrix estimation via pairwise tilting**
Presenter: **Na Huang**, London School of Economics, United Kingdom
Co-authors: Piotr Fryzlewicz

A *tilting*-based method is proposed to estimate the precision matrix of a p-dimensional random variable, $\mathbf{X}$, when p is possibly much larger than the sample size n. Each 2×2 block indexed by (i, j) of the precision matrix can be estimated by the inversion of the pairwise sample conditional covariance matrix of X_i and X_j controlling for all the other variables. However, in the high dimensional setting, including too many or irrelevant controlling variables may distort the results. To determine the controlling subsets in high dimensional scenarios, the proposed method applies the *tilting* technique to measure the contribution of each remaining variable to the variance-covariance matrix of X_i and X_j, and only puts the (hopefully) highly relevant remaining variables into the controlling subsets. We illustrate conditions under which it can successfully distinguish the highly relevant remaining variables from the rest. The simulation results will be presented under different scenarios for the underlying precision matrix. Comparison with other competing methods will also be given.

ES102 Room 349 JAPAN STATISTICAL SOCIETY: STATISTICS FOR STOCHASTIC DIFFERENTIAL EQUATIONS Chair: Masayuki Uchida

E408: **Intraday periodicity and lead-lag effects in financial markets**
Presenter: **Yuta Koike**, University of Tokyo, Japan
A new approach for modeling and estimating the lead-lag effect between two assets is proposed. The model can be accommodated to nonsynchronous trading and market microstructure noise as well as the intraday periodicity of lead-lag effects, which are essential for empirical work. In this model the lead-lag effect can be considered as a kind of endogenous microstructure noise induced by stale prices, and the time lag parameter is interpreted as the variance of this noise. An estimator for the time lag parameter is constructed in line with this idea. Its consistency is shown in a high-frequency observation setting. Unlike traditional approaches in this area, the estimator does not rely on an optimization of the covariation between two assets, and it is easier to compute.

E479: **Maximum likelihood type estimation and Bayes type estimation for nonsynchronously observed diffusion processes**
Presenter: **Teppei Ogihara**, Osaka University, Japan
Co-authors: Nakahiro Yoshida
The aim is to study maximum likelihood type estimation and Bayes type estimation for a two-dimensional diffusion process with nonsynchronous trading. The problem of nonsynchronous observations appears when estimating the covariance of security returns using high-frequency financial data. Previous authors point out the realized covariance estimator has serious bias in the case of nonsynchronous observation. They construct an unbiased estimator of the covariation of two diffusion processes by nonsynchronously observed data with a nonparametric approach, and prove consistency of the estimator. The proposed approach to the problem of nonsynchronous observations is to construct a quasi-likelihood function H and study the asymptotic behavior of the maximum-likelihood type estimator and the Bayes type estimator for H when the lengths of observation intervals converge to 0. For this purpose, we follow an approach of likelihood ratio random fields, and use polynomial type large deviation inequalities. Performances of a previous nonparametric estimator and the maximum likelihood type estimator are compared by a simple simulation.

E547: **Stable quasi-likelihood: Methodology and computational aspects**
Presenter: **Hiroki Masuda**, Kyushu University, Japan
We consider the semi-parametric model described by the stochastic differential equation $dX_t = a(X_t,\alpha)dt + c(X_{t-},\gamma)dJ_t$, where a and c are known except for the finite-dimensional parameter $\theta = (\alpha,\gamma)$, and J is a pure-jump Levy process which is "stable-like" in the sense that the law $\mathcal{L}(h^{-1/\beta}J_h)$ tends as $h \to 0$ to the symmetric β-stable distribution. The wish is to estimate the true value $\theta_0 = (\alpha_0,\gamma_0)$ based on a high-frequency sample $X_{t_0}, X_{t_1}, \ldots, X_{t_n}$, where $t_j = t_j^n = jh_n$ with $h_n \to 0$ as $n \to \infty$. A novel, tailor-made estimator is introduced, $\hat{\theta}_n = (\hat{\alpha}_n,\hat{\gamma}_n)$, based on the stable approximation of the one-step transition distribution $\mathcal{L}(X_{t_j}|X_{t_{j-1}})$. Under suitable regularity conditions, it is shown that $\mathcal{L}\{\sqrt{n}h_n^{1-1/\beta}(\hat{\alpha}_n-\alpha_0), \sqrt{n}(\hat{\gamma}_n-\gamma_0)\}$ is asymptotically mixed-normal (resp. normal) when t_n is fixed (resp. when $t_n \to \infty$ and X is ergodic). The result reveals that, in case of the stable-like J, the proposed estimator $\hat{\theta}_n$ is much more efficient than the conventional Gaussian quasi-maximum likelihood estimator, which requires $t_n \to \infty$ and leads to $\sqrt{t_n}$-asymptotic normality for both α and γ. Nevertheless, evaluation of $\hat{\theta}_n$ is computationally more involved compared with the Gaussian case. Also discussed in some detail are the computational aspects of the proposed methodology.

E832: **Inference for volatility: Some theoretical aspects**
Presenter: **Nakahiro Yoshida**, University of Tokyo, Japan
The quasi likelihood analysis (QLA) is a systematic analysis of the quasi likelihood random field and the associated QLA estimators, i.e., quasi maximum likelihood estimator and quasi Bayesian estimators, with large deviation estimates for tail probability of the random field and QLA estimators. A polynomial type large deviation (PLD) inequality was generally established for a locally asymptotically quadratic quasi log likelihood random field, and it was applied to ergodic diffusion/jump-diffusion processes in order to form QLA. This scheme works in estimation of the volatility parameter of a sampled semimartingale in finite time horizon. The convergence of moments as well as asymptotic mixed normality of the QLA estimators have been proven. Then verifying nondegeneracy of the statistical random field is crucial in the derivation of the PLD inequality in this non- ergodic setting. Recently QLA for volatility of non-synchronously observed diffusion type processes has been presented. Asymptotic expansion for volatility estimators (i.e., realized volatility, p-variation, QLA estimator) follows from martingale expansion with mixed normal limit. QLA enables us to obtain asymptotic expansion of the QLA estimator. Volatility information criteria for diffusion model selection are obtained by using QLA and methods in higher-order inference.

ES79 Room Court ROBUST ANALYSIS OF COMPLEX DATA Chair: Stefan Van Aelst

E317: **A fast algorithm for S-estimation in robust heteroscedastic regression**
Presenter: **Peter Slock**, Ghent University, Belgium
Co-authors: Stefan van Aelst, Matias Salibian-Barrera
Starting from the fast-S algorithm for homoscedastic linear regression, an efficient algorithm for S-estimators in the heteroscedastic linear regression model is developed. Heteroscedastic S-estimators minimize an M-scale of residuals which have been transformed to homoscedasticity. The method consists of 2 intertwined minimization problems: (i) Estimation of regression coefficients β that minimize the M-scale $\hat{\sigma}$ of transformed residuals, and (ii) optimally transforming the residuals by minimizing a test statistic expressing the degree of heteroscedasticity, where the variance of the response y_i is assumed to be proportional to $h(\gamma'\mathbf{x_i})$ with a known function h, and those values of heteroscedasticity parameters γ are found that result in the optimal transformation. The algorithm starts from a large number of subsamples and seeks the global optimum for β by iterative improvement steps in order to reduce the M-scale estimate $\hat{\sigma}$. A similar approach is applied in an attempt to minimize the homoscedasticity test statistic for each fixed set of β coefficients. Results of simulation studies will be presented to compare the accuracy and computation times of several variants of the heteroscedastic S-estimator algorithm. Based on the newly developed algorithm the robustness and efficiency of heteroscedastic S-estimators is also compared to the heteroscedastic MLE.

E666: **A deterministic algorithm for LTS**
Presenter: **Kaveh Vakili**, KU LEUVEN, Belgium
Co-authors: Mia Hubert, Peter Rousseeuw, Tim Verdonck
The least trimmed of squares (LTS) method is a high-breakdown robust estimator of regression. Computing the exact LTS is very hard, so in practice one resorts to approximate algorithms. Most often the FASTLTS algorithm is used. This algorithm starts by drawing many random subsets, followed by so-called concentration steps. The FASTLTS algorithm is affine and regression equivariant but not permutation invariant. A deterministic algorithm is presented, denoted as DetLTS, which does not use random subsets and is even faster. It is permutation invariant and very close to affine and regression equivariant. DetLTS is illustrated on real and simulated data sets.

E772: **P-splines quantile regression in varying coefficient models**
Presenter: **Irene Gijbels**, Katholieke Universiteit Leuven, Belgium
Co-authors: Yudhie Andriyana, Anneleen Verhasselt

Quantile regression in varying coefficient models for longitudinal data is studied. The quantile function is modeled as a function of the covariates and the main task is to estimate the unknown regression coefficient functions. We approximate each coefficient function by means of P-splines. Theoretical properties of the estimators, such as rate of convergence and an asymptotic distribution are established. The estimation methodology requires solving an optimization problem that also involves a smoothing parameter. For a special case the optimization problem can be transformed into a linear programming problem for which then a Frisch-Newton interior point method is used, leading to a computationally fast and efficient procedure. Several data-driven choices of the smoothing parameters are briefly discussed, and their performances are illustrated in a simulation study. Some real data analysis demonstrates the use of the developed method. A classical problem in quantile regression estimation, in particular for small data sets, is quantile crossing. We discuss several methods for avoiding the quantile crossing in the context of varying coefficient models.

E848: **Robust analysis of choice data**
Presenter: **Dries Benoit**, Ghent University, Belgium
Co-authors: Stefan van Aelst, Dirk van den Poel

A Bayesian method for robust estimation of multinomial choice models is presented. The method can be used for both correlated as well as uncorrelated choice alternatives. To account for outliers in the response direction, the fat tailed multivariate Laplace distribution is used. In addition, a shrinkage procedure is applied to handle outliers in the independent variables as well. By exploiting the scale mixture of normals representation of the multivariate Laplace distribution, an efficient Gibbs sampling algorithm is developed. A simulation study shows that estimation of the model parameters is less influenced by outliers compared to non-robust alternatives, even when the data generating model deviates from the assumed model. An analysis of margarine scanner data shows how our method can be used for better pricing decisions.

ES83 Room Woburn SEMI-AND-NON PARAMETRIC FUNCTIONAL DATA ANALYSIS Chair: Aldo Goia

E581: **The interval wise error rate for functional data: The interval testing procedure in action**
Presenter: **Simone Vantini**, Politecnico di Milano, Italy
Co-authors: Alessia Pini

A semi-parametric inferential procedure is presented that enables inference when testing for differences between two functional populations, in both coupled and uncoupled scenarios. The procedure is composed of three steps: (i) represent functional data on a suitable ordered functional basis; (ii) jointly perform univariate permutation tests on the expansion coefficients; (iii) combine the univariate results obtaining corrected p-values to control the Interval Wise Error Rate (IWER) that it is here defined. This type of control is particularly suited for functional data. Indeed, for instance, the control of the IWER, which lies in between the weak and the strong control of the Family Wise Error Rate, implies that, given any interval of the domain in which there is no difference between the two functional populations, the probability that at least a part of the domain is wrongly detected as significant is always controlled. The aim is to give theoretical bounds for the ITP family-wise and comparison-wise probability of rejection of the null hypothesis when true or false, and compare the ITP with other multiple testing approaches, such as the Bonferroni-Holm and Benjamini-Hochberg procedures. The ITP and some of its possible extensions are implemented in fdatest R package.

E445: **Functional kernel smoothing methods applied to segment hyperspectral images**
Presenter: **Laurent Delsol**, University of Orleans, France
Co-authors: Cecile Louchet

Splitting a picture into a set of homogeneous 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. Hyperspectral images are considered for which each pixel is described through a curve (discretized on a thin grid) and we discuss the way functional kernel density estimation and maximum a posteriori approach may be combined.

E574: **Functional data analysis for three-dimensional curves using Frenet-Serret framework**
Presenter: **Juhyun Park**, Lancaster University, United Kingdom
Co-authors: Nicolas Brunel

The main purpose of curve registration is to recover a *structural mean* of the curves by taking into account the common structure. For one dimensional curves the concept of phase and amplitude variability is easily understood as horizontal and vertical variability, respectively. Consequently, visual impression of the variability of the curves and the degree of alignment of salient features are often considered sufficient to judge the need or success of curve registration. However, such notion cannot be directly extended to higher dimensional curves. An easy way to get around such problem would be to summarize the multi-dimensional curves into one dimensional features in terms of first or second derivatives and apply one dimensional curve registration methods, either to the one-dimensional features or simultaneously to all marginal curves. An alternative approach is proposed to directly deal with three-dimensional curves borrowing ideas from shape analysis. This framework is used to remove phase variation and to define a mean shape for three dimensional curves. Shape is an intrinsic geometric feature of the object, which is independent of the underlying coordinate systems to represent the object. The shape of the curves is analyzed by means of the Frenet-Serret representation, which provides a flexible coordinate system driven by the geometric features of the curve itself. Unlike standard shape matching, the method effectively regularizes the estimation of the geometry through curvature and torsion, and does not require curve alignment to define a structural mean. The methods are demonstrated with some real data examples.

E442: **A partitioned single functional index model**
Presenter: **Aldo Goia**, University of Novara, Italy
Co-authors: Philippe Vieu

Given a functional regression model with scalar response, the aim is to present a methodology in order to approximate in a semi-parametric way the unknown regression operator through a single index approach, but taking possible structural changes into account. After presenting the methodology, its behaviour is illustrated both on simulated and real curves datasets. It appears, from an example of interest in the spectrometry, that the method provides a nice exploratory tool both for analyzing structural changes in the spectrum and for visualizing the most informative directions, still keeping good predictive power. Even if the main objective of the work is to discuss applied issues of the method, asymptotic behaviour are shortly described.

ES86 Room Deller ADVANCES IN QUANTILE REGRESSION Chair: Stanislav Volgushev

E033: **Estimation in functional linear quantile regression**
Presenter: **Kengo Kato**, University of Tokyo, Japan

The aim is to study estimation problems in functional linear quantile regressions in which the dependent variables are scalar while the covariates are functions, and the conditional quantiles for each fixed quantile index are modeled as a linear functionals of the covariate. We suppose that the covariates are discretely observed and sampling points may differ across subjects, where the number of measurements per subject increases as the sample size. Also, we allow the quantile index to vary over a given subset of the open unit interval, so the slope function is a function of two variables: (typically) time and quantile index. Likewise, the conditional quantile function is a function of the quantile index and the covariate. We consider an estimator for the slope function based on the principal component basis. An estimator for the conditional quantile function is obtained by a plug-in method. Since the so-constructed plug-in estimator not necessarily satisfies the monotonicity constraint with respect to the quantile index, we also consider a class of monotonized estimators for the conditional quantile function. We establish rates of convergence for these estimators under suitable norms, showing that these rates are optimal in a minimax sense under some smoothness assumptions on the covariance kernel of the covariate and the slope function. Empirical choice of the cutoff level is studied by using simulations.

E296: **Censored quantile regression via Box-Cox transformations under conditional independence**
Presenter: **Chenlei Leng**, University of Warwick, United Kingdom
Co-authors: Xingwei Tong

A new quantile regression model when data are subject to censoring is proposed. In comparison with some existing approaches, our model requires neither any global linearity assumption, nor independence of the covariates and the censoring time. We further develop a class of power-transformed quantile regression models such that the transformed survival time can be better characterized by linear regression quantiles. Consistency and asymptotic normality of the resulting estimators are shown. A re-sampling based approach is proposed for statistical inference. Empirically, the new estimator is shown to outperform its competitors under conditional independence, and to perform similarly under unconditional independence. The proposed method is illustrated with a real data analysis.

E753: **Quantile curve estimation for locally strictly stationary processes**
Presenter: **Wei Biao Wu**, University of Chicago, United States

Estimation of time-varying quantiles is discussed for locally strictly stationary processes in which the data-generating mechanism varies with respect to time. Uniform consistency and central limit theory will be presented, and the results are applied to environmental data sets.

E756: **Semiparametric conditional quantile estimation through copula-based multivariate models**
Presenter: **Anouar El Ghouch**, The University catholique de Louvain, Belgium
Co-authors: Hohsuk Noh, Ingrid van Keilegom

A new approach is considered in quantile regression modeling based on the copula function that defines the dependence structure between the variables of interest. The key idea of this approach is to rewrite the characterization of a regression quantile in terms of a copula and marginal distributions. After the copula and the marginal distributions are estimated, the new estimator is obtained as the weighted quantile of the response variable in the model. The proposed conditional estimator has three main advantages: it applies to both *iid* and time series data, it is automatically monotonic across quantiles and it can easily consider the case of multiple covariates in the estimation without introducing any extra complication. We show the asymptotic properties of our estimator when the copula is estimated by maximizing the pseudo log-likelihood and the margins are estimated nonparametrically including the case where the copula family is misspecified. We also present the finite sample performance of the estimator and illustrate the usefulness of our proposal by an application to the historical volatilities of Google and Yahoo companies.

Sunday 15.12.2013 10:55 - 12:35 Parallel Session I – CFE

CSI01 Room Chancellor's BAYESIAN ECONOMETRICS **Chair: Dimitris Korobilis**

C292: **Robust Bayesian methods for survival analysis using rate mixtures of Weibull distributions**
Presenter: **Mark Steel**, University of Warwick, United Kingdom
Co-authors: Catalina Vallejos
Survival models such as the Weibull or log-normal lead to inference that is not robust in the presence of outliers. They also assume that all heterogeneity between individuals can be modelled through covariates. The use of infinite mixtures of lifetime distributions is considered as a solution for these two issues. This can be interpreted as the introduction of a random effect in the survival distribution. We introduce the family of Rate Mixtures of Weibull distributions, which includes the known Lomax distribution. We implement Bayesian inference in a proportional hazard regression context under a prior that combines the structure of the Jeffreys' prior and a proper (informative) prior with an elicitation strategy. We derive conditions for the existence of the posterior distribution, also taking censoring into account. In addition, a method for outlier detection based on the mixture structure is proposed. Finally, the analysis is illustrated using real datasets.

C326: **A new index of financial conditions**
Presenter: **Dimitris Korobilis**, University of Glagow, United Kingdom
Co-authors: Gary Koop
Factor augmented vector autoregressive models with time-varying coefficients are used to construct a financial conditions index (FCI). The time-variation in the parameters allows for the weights attached to each financial variable in the index to evolve over time. Furthermore, we develop methods for dynamic model averaging or selection which allow the financial variables entering into the FCI to change over time. We discuss why such extensions of the existing literature are important and show them to be so in an empirical application involving a wide range of financial variables.

C840: **Time varying sparsity in dynamic regression models**
Presenter: **Jim Griffin**, University of Kent, United Kingdom
Co-authors: Maria Kalli
Regression models can be useful for prediction. If observations are made over time, the effect of the regressors can be changing over time. A novel Bayesian method for regression models where both the value of the regression coefficients and the variables selected change over time is presented. Inference can be made using an efficient Markov chain Monte Carlo method. The method will be illustrated by application to forecasting problems in econometrics.

CS10 Room Russell MULTIVARIATE VOLATILITY MODELLING **Chair: Angeles Carnero**

C134: **Measuring persistence in volatility spillovers**
Presenter: **Christian Conrad**, Heidelberg University, Germany
Co-authors: Enzo Weber
The aim is to analyze volatility spillovers in multivariate GARCH-type models. We show that the cross-effects between the conditional variances determine the persistence of the transmitted volatility innovations. In particular, the effect of a foreign volatility innovation on a conditional variance is even more persistent than the effect of an own innovation unless it is offset by an accompanying negative variance spillover of sufficient size. Moreover, ignoring a negative variance spillover causes a downward bias in the estimate of the initial impact of the foreign volatility innovation. Applying the concept to portfolios of small and large firms, we find that shocks to small firm returns affect the large firm conditional variance once we allow for (negative) spillovers between the conditional variances themselves.

C138: **The uncertainty of conditional correlations in DCC models**
Presenter: **Esther Ruiz**, Universidad Carlos III de Madrid, Spain
Co-authors: Diego Fresoli
When forecasting conditional correlations that evolve according to a Dynamic Conditional Correlation (DCC) model, only point forecasts can be obtained at each moment of time. We use bootstrap procedures to approximate the uncertainty around future conditional correlations. We implement the procedure to forecast the correlations among daily returns of spot exchange rates of the Euro, British Pound, Canadian Dollar, Japanese Yen and Australian Dollar against the US Dollar.

C438: **Dynamic principal Component: A new MGARCH model for large systems**
Presenter: **Massimiliano Caporin**, University of Padova, Italy
Co-authors: Gian Piero Aielli
A new MGARCH model, the Dynamic Principal Component, DPC, is introduced. The DPC generalizes the OGARCH by allowing for dynamic evolution of both the eigenvalues and eigenvectors of the covariance matrix spectral decomposition. The DPC model nests as special cases the BEKK and OGARCH models. Model construction, identification, estimation and interpretation of its components are discussed. Different specifications for the model dynamic elements are provided, including also reduced rank cases (dynamic in few principal components only) and the introduction of variance targeting. Simulated experiments and an application to real data show the flexibility of the model in comparison to traditional competitors.

C813: **Multivariate variance targeting in the BEKK-GARCH model**
Presenter: **Anders Rahbek**, Copenhagen University, Denmark
Asymptotic inference is considered in the multivariate BEKK model based on (co-)variance targeting (VT). By definition the VT estimator is a two-step estimator and the theory presented is based on expansions of the modified likelihood function, or estimating function, corresponding to these two steps. Strong consistency is established under weak moment conditions, while sixth order moment restrictions are imposed to establish asymptotic normality. Included simulations indicate that the multivariately induced higher-order moment constraints are indeed necessary.

CS25 Room Montague WAVELET APPLICATIONS IN ECONOMICS Chair: Marco Gallegati

C064: **Studies in nonlinear dynamics and wavelets for business cycle analysis**
Presenter: **Peter Martey Addo**, EDEEM, France
Co-authors: Monica Billio, Dominique Guegan
A signal modality analysis to characterize and detect nonlinearity schemes in the US Industrial Production Index time series is provided. The analysis is achieved by using the recently proposed 'delay vector variance' (DVV) method, which examines local predictability of a signal in the phase space to detect the presence of determinism and nonlinearity in a time series. Optimal embedding parameters used in the DVV analysis are obtained via a differential entropy based method using Fourier and wavelet-based surrogates. A complex Morlet wavelet is employed to detect and characterize the US business cycle. A comprehensive analysis of the feasibility of this approach is provided. Our results coincide with the business cycles peaks and troughs dates published by the National Bureau of Economic Research (NBER).

C557: **Gold, oil, and stocks**
Presenter: **Jozef Barunik**, Charles University in Prague, Czech Republic
Co-authors: Evzen Kocenda, Lukas Vacha
The aim is to analyze the dynamics of the prices of gold, oil, and stocks over 26 years (1987–2012) using both intra-day and daily data and employing a variety of methodologies including a novel time-frequency approach based on wavelets. It accounts for structural breaks and shows radical change in correlations between assets after the 2007-2008 crisis in terms of time-frequency behavior. No strong evidence for a specific asset leading any other one emerges and the assets under research do not share the long-term equilibrium relationship. Strong implication is that after the structural change gold, oil, and stocks cannot be used together for risk diversification.

C053: **Wavelet analysis and the forward premium anomaly**
Presenter: **Michaela Kiermeier**, University of Applied Sciences Darmstadt, Germany
Forward and corresponding spot rates on foreign exchange markets differ so that forward rates cannot be used as unbiased predictors for future spot rates which is suggested by rational expectation theory. This phenomenon has entered the literature under the heading of the forward premium anomaly. We argue that standard econometric analyses implicitly assume that the relationship is time scale independent which might be responsible for the puzzle. We use wavelet analysis to decompose the exchange rate changes and the forward premiums using the maximum overlap discrete wavelet transform and estimate the relationship on a scale-by-scale basis. The results show that the relationship does in fact hold for certain time scales, but not for all. Weekly data concerning Euro, U.S. dollar and British Pound is analysed. Performing the analysis at different time scales we cannot reject the unbiasedness hypothesis for adjustment time periods of two to three months.

C1190: **Money growth, loan growth and consumer price inflation in the euro area: A wavelet analysis**
Presenter: **Michael Scharnagl**, Deutsche Bundesbank, Germany
Co-authors: Martin Mandler
The relationship between money growth and consumer price inflation in the euro is studied using wavelet analysis. Wavelet analysis allows us to account for variations in the money growth - inflation relationship both across the frequency spectrum and across time. Our results indicate that over the sample period 1970-2012 there was no stable significant relationship between fluctuations in money growth and consumer price inflation at low frequencies. In contrast, most of the literature, by failing to account for the effects of time variation, estimated stable long-run relationships between money growth and inflation well into the 2000s. We also analyze the relationship between loan growth and inflation in the euro area, since bank loans represent the most important counterpart to monetary developments but find no evidence for a stable relationships between loan growth and inflation at any frequency.

CS29 Room Senate DYNAMIC CONDITIONAL SCORE MODELS Chair: Andrew Harvey

C469: **A score-driven approach for autoregressive models with time-varying parameters and heavy-tails**
Presenter: **Davide Delle Monache**, Queen Mary - University of London, United Kingdom
Co-authors: Ivan Petrella
The aim is to propose a flexible framework for estimating autoregressive models with time-varying parameters based on the recent developments on the Score-driven models. It is highlighted how the recursive algorithms commonly used in the macro-econometrics literature, such as learning expectations algorithms and forgetting factor algorithms, are nested within the framework under Gaussian innovations and they are generelised along various directions, in particular Student-t distribution is included. General transformations are proposed in order to impose stationarity and bounded mean in the reduced form autoregressive model. In contrast to the today mainstream time-varying-parameters models, the porposed approach does not involve Bayesian simulation techniques and it can easily be estimated by maximum likelihood. The model is applied to the analysis of inflation dynamics. The analysis provides new insights on recent debate surrounding the role of the trend-inflation, and the time variation in persistence and volatility of inflation. The model also provides accurate forecasts of inflation.

C281: **Maximum likelihood estimation for generalized autoregressive score models**
Presenter: **Francisco Blasques**, VU University Amsterdam, Netherlands
Co-authors: Siem Jan Koopman, Andre Lucas
Consistency and asymptotic normality of the maximum likelihood estimator (MLE) for a broad class of non-linear score driven time series models are considered. These models not only encompass many familiar time series models such as GARCH, ACD, ACM, and MEM, but also include many new empirically useful models, such as observation driven mixed measurement factor models and volatility and correlation models for fat-tailed observations. We formulate primitive rather than high-level conditions for global identification, consistency, and asymptotic normality of the MLE. We study both cases of well-specified and misspecified models, such that our results are also applicable outside an (arbitrarily small) neighbourhood of the true parameter value. We illustrate how to apply the theory using non-linear score driven time series models for conditional means and variances.

C410: **Asymptotic theory for Beta-t-GARCH**
Presenter: **Ryoko Ito**, Cambridge University, United Kingdom
The consistency and asymptotic normality of the maximum likelihood estimator is established for the first order Beta-t-GARCH model, which is a special class of the dynamic conditional score models. The asymptotic results hold for both stationary and nonstationary versions of the model.

C264: **Time series models with an EGB2 conditional distribution**
Presenter: **Andrew Harvey**, University of Cambridge, United Kingdom
Co-authors: Michele Caivano
A time series model in which the signal is buried in noise that is non-Gaussian may throw up observations that, when judged by the Gaussian yardstick, are outliers. We describe an observation driven model, based on an exponential generalized beta distribution of the second kind (EGB2),

in which the signal is a linear function of past values of the score of the conditional distribution. This specification produces a model that is not only easy to implement, but which also facilitates the development of a comprehensive and relatively straightforward theory for the asymptotic distribution of the maximum likelihood estimator. The model is fitted to US macroeconomic time series and compared with Gaussian and Student-t models. A theory is then developed for an EGARCH model based on the EGB2 distribution and the model is fitted to exchange rate data. Finally, dynamic location and scale models are combined and applied to data on the UK rate of inflation.

CS96 Room 349 COMPUTATIONAL ECONOMETRICS I Chair: Robert Hudson

C045: **Artificial stock market dynamics and market efficiency: An econometric perspective**
Presenter: **Viktor Manahov**, Newcastle University, United Kingdom
Co-authors: Robert Hudson
Various artificial stock markets populated with different numbers of traders are developed by using a special adaptive form of the Strongly Typed Genetic Programming (STGP)-based learning algorithm. We applied the STGP technique to real-life historical data of three indices -FTSE 100, S&P 500 and Russell 3000- to investigate the formation of stock market dynamics and market efficiency. We used several econometric techniques to investigate the emergent properties of the stock markets. We have found that the introduction of increased heterogeneity and greater genetic diversity leads to higher market efficiency in terms of the Efficient Market Hypothesis. We have also found that stock market dynamics and nonlinearity are better explained by the evolutionary process associated with the Adaptive Market Hypothesis. Hence, market efficiency exists simultaneously with the need for adaptive flexibility. Our empirical results, generated by a reduced number of boundedly rational traders in six of the stock markets, for each of the three financial instruments do not support the laissez-faire school of thought, indicating the need for governmental intervention in stock markets.

C853: **The enhanced SSpace Matlab toolbox**
Presenter: **Diego Pedregal**, University of Castilla-La Mancha, Spain
The utility of SSpace, a MATLAB toolbox for the analysis of State Space systems, is illustrated. The toolbox has been available during a number of years, but recently has been enhanced with new capabilities, like exact filtering, smoothing, disturbance smoothing, likelihood estimation, nested systems, system concatenation, etc. The key advantage of this particular toolbox over other pieces of software is its generality, flexibility and ease of use. Regarding generality, different specifications of the same dynamic system are possible because all system matrices may be time variable, covariances between state and observed noises are allowed, etc. The flexibility comes from the way the user communicates with the computer, since model implementation requires writing a Matlab function in which any sort of standard programming is allowed. This may look irrelevant at a first glance, but is a powerful way of implementing models, because it opens up the possibility to some non-linear models, different parameterization of the same models, any sort of constraints among parameters, etc. The toolbox is rather easy to use because only a few functions are necessary to take full advantage of the power of SS system modelling. An overview of SSpace is provided and its usage is demonstrated with several examples.

C934: **The efficiency of portfolio optimization: a multi-agents ecological competition analysis**
Presenter: **Iryna Veryzhenko**, CNAM, France
Previous research is extended to find out if a sophisticated portfolio optimization strategy can be outperformed by simpler assets allocation rules. We study the relative performance of investment strategies scrutinizing their behavior in an ecological competition where populations of investors co-evolve in an artificial stock market framework. We test different variations around the canonical modern portfolio theory of Markowitz, strategies based on the naive diversification principles and the combination of several strategies. We compare these diversification rules in terms of dollar wealth, Sharpe ratio and portfolio turnover. Contrary to most previous research relying on backtesting, we use agent-based simulations to study this aspect. The major benefit of our methodology is that the agent trading strategies define the market dynamics, which then impact agents' behavior and performance. We show, among others, that the best possible strategy over the long run relies on a mix of Mean-Variance sophisticated optimization and a naive diversification. We show that this result is robust when short selling is allowed in the market and whatever the performance indicator chosen to gauge the relative interest of the studied investment strategies.

C1200: **Adaptive learning and approximate aggregation in heterogeneous agent models**
Presenter: **Jan Acedanski**, University of Economics in Katowice, Poland
A standard heterogeneous agent model with aggregate uncertainty and borrowing constraints with adaptive learning schemes instead of fixed almost fully rational linear law of motion for predicting aggregate capital is studied. It is assumed that agents do not know the law of motion for aggregate capital that is derived under the approximate aggregation assumption but employ econometric techniques to predict the capital stock when taking their individual decisions. It is examined how various learning schemes, initial condition determination mechanisms as well as functional forms affect dynamics of individual and aggregate variables in the model. It is also checked if the law of motion for capital under the adaptive learning converges to the standard, fixed law of motion, i.e. whether the approximate aggregation solution is learnable.

CS41 Room Jessel BAYESIAN NONPARAMETRIC ECONOMETRICS Chair: John Maheu

C395: **Copula based factorization in Bayesian multivariate infinite mixture models**
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. 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. A factorization scheme of multivariate dependence structures based on the copula modeling framework is proposed, whereby each marginal dimension in the mixing parameter space is modeled separately and the marginals are then linked by a nonparametric random copula function. Specifically, nonparametric univariate Gaussian mixtures are considered for the marginals and a multivariate random Bernstein polynomial copula for the link function, under the Dirichlet process prior. It is shown that in a multivariate setting this scheme leads to an improvement in the precision of a density estimate relative to the commonly used multivariate Gaussian mixture. Weak posterior consistency of the copula-based mixing scheme is derived for general kernel types under high-level conditions, and strong posterior consistency for the specific Bernstein-Gaussian mixture model.

C414: **Beta-product dependent Pitman-Yor processes for Bayesian inference**
Presenter: **Roberto Casarin**, University Ca Foscari of Venice, Italy
Co-authors: Federico Bassetti, Fabrizio Leisen
Multiple time series data may exhibit clustering over time and the clustering effect may change across different series. The work is motivated by the Bayesian non-parametric modelling of the dependence between clustering effects in multiple time series analysis. A Dirichlet process mixture approach is followed and a new class of multivariate dependent Pitman-Yor processes (DPY) is defined. The proposed DPY are represented in

terms of vectors of stick-breaking processes which determine dependent clustering structures in the time series. A hierarchical specification of the DPY base measure is followed to account for various degrees of information pooling across the series. Some theoretical properties of the DPY are discussed and they are used to define Bayesian non-parametric repeated measurement and vector autoregressive models. Efficient Monte Carlo Markov Chain algorithms are provided for posterior computation of the proposed models and the effectiveness of the method is illustrated with a simulation study and an application to the United States and the European Union business cycles.

C573: **Bayesian semiparametric vector autoregressive models**
Presenter: **Maria Kalli**, Canterbury Christ Church University, United Kingdom
Co-authors: Jim Griffin
Vector autoregressive models (VARs) are the working horse for much macroeconomic forecasting. The two key reasons for their popularity are: their ability to describe the dynamic structure of many variables, and their ability to conditionally forecast potential future paths of specified subsets of those variables. Whether a classical or a Bayesian approach is adopted in VAR modelling most (if not all) models are linear with normally generated innovations. A Bayesian semiparametric approach is proposed which uses the Dirichlet process mixture to construct non-linear first order stationary multivariate VAR processes with non-Gaussian innovations. The method is applied to macroeconomic time series.

C717: **A flexible time-varying mixture model for economic time series**
Presenter: **Markus Jochmann**, Newcastle University, United Kingdom
Co-authors: John M. Maheu
Bayesian nonparametric mixture models based on the Dirichlet process prior have recently become more common in econometric time series applications. However, in most cases the prior is assumed to be constant over time. This is due to the fact that extensions of the Dirichlet process that allow for time-varying behavior are complex and, thus, hard to apply in practice. An alternative setup is discussed that is easier to implement but can still account for a time-varying mixture. It uses a recently introduced random distribution function that is based on a decreasing set of weights. We illustrate the proposed model framework with an application using macroeconomic data.

CS55 Room Torrington LONG MEMORY IN ECONOMIC AND FINANCIAL TIME SERIES Chair: Tommaso Proietti

C839: **Level shifts and long memory: A state space approach**
Presenter: **Stefano Grassi**, Aarhus University, Italy
Co-authors: Davide Delle Monache, Paolo Santucci de Magistris
It is well known that fractionally integrated processes are hard to distinguish from short-memory models when they are subject to structural breaks, e.g. a level shift. A robust estimation strategy is proposed to disentangle the level shift from the long-memory process. In particular, we use a state space model approach that provides corrected estimates of the long memory parameter also in the presence of regime changes. In addition, we estimate the probability and the size of the random level shifts and we test their joint nullity for the presence of level shift. Furthermore, we are able to track the level shifts and identify the unknown shifting dates. A set of Monte Carlo simulations shows that the proposed strategy produces unbiased estimates of the memory parameter in finite samples also when shifts in the mean, or other slowly varying trends, are present in the data. Our robust version of the KPSS test for the presence of level shift has proper size and by far the highest power compared to the other existing tests. In the empirical application we consider a set of realized volatility series of US stocks and it emerges that they are likely to be characterized by both long-memory and level shifts.

C919: **A bootstrap approximation for the distribution of the local whittle estimator**
Presenter: **Josu Arteche**, University of the Basque Country, Spain
Co-authors: Orbe Jesus
The Local Whittle estimator of the memory parameter d is characterized by a pivotal Gaussian asymptotic distribution, which makes inference very simple to implement. However, the Gaussian approximation may not be reliable in several cases, as for example with small sample sizes or even with larger samples when $d > 0.75$. In other situations the asymptotic distribution is unknown, as for example in a noninvertible context or with some nonlinear transformations of long memory processes. For all these cases a local bootstrap strategy based on resampling a standardized periodogram is proposed. It is shown in a Monte Carlo study that this strategy leads to a good approximation of the distribution of the Local Whittle estimator in those situations where the asymptotic Gaussianity is not reliable.

C1236: **Quantile regression for long memory testing: A case of realized volatility**
Presenter: **Uwe Hassler**, Goethe University Frankfurt, Germany
Co-authors: Paulo Rodrigues, Antonio Rubia
A quantile regression approach is derived to formally test for long memory in time series. We propose both individual and joint quantile tests which are useful to determine the order of integration along the different percentiles of the conditional distribution and, therefore, allow us to address more robustly the overall hypothesis of fractional integration. The null distributions of these tests obey standard laws (e.g., standard normal) and are free of nuisance parameters. The finite sample validity of the approach is established through Monte Carlo simulations, showing, for instance, large power gains over several alternative procedures under non-Gaussian errors. An empirical application of the testing procedure on different measures of daily realized volatility is presented. Our analysis reveals several interesting features, but the main finding is that the suitability of a long-memory model with a constant order of integration around 0.4 cannot be rejected along the different percentiles of the distribution, which provides strong support to the existence of long memory in realized volatility from a completely new perspective.

C939: **On the identification of fractionally cointegrated VAR models with the F(d) condition**
Presenter: **Federico Carlini**, Aarhus University, Denmark
Co-authors: Paolo Santucci de Magistris
An identification problem in the fractionally cointegrated system is discussed. The identification problem arises when the lag structure is over-specified, such that there exist several equivalent parametrizations of the model corresponding to different degrees of fractional integration and cointegration. We study the nature of these multiple solutions and we provide a sufficient condition that allows to correctly identify the fractional order of the system. We name this the F(d) condition. The assessment of the F(d) condition in the empirical analysis is relevant for the determination of the fractional orders in fractional VAR models.

CS62 Room Woburn RECENT DEVELOPMENTS IN SEASONAL ADJUSTMENT I Chair: Rosa Ruggeri-Cannata

C182: **Harmonizing the direct with the indirect approach in seasonal adjustment**
Presenter: **Marcus Scheiblecker**, WIFO, Austria
While the indirect approach compiles a seasonally adjusted series of an aggregate by summing up its seasonally adjusted components, the direct approach adjusts directly the aggregate. As the latter method ignores the possible restrictions imposed by its components, differences between the directly adjusted series and the sum of its adjusted components may occur. In order to identify one consistent seasonal and calendar adjusted aggregate, a procedure based on several criteria is proposed here. Apart from the theoretical consistency of the time series structures, the criterion of the consistency of calendar and deterministic effects and the one of seasonality are used. For chain-linked time series which do not add up even before adjustment, an approach for deriving a consistent seasonal and calendar adjusted aggregate is proposed as well. As an example the new approaches are demonstrated with European GDP data.

C271: **Choosing an appropriate seasonal adjustment approach: A data-driven way**
Presenter: **Karsten Webel**, Deutsche Bundesbank, Germany
When choosing a software package for conducting seasonal adjustment as part of their daily routines, many statistical agencies base their decisions on pragmatic reasons, such as employees' individual backgrounds, data users' demands (e.g. for low revisions) and the program's suitability for statistical mass production. Then, usually all time series, or at least broad subsets thereof, are seasonally adjusted according to the approach implemented in the software package chosen. Recent releases of X-13ARIMA-SEATS and Demetra+ may change habits as these programs include both the nonparametric X-11 approach and the parametric ARIMA model-based approach. Users may thus select between both approaches for each particular time series under review. Accordingly, the question immediately arises which criteria one should rely on when making this choice. We suggest a decision tree that combines theoretical considerations regarding differences between both approaches with empirical findings. In particular, we compare squared gains of final seasonal adjustment filters and diverse revision measures. Using German turnover data, we also demonstrate that running a seasonal adjustment algorithm with default options may lead to results that are somewhat misleading.

C384: **New features of JDEMETRA+**
Presenter: **Dario Buono**, Eurostat, Luxembourg
Co-authors: Enrico Infante, Sylwia Grudkowska
JDemetra+ is the Eurostat tool for seasonal adjustment enabling the compliance with the revised "ESS Guidelines on SA". It was developed to provide reusable and extensible components, compatible with the environment of NSOs. JDemetra+ is a user-friendly graphical interface built on open source Java libraries usable to deal with additional time series related issues like temporal disaggregation and benchmarking. It includes the most widespread seasonal adjustment methods: X-13-ARIMA and TRAMO/SEATS and structural models. The original routines were rewritten using NetBeans application. The tool enables the users to add extensions installable as plug-ins, with no impact on the core engine. JECOTRIM is an example of such plug-in that was developed by Eurostat to perform benchmarking compliant with the Eurostat "Handbook on quarterly national accounts". The aim is to present a practical case study on using both tools to obtain seasonally adjusted and benchmarked EU aggregates.

C617: **Variance estimates for seasonally adjusted unemployment level estimators for Wales and the UK**
Presenter: **Mark Hogan**, Office for National Statistics, United Kingdom
National Statistical Organisations routinely publish variance estimates for non-seasonally adjusted series to aid user interpretation, but not for seasonally adjusted series. Seasonally adjusted series are of prime interest to users and several methods for calculating the variance estimates are available. The need for variance estimates for seasonally adjusted series has been highlighted by researchers and governments as the lack of variance estimates frustrates the user in deciding how accurate the estimates are and thus how much importance to give them when using the series to make decisions. The leading methods for calculating variance estimates for seasonally adjusted series are briefly reviewed before being applied to unemployment level estimates for Wales and the UK taken from the Labour Force Survey of the Office for National Statistics. A suggestion is made as to which method could best be used to provide the variance estimates for the above series in a production setting.

CS65 Room Holden PARTIAL INFORMATION AND MODEL CHOICE IN PORTFOLIO OPTIMIZATION Chair: Joern Sass

C728: **Estimation risk in portfolio optimization**
Presenter: **Luitgard Veraart**, London School of Economics, United Kingdom
Co-authors: Mathieu Dubois
The effect of parameter uncertainty in a high-dimensional portfolio optimization problem is investigated. In particular, we focus on the relationship between the investor's risk aversion and estimation risk. We show that the loss in expected utility is large when simple plug-in strategies are used. To reduce the effect of estimation we use an L_1-norm constraint which can lead to sparse portfolios. We show that the sparsity of the constrained portfolio depends on the investor's coefficient of relative risk aversion. Based on a simulation study we discuss how a suitable L_1-norm constraint can be chosen.

C311: **American option pricing and filtering with a hidden regime-switching jump diffusion model**
Presenter: **Tak Kuen Siu**, Cass Business School, United Kingdom
The valuation of an American-style contingent claim is discussed in a hidden Markov regime-switching jump-diffusion market, where the evolution of a hidden economic state process over time is described by a continuous-time, finite-state, hidden Markov chain. A two-stage procedure is introduced to discuss the American option valuation problem. Firstly, filtering theory is employed to transform the market with hidden quantities into a filtered market with complete observations. Then the valuation problem is discussed using a generalized version of the Esscher transform in the filtered market. A probabilistic approach to the American option pricing is considered, where a decomposition formula for the price of an American put option is decomposed as the sum of its European counterpart and an early exercise premium.

C488: **Asset allocation in a regime-switching model with a robustified EM-algorithm**
Presenter: **Christina Erlwein-Sayer**, Fraunhofer Institute for Industrial Mathematics ITWM, Germany
Co-authors: Peter Ruckdeschel
The proposed asset allocation model allows for regime shifts in parameters of asset returns. Changing market conditions are therefore captured through switching volatilities and drifts of the asset price process. In this HMM framework, parameters of the model are guided by a Markov chain in discrete time. The parameters are estimated through an on-line algorithm, which utilizes incoming information from the market and leads to adaptive optimal estimates. A robust version of this filter-based EM-algorithm is established, thus, outliers, peaks or missings from historical asset prices have a limited effect on forecasts of the distribution of asset returns. Investment strategies are developed taking into account these robustified estimates. The obtained portfolios realistically handle regime-shifts in markets as well as additive outliers in historical data.

C158: **Continuous-time regime switching, discretization and portfolio optimization**
Presenter: **Joern Sass**, University of Kaiserslautern, Germany

In a continuous-time Markov switching model (MSM) the observation process is a diffusion whose drift and volatility coefficients jump governed by a continuous-time Markov chain. Due to the switching volatility, in continuous time the underlying Markov chain could be observed and no filtering is needed, while for the discretely observed model we would have to filter. Therefore, results in portfolio optimization in the continuous-time MSM often provide a poor approximation for the discretely observed model in which explicit optimization results are difficult to obtain. On the other hand, for a continuous-time hidden Markov model (HMM), where only the drift jumps and the volatility is constant, the discretized model provides a more consistent approximation for the continuous-time model. The HMM allows for explicit calculations, but does not have such good econometric properties. We discuss observability, filtering and portfolio optimization in both models. To combine useful aspects of both models, we then look at a HMM where the volatility is a function of an observable process, e.g. the filter. We analyze its relation to MSMs and show that we can get quite explicit filtering and optimization results which provide a good approximation to the discretely observed model.

CS78 Room Athlone REALIZED CORRELATIONS Chair: David Veredas

C058: **Econometric analysis of multivariate realised QML: Efficient positive semi-definite estimators of the covariation of equity prices**
Presenter: **Dacheng Xiu**, University of Chicago, United States
Co-authors: Neil Shephard

Estimating the covariance and correlation between assets using high frequency data is challenging due to market microstructure effects and Epps effects. We extend Xiu's univariate QML approach to the multivariate case, carrying out inference as if the observations arise from an asynchronously observed vector scaled Brownian model observed with error. Under stochastic volatility the resulting QML estimator is positive semi-definite, uses all available data, is consistent and asymptotically mixed normal. The quasi-likelihood is computed using a Kalman filter and optimised using a relatively simple EM algorithm which scales well with the number of assets. We derive the theoretical properties of the estimator and prove that it achieves the efficient rate of convergence. We show how to make it achieve the non-parametric efficiency bound for this problem. The estimator is also analysed using Monte Carlo methods and applied on equity data that are distinct in their levels of liquidity.

C356: **Disentangled jump-robust realized covariances and correlations with non-synchronous prices**
Presenter: **Harry Vander Elst**, Universite libre de Bruxelles, Belgium
Co-authors: David Veredas

A new jump-robust method to estimate ex-post quadratic covariation of Brownian semimartingales with finite activity jumps using high frequency data is introduced. The approach is based on disentangling correlations and volatilities to measure them at their optimal sampling frequencies. Quantile-and-median-based realized measure and a Hayashi-Yoshida cumulative scheme for asynchronicity are used. Consistency and asymptotic theory are provided in absence of microstructure noise. The goodness of the methodology is illustrated in a thorough Monte Carlo study and an application on real high-frequency data is provided.

C759: **Long term component dynamic modeld for realized covariance matrices**
Presenter: **Giuseppe Storti**, University of Salerno, Italy
Co-authors: Luc Bauwens, Manuela Braione

Dynamic models for realized covariance matrices are proposed, which include a secular component that captures the changing levels of realized variances and correlations. They generalize a previous realized DCC models where the long term level is assumed to be constant. The long term component is specified either as a nonparametric function or as a MIDAS term. Estimation can be done in steps for large dimensional matrices.

C865: **Central limit theorem for integrated covariation in the presence of infinite variation Levy jumps**
Presenter: **Cecilia Mancini**, University of Florence, Italy

Two processes driven by Brownian motions plus drift and jumps with possibly infinite activity and infinite variation are considered. Given discrete observations it is possible to separately estimate the integrated covariation IC between the two Brownian parts and the sum of the co-jumps by using a threshold principle (truncated realized covariance) allowing us to isolate the jumps over a given threshold. We establish here the speed of convergence of the estimator of IC when the jump component of the bivariate process is Levy. We find that such a speed is heavily influenced by the small jumps dependence structure other than by their jump activity indices. Previous results are extended, where the asymptotic normality of the estimator was obtained when the jump components have finite activity or finite variation. This result gives insight into the dependence structure of the processes and has important applications in finance.

CS90 Room Court FINANCIAL RISK MODELING AND MANAGEMENT Chair: Jiri Witzany

C096: **Estimating default and recovery rate correlations**
Presenter: **Jiri Witzany**, University of Economics in Prague, Czech Republic

A two-factor credit risk model is analyzed which captures default and recovery rate variation, their mutual correlation, and dependence on various explanatory variables. At the same time, it allows the analytical computation of the unexpected credit loss. We propose and empirically implement an estimation of the model based on aggregate and exposure level Moody's default and recovery data. The results confirm the existence of significantly positive default and recovery rate correlation. We empirically compare the unexpected loss estimates based on the reduced two-factor model with Monte Carlo simulation results, and with the current regulatory formula outputs. The results show a very good performance of the proposed analytical formula which could feasibly replace the current regulatory formula.

C101: **Pricing of options on defaultable bonds using a binomial tree**
Presenter: **Marek Kolman**, University of Economics Prague, Czech Republic

A flexible tree model for pricing of options on (coupon-bearing) defaultable bonds is proposed. The model comprises two separate building blocks. Firstly, we use a tree model Black-Derman-Toy for default-free term-structure of interest rates. Secondly, we model a default/survival process. These two blocks combined result into a two-layer tree representing survival and default states of the bond. An option pricing framework follows on these two layers. Eventually, we also show how a barrier option on defaultable claim can be priced within this framework using a decomposition of the tree into separate states-of-the-world and then these are priced one by one. A simple proof based on verification of the parity between barrier and vanilla plain options concludes the research effort.

C359: **Interest rate swap credit value adjustment**
Presenter: **Jakub Cerny**, Charles University in Prague, Czech Republic
Co-authors: Jiri Witzany

The credit value adjustment (CVA) of OTC derivatives is an important part of the Basel III credit risk capital requirements and current accounting rules. Its calculation is not an easy task - not only it is necessary to model the future value of the derivative, but also the probability of the default of the counterparty. Another complication arises in the calculation incorporating the wrong-way risk, i.e. the negative dependence between the

underlying asset and the default time. A semi-analytical CVA formula simplifying the interest rate swap (IRS) valuation with counterparty credit risk (CCR) including the wrong-way risk is derived and analyzed. The formula is based on the fact that the CVA of an IRS can be expressed by the swaption price. The link between the interest rates and the default time is represented by a Gaussian copula with constant correlation coefficient. Finally, the results of the semi-analytical approach are compared with results of a complex simulation study.

C597: **On discrimination indices for survival models**
Presenter: **Chiara Gigliarano**, Department of Economics and Social Sciences, Italy
Co-authors: Silvia Figini, Pietro Muliere
The Harrell's C discrimination index is an extension of the area under the ROC curve to the case of censored survival data. It has been widely used to assess and compare prediction models with respect to their ability to discriminate individual risks. The main objective is to propose a novel approach for model comparisons and performance evaluation. In particular, a new class of performance measures is defined that captures time trends for models that allow crossing of survival curves. The theoretical proposal is then tested on a real financial data set.

C313: **Risk estimation and assessment using survival models for credit risk data**
Presenter: **Silvia Figini**, University of Pavia, Italy
Co-authors: Filomena Madormo
Credit risk models are used to evaluate the insolvency risk caused by credits that enter into default. Many models for credit risk have been developed over the past few decades. We focus on those models that can be formulated in terms of the probability of default by using survival analysis techniques. In order to write the default probability we use the Cox regression model and Random Survival Forest. We compare the results of the survival models with respect to classical models, such as generalised linear models based on logistic regression and non-parametric techniques based on classification trees in terms of cross validation. An empirical study, based on real data, illustrates the performance of each model.

CS102 Room Gordon FINANCIAL ECONOMETRICS I Chair: Engelbert Dockner

C365: **Refined estimation of parametric models by functional approximation**
Presenter: **Gareth Liu-Evans**, University of Liverpool, United Kingdom
An alternative refinement method is proposed for estimation of parametric models, particularly diffusions. The method is simulation based and results in closed-form refinements to the existing estimator. It performs well in simulations. It is compared with indirect inference and the bootstrap for estimation of the mean reversion parameter in CIR interest rate models. Unrefined ML-type estimators are known to perform poorly for this parameter when the level of mean reversion is low.

C923: **The impact of the financial crisis on transatlantic information flows: An intraday analysis**
Presenter: **Thomas Dimpfl**, University of Tubingen, Germany
Co-authors: Franziska J. Peter
Intraday stock index data from both sides of the Atlantic during overlapping trading hours are used to analyze the dynamic interactions between European and US stock markets. We are particularly interested in differences of information transmission before, during, and after the financial crisis of 2007 to 2009. Our analysis draws on the concept of Rényi transfer entropy to allow for a flexible and model-free empirical assessment of linear as well as non-linear market dependencies. Thereby the importance of extreme (tail) observations of the return distributions is highlighted. The results show significant bi-directional information transfer between the US and the European markets with a dominant flow from the US market. During the crisis dynamic interactions increase. At the same time the European markets' contribution to price discovery rises. The US market does not entirely regain its leading role in the after crisis period.

C1161: **Basel III and the prediction of financial crises**
Presenter: **Simon van Norden**, HEC Montreal, Canada
Co-authors: Marc Wildi
A key new feature of financial regulation under Basel III is the Counter-Cyclical Capital Buffer, which requires regulators to adjust capital requirements well in advance of periods of financial stress. We investigate the extent to which regulators can anticipate such periods using the credit aggregates suggested by BIS and IMF studies. We present new evidence (a) on the trade-off regulators can hope to achieve between correct prediction of financial crises and costly false alarms using ROC curves and related inference methods, and (b) the benefits of using Direct Filter Analysis to optimize filter design in terms of both noise suppression and speed in signal detection. We show that a broad range of filter designs provide an improvement over existing aggregates and provide statistically significant predictive power for banking crises.

C807: **Errors in higher order risk neutral moments estimation**
Presenter: **Guillaume Bagnarosa**, University Paris Pantheon Sorbonne, United Kingdom
While the risk neutral density has always aroused interest in the academic literature, the developments over the last decade in nonparametric extraction of the associated risk neutral higher order moments have been at the origin of very promising applications in areas such as portfolio allocation, realized moments forecasting or risk aversion measurements. Although the extraction of risk neutral higher-order moments does not present technical difficulties any more, the quality of the market data commonly used and the high sensitivity to noise of the risk neutral measures render the task more perilous than it looks like. The risk neutral moments being extracted from cross-sectional options prices at a given time, the microstructure phenomena affecting options data on a high frequency scale affect accordingly the end of the day options prices commonly used in the literature. The first contribution is to provide an exhaustive and innovative nomenclature of the sources of microstructure noises specifically affecting options high frequency data. The second contribution is to provide nonparametric confidence intervals for the inferred risk neutral density used for the valuation of any derivative asset. Eventually, the consecutive impact and uncertainty over the nonparametrically estimated implied risk neutral moments is measured and discussed.

C831: **A non-parametric test for dependence based on the entropy rate**
Presenter: **Galen Sher**, University of Oxford, United Kingdom
Co-authors: Pedro Vitoria
A non-parametric test for dependence between sets of random variables based on the entropy rate is proposed. The test has correct size, unit asymptotic power, and can be applied to test setwise cross sectional and serial dependence. Using Monte Carlo experiments, we show that the test has favourable small-sample properties when compared to other tests for dependence. The 'trick' of the test relies on using universal codes to estimate the entropy rate of the stochastic process generating the data, and simulating the null distribution of the estimator through subsampling. This approach avoids having to estimate joint densities and therefore allows for large classes of dependence relationships to be tested. Potential economic applications include model specification, variable and lag selection, data mining, goodness- of-fit testing and measuring predictability.

CS107 Room Bloomsbury FINANCIAL APPLICATIONS I Chair: Joao Victor Issler

C180: **Optimal VWAP tracking**
Presenter: **Jedrzej Bialkowski**, University of Canterbury, New Zealand
Co-authors: Daniel Mitchell, Stathis Tompaidis
The problem of finding a strategy that tracks the volume weighted average price (VWAP) of a stock, a key measure of execution quality for large orders used by institutional investors, is considered. We obtain the optimal, dynamic, VWAP tracking strategy in closed form in a model without trading costs with general price and volume dynamics. We build a model of intraday volume using the Trade and Quote dataset to empirically test the strategy, both without trading costs and when trading has temporary and permanent effects, and find that the implementation cost is lower than the cost charged by brokerage houses.

C229: **IPO and freeze-out pricing: An empirical analysis of the French market**
Presenter: **Hanane Dakhli**, Sorbonne, France
The valuation and the pricing of Initial Public Offering (IPO) and Freeze-out by underwriters are investigated for a sample of 145 IPOs and 169 Freeze-outs on the French Stock Market between 2000 and 2011. Underwriters generally use several valuation methods and assign weights to the estimated values in order to assess the fair value. We study three families of valuation methods: direct valuation, comparable or indirect valuation and patrimonial valuation methods. Firstly, the aim is to explain the choice of underwriters for a specific valuation method according to the characteristics of the firm, the expected market return and its volatility during the period before each offer and the reputation of the underwriter. Secondly, we analyze the weightings allocated to the various estimated values according the same variables. Finally, we try to specify the degree of accuracy of the estimated values by each valuation method comparably to the IPO or Freeze-out prices. Our results then allow us to formulate several main recommendations on the choices of valuation methods and weightings of the estimated values.

C915: **On strong law of large numbers for approximation of LU-fuzzy numbers**
Presenter: **Tomas Tichy**, VSB-TU Ostrava, Czech Republic
Co-authors: Michal Holcapek
Financial problems can be analyzed by several ways. One of them is based on simulation of scenarios that might happen. In the standard case, we assume that a given financial quantity in question is a stochastic (or random) variable, ie. it follows some probability distribution and its possible states can get prescribed particular probabilities. The analysis is basically related to the Law of Large Numbers, which allows one to show that as the number of scenarios approaches to infinity, the modeled results should approaches to the reality. However, in financial modeling there appear two problems. The first one is the complexity of stochastic models that causes impossibility to solve them using conventional methods (stochastic differential equations). The second problem concerns the estimation of parameters of such models (e.g. volatility). One can recognize an unnatural simplification of parameters that can lead to a loss of important information hidden in data. If one wants to apply Monte Carlo simulation to analyze a financial problem with values expressed by imprecisely defined numbers, it is important to show that random variables with imprecisely defined numbers satisfy the strong law of large numbers, as well. Otherwise such approach would have no sense. We analyze the imprecisely defined numbers using parameterized LU-fuzzy numbers that well approximate LU-fuzzy numbers and analyze the convergence of the results in comparison with other approaches using financial data. We show that the almost sure convergence is preserved by splines.

C973: **Volatility persistence: On the use of structural breaks**
Presenter: **Yuhan Zhang**, University of Leicester, United Kingdom
Co-authors: Alexandra Dias
The existence and use of structural breaks in volatility estimation is still an open question. We explore the effect of structural breaks on the estimation of asset volatility in the case of three stock indices: S&P 500 (US), FTSE 100 (UK), and SSE Composite Index A-shares (China). Using a GARCH model with dummy variables, we compare the volatility persistence forecasts with and without incorporating structural breaks. The study reveals the existence of structural breaks in the volatility persistence in both developed and developing markets. There is evidence of approximate simultaneity on breaks in the volatility especially between UK and US since 2007, revealing a growing market integration of UK and US. In the end of 2008 China has a structural break following UK and US, showing the global impact of the subprime crisis. The volatility estimates do not change significantly when considering the existence of structural breaks. We complement our study with a simulation study where we investigate the robustness of the methods used here for incorporating breaks on the estimation of volatility.

C1230: **Estimation and calculation procedures of the provisions for outstanding insurance claims**
Presenter: **Andrejs Matvejevs**, Riga Technical university, Latvia
Co-authors: Oksana Pavlenko, Normunds Gutmanis
The purpose is to present algorithms for insurance technical provisions taking into account losses which are incurred but not reported. Rating an excess of loss treaty for long-tail business such as motor liability insurance or general third party liability is an almost impossible task, since it involves estimating the loss amount that will be awarded by a court in 10 to 20 years' time. Typically major losses in motor third party liability insurance are personal injury claims that take a long time to settle. Often, life-long treatment and nursing costs have to be paid. To do so, it is necessary to know how the economic environment will look at that time. Despite these difficult conditions, however, one can take up the challenge, at every renewal, of estimating future loss amounts on the basis of past data or current risk profiles. Some procedures are presented for the most accurate forecast possible for the expected excess of loss amount for a treaty year, using the loss experience of previous years of occurrence with their development. It is described how the past premiums and outstanding losses have to be projected to the renewal year so that they correspond to the current values.

CS05 Room Bedford CONTRIBUTIONS TO HIGH-FREQUENCY VOLATILITY FORECASTING Chair: Timo Terasvirta

C093: **A GARCH analysis of dark-pool trades**
Presenter: **Oren Tapiero**, Bar - Ilan, Israel
Co-authors: Philippe de Peretti
The ability to trade in dark-pools without publicly announcing trading orders, concerns regulators and market participants alike. The information contribution of dark trades to the intraday volatility process is analyzed. The analysis is conducted by performing a GARCH estimation framework where errors follow the generalized error distribution (GED) and two different proxies for dark trading activity are separately included in the volatility equation. Results indicate that dark trades convey important information on the intraday volatility process. Furthermore, the results highlight the superiority of the proportion of dark trades relative to the proportion of dark volume in affecting the one-step-ahead density forecast.

C830: **The out-of-sample significance of macroeconomic announcements, linearity, long memory, heterogeneity and jumps**
Presenter: **Dimitrios Vortelinos**, University of Lincoln, United Kingdom
The purpose is to examine the significance of macroeconomic announcements, linearity, long memory, heterogeneity and jumps to realized volatility forecasting, in three U.S. mini-futures markets: equity index, energy, and commodities. The research questions answered, are: Do linear, long

memory, heterogeneous or jumps models forecast volatility more accurately than the benchmark AR linear model? Which linear, long memory, heterogeneous or jumps model performs best in each corresponding category? Do linear, long memory, heterogeneous or jumps models forecast volatility better in announcement days than in days without announcements? Which linear model more accurately forecasts realized volatility either in the entire sample or in days with announcements? Is it better intraday volatility forecasts to take place on out-of-sample days with macroeconomic announcements rather than on all the out-of-sample days?

C1166: **An empirical evaluation of pseudo-long-memory time-series models to predict the S&P 500 index-futures realized volatility**
Presenter: **Benoit Sevi**, Aix-Marseille School of Economics, France

The forecasting accuracy of time-series models based on high-frequency data to predict future volatility is now widely recognized. Among these models, those which take account of the long-memory behavior seem to deliver the best forecasts of conditional volatility. However, for these models, numerous specifications have been suggested which (1) consider the past realized volatility at various horizons, (2) separately consider the jump component, (3) consider the realized negative/positive semivariance, (4) consider signed jumps and (5) consider possible leverage effect. We evaluate the in and out-of-sample performance of 12 different models using statistical criteria that are robust to the noisy feature of realized volatility. Our results indicate that considering these components does not statistically improve the out-of-sample performance. Importantly, the standard HAR model has a good forecast accuracy and is not outperformed by more sophisticated models. This finding is robust to the forecast horizon under consideration and the use of a variety of loss functions. Overall, our results seriously call into question the principle of disentangling the various components of the realized volatility to enhance forecast accuracy.

C1167: **Dynamic HAR model with jumps and leverage effect**
Presenter: **Katja Gisler**, University of St Gallen, Switzerland
Co-authors: Daniel Buncic

The Heterogeneous Autoregressive (HAR) model has become the benchmark model for forecasting realized volatility. We use a flexible dynamic modelling approach allowing for time varying parameters as well as time varying model sizes to forecast realized volatility of the S&P 500 futures. We add leverage and jumps to the standard HAR model as predictors, and also allow for 'day of the week' effects by including dummy variables in the predictor set. Using standard forecast evaluation tests, we show that the flexible dynamic model significantly outperforms the standard HAR model and also the leverage HAR with continuous volatility and jumps (LHAR-CJ) model in predicting realized volatility of the S&P 500 futures at forecast horizons ranging from 1 to 22 days ahead. We also illustrate the economic value of the forecast gains with a simple option pricing application.

C1218: **Asymmetries in high frequency foreign exchange markets: On when the leverage effects kick in**
Presenter: **Felix Matthys**, University of Zurich, Switzerland

Standard Threshold GARCH models divide the distribution of innovations into two subintervals at a threshold level set at zero such that the influence of positive and negative innovations exhibit differential effects on volatility. We introduce a novel flexible Generalized Leverage GARCH model (GL-GARCH) which treats the threshold level, after which innovations have asymmetric effects on volatility, as endogenous. We derive stationarity conditions, autocorrelation functions and moments up to fourth order of the GL-GARCH process and discuss their dependence upon the threshold level in detail. In an extensive finite sample Monte Carlo simulation study we investigate the performance of the maximum likelihood estimator and provide an efficient numerical scheme to estimate the model. As an empirical application, we estimate the GL-GARCH model using high frequency foreign exchange data and show that, first the threshold level is not equal to zero but tends to be negative for higher frequently sampled returns and second that the leverage effect coefficient is substantially larger as compared to the classical GJR-GARCH model. Finally, we conduct a forecast comparison between the GL-GARCH and standard asymmetric GARCH models and show that the GL-GARCH is in general able to outperform its competitors.

CS76 Room Deller MODELLING THE TERM STRUCTURE OF INTEREST RATES Chair: Rochelle Edge

C903: **Modeling the term structure of interest rates and the Eurozone: Some economic implications from principal components analysis**
Presenter: **Januj Juneja**, San Diego State University, United States

One approach for extracting factors from several yield curves, called pooled principal component analysis (PPCA), involves pooling the data from multiple groups. It assumes that each group has identical variability which leads to the group with the highest variability determining common factor structure—an undesirable feature if the groups do not have identical variability. A multivariate extension to principal components analysis (PCA), known as common principal components analysis (CPCA) enables several groups to have different variability in the determination of common factor structure. However, both methods have received less attention because extracted factors lack economic meaning and identification. Utilizing recent advances, we overcome this drawback by designing and implementing a simulation algorithm to identify and compare common factors obtained from CPCA and PPCA using correlations. We also assess economic implications of their statistical differences using a term structure model employing common factors extracted from PPCA and CPCA as the state vector to model the interest rate dynamics of several countries in the Eurozone. The models are estimated using monthly data following the institutionalization of the Eurozone. We focus on the Eurozone because it has received quite a bit of publicity questioning its stability as a group.

C992: **Nonparametric estimation of yield curves with L-spline**
Presenter: **Yoshinori Kawasaki**, The Institute of Statistical Mathematics, Japan
Co-authors: Yasunori Horikoshi

A method to estimate yield curves using L-spline is proposed. In the context of L-spline smoothing, a roughness penalty is constructed by a linear differential equation which the true function should satisfy approximately. We derive a smoothing constraint from the linear differential operators together with the boundary conditions that the Nelson-Siegel model obeys, and use it for nonparametric estimation of yield curves based on coupon bonds. Also derived are the inner product and the corresponding reproducing kernel for our purpose. Worked examples with traded coupon bond data will be given.

C1122: **Economic relevance of hidden factors in international bond risk premia**
Presenter: **Luca Tiozzo Pezzoli**, Banque de France - Paris Dauphine University, France

The relevance of hidden factors is investigated in international bond risk premia to forecast future excess bond returns and macroeconomic variables such as economic growth and inflation. Using maximum likelihood estimation for a linear Gaussian state-space model, adopted to explain the dynamics of expected excess bond returns of a given country, we detect as relevant factors otherwise judged negligible by classical explained variance approach. We call these factors hidden, meaning that they are not visible through the lens of a principal component analysis of expected excess bond returns. We find that these hidden factors are useful predictors of both future economic growth and inflation given that they add forecasting power over and above the information contained in curve factors. These empirical findings are robust across different sample periods and countries and with respect to the interpolation technique used in the construction of the international bond yields data sets.

C1135: **Estimating a term-structure model of commodity prices with heteroskedastic measurement error**

Presenter: **Hiroaki Suenaga**, Curtin University, Australia

Term structure of commodity futures prices has been studied extensively in the literature where the standard approach is to specify the stochastic dynamics of underlying factors and derive valuation formulas of derivative contracts as functions of the underlying factor values. The parameters in these term-structure models are usually estimated with panel data comprised from multiple contract prices observed per day with differing maturity dates. The estimation requires measurement errors to be added to the model to allow for discrepancies between the models' implied and observed prices. These measurement errors are commonly assumed as independently and identically distributed (iid). However, a recent study shows that the iid assumption rarely holds and allowing cross-sectional and serial correlation in measurement errors results in substantially different parameter estimates in the exponential affine model. We extend this line of research and examine how the specification of the variance structure of the measurement error affects parameter estimates. Using data from four commodity markets, we illustrate empirically that the variance of measurement error differs substantially across seasons and by time-to-maturity. Incorporating these patterns affects substantially the estimates of the other key parameters in the model.

Sunday 15.12.2013 10:55 - 12:10 Parallel Session J – ERCIM

ES10 Room B18 ADVANCES ON ROBUST ANALYSIS OF FUNCTIONAL DATA AND COMPLEX STRUCTURES Chair: Graciela Boente

E047: **Functional regression models with temporal and/or spatial dependence**
Presenter: **Manuel Febrero-Bande**, University of Santiago de Compostela, Spain
Co-authors: Manuel Oviedo de la Fuente

Functional regression models affected by temporal or spatial dependence are considered. We propose an iterative procedure that can be applied to linear and non-linear models and provides similar results as the classical gls procedure. This is shown with an extensive simulation study. The procedure is applied to real data with the goal of predicting the incidence rate of influenza.

E342: **High dimension low sample size asymptotics of robust PCA**
Presenter: **Yi-Hui Zhou**, North Carolina State University, United States
Co-authors: J.S. Marron

Conventional principal component analysis is highly susceptible to outliers. In particular, a sufficiently outlying single data point can draw the leading principal component toward itself. The effects of outliers for high dimension and low sample size data are studied using asymptotics. The non-robust nature of conventional principal component analysis is verified through inconsistency under multivariate Gaussian assumptions with a single spike in the covariance structure, in the presence of a contaminating outlier. In the same setting, the robust method of spherical principal components is consistent with the population eigenvector for the spike model, even in the presence of contamination.

E828: **Robust lasso variable selection for functional data**
Presenter: **Juan Romo**, Universidad Carlos III de Madrid, Spain
Co-authors: Nicola Mingotti, Rosa Lillo

Functional regression has been an active field of research for the last two decades but still lacks a secure variable selection methodology. Lasso is a well known effective technique for parameters shrinkage and variable selection in regression problems. We generalize the lasso technique to select variables in the functional regression framework and we show its good behaviour in this setting. We focus on functional regression with scalar regressors and functional response and reduce the associated functional optimization problem to a convex optimization on scalars, finding the solution and stressing its interpretability. We also present a robust version of this new lasso proposal. The practical use and the effectivennes of the proposed technique is demonstrated by applying it to simulated data sets as well as to a real data set of car velocity functions in low speed car accidents, a frequent cause of whiplash injuries.

ES12 Room B36 HIGH-DIMENSIONAL STATISTICS Chair: Thomas Kneib

E332: **Large-scale regression with sparse data**
Presenter: **Rajen Shah**, University of Cambridge, UK
Co-authors: Nicolai Meinshausen

The information age in which we are living has brought with it a combination of statistical and computational challenges that often must be met with approaches that draw on developments from both the fields of statistics and computer science. A method for performing regression will be presented where the n by p design matrix may have both n and p in the millions, but where the design matrix is sparse, i.e. most of its entries are zero; such sparsity is common in many large-scale applications such as text analysis. In this setting, performing regression using the original data can be computationally infeasible. Instead, we first map the design matrix to an $n \times L$ matrix with $L \ll p$, using a scheme based on a technique from computer science known as min-wise hashing. From a statistical perspective, we study the performance of regression using this compressed data, and give finite sample bounds on the prediction error. Interestingly, despite the loss of information through the compression scheme, we will see that least squares or ridge regression applied to the reduced data can actually allow us to fit a model containing interactions in the original data.

E742: **Two-sample testing in high-dimensional models**
Presenter: **Nicolas Stadler**, NKI, Netherlands
Co-authors: Sach Mukherjee

A novel methodology for testing equality of model parameters between two high-dimensional populations is proposed. The technique is very general and applicable to a wide range of models. The method is based on sample splitting: the data is split into two parts; on the first part we reduce the dimensionality of the model to a manageable size; on the second part we perform significance testing ($p-$value calculation) based on a restricted likelihood ratio statistic. Assuming that both populations arise from the same distribution, we show that the restricted likelihood ratio statistic is asymptotically distributed as a weighted sum of chi-squares with weights which can be efficiently estimated from the data. Arbitrary splitting of the data results in a "$p-$value lottery". In order to get reproducible results we repeat the splitting procedure various times and aggregate the resulting $p-$values. This multi-split approach provides improved $p-$values. We illustrate the use of our general two-sample approach in two-sample comparison of high-dimensional regression models (differential regression) and graphical models (differential network). In both cases we show results on simulated data as well as real data from recent, high-throughput cancer studies.

E995: **hdi: High-dimensional inference**
Presenter: **Ruben Dezeure**, ETH Zurich, Switzerland
Co-authors: Peter Buehlmann, Lukas Meier

The focus of a 'classical' data analysis is not only on point estimation, but more importantly on inference, i.e. on assigning uncertainty (or significance). For high-dimensional problems, most of the research only dealt with point estimation so far. Recently, some progress has been made with respect to assigning uncertainty to point estimates. We illustrate some generic methods, their properties and show how they can be easily applied to a broad range of real word problems using the R-package 'hdi'.

ES30 Room B20 ADVANCES IN DYNAMIC MODELLING AND COMPUTATIONAL INFERENCE Chair: Richard Gerlach

E185: **Sparse sieve MLE**
Presenter: **Artem Prokhorov**, Concordia U and U Sydney, Canada
Co-authors: Di Liu
The Dantzig selector is traditionally used for point estimation in least squares when the number of parameters exceeds the number of observations. We use it to get smaller standard errors in a sieve maximum likelihood estimation. We assume correctly specified likelihood-based models for the marginals and the Bernstein polynomial serves as the sieve copula capturing dependence between them. This leads to an estimator that dominates conventional QMLE, but the number of parameters in the sieve is close to and often exceeds the sample size in applications. We show in simulations that the Dantzig selector produces a sparse sieve MLE estimator with properties very similar to the non-sparse alternative. Thus, the sparsity employed by the Dantzig selector is innocuous with respect to the finite sample behavior of the sieve MLE; it also permits a substantial increase in computational efficiency compared to the unrestricted SMLE.

E441: **Bayesian model selection of regular vine copulas**
Presenter: **Lutz Gruber**, Munich University of Technology, Germany
Co-authors: Claudia Czado
Regular vine copulas can describe a wider array of dependency patterns than the multivariate Gaussian copula or the multivariate Student's t copula. The first fully Bayesian approach to model selection of regular vine copulas is presented. The reversible jump Markov chain Monte Carlo algorithm employs a mixture of different proposal distributions to achieve outstanding acceptance rates and rapid convergence of the sampling chain. Another major contribution to the area of model selection of regular vine copulas is the algorithm's capability to estimate all levels of vine copulas simultaneously. The feasibility of the approach is established in a simulation study that benchmarks against methods suggested in current literature. A case study shows how the pricing of financial derivatives with multiple underlying assets is improved when their dependence characteristics are modeled by a vine copula instead of Student's t copula.

E627: **Optimal insurance purchase strategies via multiple optimal stopping time rules**
Presenter: **Rodrigo Targino**, University College London, United Kingdom
Co-authors: Gareth Peters, Georgy Sofronov, Pavel Shevchenko
The aim is to study the class of multiple optimal decision rules for purchases of insurance linked security products and other insurance products. This topic has direct links to purchase strategies of insurance for Operational risk under Basel II/III as well as purchase strategies for catastrophe bonds. In particular the focus is on the families of Loss Distributional Models for flexible heavy tailed, skew-kurtosis severity models and generalized Poisson frequency models. Closed form solutions for the multiple stopping rules and value functions for some models are developed, as well as closed form Askey-orthogonal polynomial expansion solutions for more general insurance and LDA model combinations. In addition alternative recursive numerical procedures are discussed for evaluation of the optimal purchase decision rules.

ES129 Room B30 DIRECTIONAL STATISTICS II Chair: Thomas Verdebout

E835: **Central limit theorems for directional and linear data with applications**
Presenter: **Eduardo Garcia-Portugues**, Santiago de Compostela, Spain
Co-authors: Rosa M. Crujeiras, Wenceslao Gonzalez-Manteiga
A central limit theorem for the integrated squared error of the kernel density estimator for directional-linear data is given in this work. This result enables the design of testing procedures on the density, based on L2-distances: a goodness-of-fit test for parametric families of directional-linear densities and a completely nonparametric independence test. The limit distributions of the test statistics are derived, and a suitable bootstrap strategy is designed for its application in practice. Furthermore, the proposed methodology can be extended for handling two directional components. Simulation results are provided to illustrate the finite sample performance and the asymptotic behaviour of the tests. Illustrations in real data examples from forestry and proteomics will be also provided.

E092: **A new concept of quantiles for directional data**
Presenter: **Christophe Ley**, Universite libre de Bruxelles, Belgium
Co-authors: Camille Sabbah, Thomas Verdebout
A new concept of quantiles for directional -circular and (hyper-)spherical- data is introduced. Their definition, which also leads to a novel concept of depth for directional data, can serve as an interesting alternative to the more complicated angular simplicial and angular Tukey depth. We establish the typical quantile-asymptotics, namely a Bahadur-type representation and asymptotic normality. These results are corroborated by a Monte Carlo simulation study. We also introduce new user-friendly statistical tools such as a directional QQ-plot and a quantile-based goodness-of-fit test. We illustrate the power of our new procedures by analyzing a cosmic rays data set. We conclude by briefly discussing some further applications of our quantiles in trimming and classification issues.

E544: **The Le Cam methodology for directional data**
Presenter: **Yvik Swan**, University Luxembourg, Luxembourg
Co-authors: Thomas Verdebout, Christophe Ley
We describe how to use the Le Cam methodology for estimation and testing problems for directional data. In particular we discuss the local asymptotic normality property of a sequence of rotationally symmetric models; this is a non-standard result due to the curved nature of the unit sphere. Locally and asymptotically most stringent parametric tests for ANOVA are constructed for directional data within the class of rotationally symmetric distributions. Asymptotic relative efficiencies are calculated and the finite-sample behaviours of the proposed tests are investigated by means of a Monte Carlo simulation.

ES54 Room B34 GOODNESS-OF-FIT TESTS Chair: Simos Meintanis

E294: **Characterizations and goodness-of-fit tests for multivariate normal and Cauchy distributions**
Presenter: **Emanuele Taufer**, University of Trento, Italy
Co-authors: Simos Meintanis
A characterization for stable distributions, based on the form of the characteristic function, is proven and used to define new goodness-of-fit tests for the multivariate normal and Cauchy distribution. The new tests are shown to be consistent and closed-form formulas for their computations are provided. The small sample performance of the tests is analyzed by means of simulations.

E702: **Estimation and goodness-of-fit tests on stress test models for banks**
Presenter: **Savas Papadopoulos**, Democritus University of Thrace, Greece
Co-authors: George Papadopoulos
The stability of banking systems is evaluated by stress testing (ST). Capital ratios are estimated under unfavorable macroeconomic conditions for several groups of banks. We estimate several panel-data models for loan loss provision, profit, credit, deposit, and capital levels linked to GDP, inflation and unemployment, and select the one that performs better in out-of-sample predictions. The models are statistically satisfactory and predict with small bias and variability. Dynamic panel data models were estimated using several methods including a novel one. The new method is based on a proved theorem and has various advantages in comparison to standard methods, especially to small samples for N and to unbalanced data sets. The models are simulated to assess the bias and the variability of the estimates. The new method excels the standard methods in many cases and it can be applied when the other methods cannot be executed or cannot provide satisfactory results due to small N. In the application to ST, standard methods do not give the desirable estimates due to the highly unbalanced data set, while the new method performs satisfactorily.

E495: **Fourier methods for model selection**
Presenter: **M Dolores Jimenez-Gamero**, Universidad de Sevilla, Spain
Co-authors: Apostolos Batsidis, M. Virtudes Alba-Fernandez
A test approach to model selection is proposed. The scheme is close to that developed by Vuong, which is based on comparing the maximized likelihood of the data under the competing models, that estimates the Kullback-Leibler distance between the models and the true population. Other discrepancy measures could be used to measure the closeness between each competing model and the population. This is specially appealing in cases where the likelihood of a model cannot be calculated or even, if it has a closed expression, it is not easily tractable. Such closeness is measured by means of a distance based on the characteristic functions (cfs). As a prerequisite to the model selection problem based on cfs, some asymptotic properties of the minimum ISE estimators are studied. From these properties, consistent tests for model selection based on cfs are given for separate, overlapping and nested models.

ES71 Room B33 SPATIAL FUNCTIONAL DATA ANALYSIS Chair: Laura Sangalli

E526: **Population density estimation over complex domains using bivariate splines**
Presenter: **Serge Guillas**, University College London, United Kingdom
Co-authors: Ming-Jun Lai, Jean Gaudart, Bree Ettinger
The aim is to consider the estimation of a non-negative surface over a complex domain with an irregular boundary and interior holes. Bivariate splines over triangulations are employed to build the surface. The non-negativity condition can be satisfied by recasting the estimation of the spline's coefficients as a constrained penalized least squares problem over the spline basis. In addition, boundary conditions are enforced in accordance with the given application. In the absence of replicates of a given surface, the computation of spatial uncertainties in the estimation is addressed by bootstrapping the original surface. As a result, boxplots for the distribution of surfaces are computed and displayed that rely on functional data depth. An illustration to the estimation of population density of a town separated by a river highlights the skills of the method.

E648: **Spatial regression with PDE penalization**
Presenter: **Laura Azzimonti**, Politecnico di Milano, Italy
Co-authors: Laura M. Sangalli, Piercesare Secchi
The class of Spatial Regression models with PDE penalization (SR-PDE) is considered and it is extended to include the time dimension. These models have a penalized regression framework where the penalty term involves a differential operator, which describes the prior knowledge on the phenomenon under study coming from physics, physiology or mechanics of the problem at hand. In particular, phenomena that are well described by means of parabolic partial differential equations are considered, involving second-order derivatives in space and first order derivatives in time. SR-PDE allows for important modeling flexibility, accounting for space anisotropy and non-stationarity in space and time in a straightforward way. These models exploit advanced scientific computing techniques and specifically make use of the Finite Element method and of the Finite Difference method for the discretization of the estimation problem respectively in space and time. The motivating applied problem concerns the estimation of the time-dependent blood-flow velocity field in a section of the common carotid artery, using data provided by echo-color doppler acquisitions. This applied problem arises within the research project MACAREN@MOX that aims at studying atherosclerosis pathogenesis.

E1253: **A new depth measure for ordering spatially dependent functional data**
Presenter: **Elvira Romano**, Seconda Universita degli Studi di Napoli, Italy
Co-authors: Antonio Balzanella
Depth functions have shown its usefulness in functional data analysis where some ordering is needed. Several notions of depth for functional data have been proposed for outliers detection, for measuring the centrality of a function, for providing graphical representations of functional data. Often functional data arise from measurements of environmental variables which are characterized by spatial dependence. We provide a definition of depth for geostatistical functional data which considers both the functional and the spatial nature of the data. By means of a simulation study we will show how the proposed depth measure is able to catch the centrality of a function and to provide a center to outward ordering of a sample of georeferenced curves.

ES87 Room B35 NONPARAMETRIC COPULA MODELS Chair: Stephen Walker

E587: **Bayesian copula-based spatial modeling using penalized splines**
Presenter: **Hannes Kazianka**, University of Klagenfurt, Austria
Copulas have recently attracted much attention in spatial statistics and are used as a flexible alternative to traditional methods for non-Gaussian spatial modeling. It is assumed that the dependence structure of the observations can be described by a Gaussian copula while the marginal distributions are non-Gaussian. In a first step, approaches to Bayesian spatial interpolation, e.g. to predict the value of the underlying spatial process at ungauged locations based on the information derived from the data, are presented that cover both continuous and discrete marginal distributions. The prior distributions for the copula parameters are chosen in an objective Bayesian way. In a second step, the use of (multi-dimensional) penalized splines is proposed to additionally consider a spatial trend or covariates in the model and suggest a hybrid MCMC algorithm to sample the posterior. The model is applied to a data set consisting of right-skewed radioactivity measurements supplemented with covariates. It is shown that the flexibility introduced by the penalized splines on the one hand enhances predictive power compared to simple parametric modeling of the covariates and on the other hand improves the physical interpretation of the model.

E894: **Copula modelling of dependence in multivariate time series**
Presenter: **Michael Smith**, University of Melbourne, Australia
Almost all existing nonlinear multivariate time series models remain linear, conditional on a point in time or latent regime. Here, an alternative is proposed, where nonlinear serial and cross-sectional dependence is captured by a copula model. The copula defines a multivariate time series on the unit cube. A D-vine copula is employed, along with a factorization which allows the marginal and transitional densities of the time series to be

expressed analytically. It also provides for simple conditions under which the series is stationary and/or Markov, as well as being parsimonious. A parallel algorithm for computing the likelihood is given, along with a Bayesian approach for computing inference based on model averages over parsimonious representations of the copula. The model average estimates are shown to be more accurate in a simulation study. Two five-dimensional time series from the Australian electricity market are examined. In both examples, the fitted copula captures substantial asymmetric tail dependence, both over time and across elements in the series.

E1181: **Flexible sampling for the Gaussian copula extended rank likelihood model**
Presenter: **Ricardo Silva**, UCL, United Kingdom
Co-authors: Alfredo Kalaitzis

The framework of copula modeling has gained popularity due to its modular parametrization of joint distributions. Among other properties, copulas provide a recipe for combining flexible models for univariate marginal distributions with parametric families suitable for potentially high dimensional dependence structures. More radically, a previous extended rank likelihood approach bypasses learning marginal models completely when such information is ancillary to the learning task at hand as in, e.g., standard dimensionality reduction problems or copula parameter estimation. The main idea is to represent data by their observable rank statistics, ignoring any other information from the marginals. Inference is typically done in a Bayesian framework with Gaussian copulas, and it is complicated by the fact this implies sampling within a space where the number of constraints increases quadratically with the number of data points. The result is slow mixing when using off-the-shelf Gibbs sampling. We present an efficient algorithm based on recent advances on constrained Hamiltonian Markov chain Monte Carlo that is simple to implement and does not require paying for a quadratic cost in sample size.

ES111 Room B29 COMPUTATIONAL STATISTICS I Chair: Niels Richard Hansen

E871: **Estimation of the typical ranks of $2 \times 2 \times \cdots \times 2$ tensors through intrinsic dimension estimators**
Presenter: **Toshio Sakata**, Kyushu University, Japan
Co-authors: Toshio Sumi, Mitsuhiro Miyazaki

The tensor data, multi-way data, are recently applied to many field successfully. The rank of a tensor T is a measure of complexity of tensors. The maximal tensor rank has attracted many researchers' interest, however, statistically speaking, the typical rank is more important. The typical rank of a given type of tensors is defined to be an integer r such that the set of tensors with the rank r has a Lebesgue measure greater than 0. To determine a typical rank is difficult and attracted many researchers' concern, too. We investigate the possibility of determining the typical rank of $2 \times 2 \times \cdots \times 2$ tensors by estimating statistically the intrinsic dimension of the manifold formed of tensors of a rank less than or equal to r. In fact this is the situation that we have a manifold and points over the manifold and estimate its intrinsic dimension by the data points. This is the first such challenge. We will show by simulation that the maximum likelihood estimation of the intrinsic dimension is useful to estimate the typical rank. Several methods other than the maximum likelihood estimation will be also compared. Finally we note that our method is also applicable to the other types of tensors.

E1204: **Challenges in extremal index estimation through computer intensive procedures**
Presenter: **Manuela Neves**, University of Lisbon and CEAUL, Portugal
Co-authors: Dora Prata Gomes

In statistics of extremes there are a few parameters of particular interest characterizing the behavior of extreme or even rare events. Among these parameters the extremal index appears for dependent sequences. It measures the relationship between the dependence structure of the data and the behavior of the exceedances over high thresholds. It can be roughly defined as the reciprocal of the 'mean time of duration of extreme events' and needs to be adequately estimated not only by itself, but also because its influence in the estimation of other parameters of interest. Most semi-parametric estimators of this parameter show the same behavior: nice asymptotic properties but a high variance for small values of k, the number of upper order statistics used in the estimation and a high bias for large values of k. The Mean Square Error, a measure that encompasses bias and variance, usually shows a very sharp plot, needing an adequate choice of k. After reviewing some classical extremal index estimators, the emphasis is now given to explore new heuristic procedures, helped by computational techniques, for obtaining efficient estimates. A simulation study will illustrate the properties of the estimators and the performance of the adaptive algorithm proposed.

E887: **A comparison of computational approaches for maximum likelihood estimation of the Dirichlet parameters on high dimensional data**
Presenter: **Marco Giordan**, Fondazione Edmund Mach, Italy
Co-authors: Ron Wehrens

Likelihood estimates of the Dirichlet distribution parameters can be obtained only through numerical algorithms. Even though the Dirichlet distribution belongs to the Exponential Family, such algorithms can give estimates outside the correct range for the parameters. In addition, they can require a large amount of iterations to reach convergence. These problems can be exasperated if good starting points are not provided. We discuss several approaches looking at the trade-off between speed and stability. We illustrate the combinations of initialization and estimation methods using both real and simulated data of high dimension.

E858: **Two depth based strategies for robust estimation of a predictive distribution of a data stream**
Presenter: **Daniel Kosiorowski**, Cracow University of Economics, Poland

Data stream analysis can be informally defined as a computationally challenging sequence of stochastic process analyses conducted on-line basing on sliding window or windows from the process. They may lead to a sequence of investment decisions in case of an algorithmic trading, to dangerous behaviour detections in case of a city monitoring, to mining in astronomic data sets, etc. Data streams are generated by nonstationary and multiregime processes. Statistical procedures used for the data stream analysis should be nonparametric, be able to cope with "curse of dimensionality", be computationally feasible and robust but sensitive to changes of a majority of observations in the window (the regime of the underlying process detection). We present two depth based strategies for predictive distribution of the stream estimation. Our first strategy "dynamic robust rule of thumb" appeals to general considerations concerning Student depth. Our second strategy appeals to an idea of adjusted conditional distribution estimator previously proposed. We adjust this estimator using projection depth that protects us against outliers and inliers. We show properties of our proposals using multiregime econometric models with deterministic and stochastic transitions between regimes as well as empirical examples.

ES127 Room G16 CONTRIBUTIONS TO STATISTICS IN ACTUARIAL SCIENCES Chair: M. Dolores Martinez Miranda

E454: **Love and death: A Freund model with frailty**
Presenter: **Yang Lu**, University Paris Dauphine, France
Co-authors: Christian Gourieroux
New models for analyzing the mortality dependence between individuals in a couple are introduced. The mortality risk dependence is usually taken into account in the actuarial literature by introducing an Archimedean copula. This practice implies symmetric effects on the remaining lifetime of the surviving spouse. The new model allows for both asymmetric reactions by means of a Freund model, and risk dependence by means of an unobservable common risk factor (or frailty). These models allow for distinguishing in the lifetime dependence the component due to common lifetime (frailty) from the broken-heart syndrom (Freund model). The model is applied to insurance products such as joint life policy, last survivor insurance, or contracts with reversionary annuities.

E335: **Distorted bimodal distributions using an inverse *S*-shaped distortion function**
Presenter: **Fanirisoa Zazaravaka R Hasinavonizaka**, Pantheon Sorbonne, France
Co-authors: Hanjarivo Lalaharison, Dominique Guegan
Distorted distributions have been used in pricing of insurance contracts for a long time under the concept of distortion risk measure. The transformation by distortion converts the initial distribution to an equivalent distribution by re-weighting the initial probability. They were originally applied to a wide variety of financial problems such as the determination of insurance premiums, economic capital, and capital allocation. A new family of bimodal distributions generated by distortion using an inverse *S*-shaped function with one saddle point is investigated. This function generates two peaks on the repartition of the probability due to the saddle point. This process converts the initial distribution to a new distribution which is bimodal. A characterization is presented and a comparison is proposed between this family and some well-known bimodal distributions, like a mixture of two normal distributions, and beta-normal distributions. The goodness-of-fit for a data set of rolling annual returns from 2000 to 2011 for U.S. equities, commodities and developed market equities and U.S. REITs is presented.

E1238: **De-risking strategy: Modeling buy-in**
Presenter: **Valeria D Amato**, University of Salerno, Italy
Co-authors: Emilia Di Lorenzo, Marilena Sibillo
Buy-in and buy-out market is growing. Insurance companies and investment banks are offering the instruments under consideration to allow longevity risk to be managed. Over the past few years, pension funds have become an increasingly important issue for companies. Pension de-risking helps corporate sponsors and trustees remove or reduce the risk of additional financial costs associated with pension schemes. The aim is the valuation of the longevity risk impact on pension schemes, to establish if buy-in and buy-out can improve the fund performance, as well as the increase of the benefits to be corresponded to the pensioner. Empirical applications are presented by using a graphical and numerical analysis.

ES118 Room G15 THE ECONOMETRICS OF BUSINESS CYCLE: SOME RECENT DEVELOPMENTS Chair: Mateusz Pipien

E779: **Discrete spectrum and subsampling in testing business cycle similarity**
Presenter: **Lukasz Lenart**, Cracow University of Economics, Poland
Co-authors: Mateusz Pipien
A nonstandard subsampling procedure is proposed to detect almost periodic structure in mean function for multivariate almost periodically correlated time series. The asymptotic normality was proved for normalized Fourier coefficient estimator and the consistency of subsampling procedure was shown. Next the asymptotically consistent test for frequency significance was proposed. It was shown that some characteristics of business cycle comovement can be modelled by discrete spectrum parameters of the Almost Periodically Correlated time series. Finally, the proposed test was applied in order to make formal statistical inference about the business cycle (one of the most important unobserved features characterizing fluctuations of economic growth) comovement.

E781: **A Bayesian approach to business cycle synchronization: Testing for common almost periodic fluctuations**
Presenter: **Blazej Mazur**, Cracow University of Economics and National Bank of Poland, Poland
The problem of testing hypotheses corresponding to various degrees of business cycle synchronization within a group of countries is considered. A Bayesian parametric framework is adopted, which allows for obtaining density forecasts for the variables of interest. In order to model economic fluctuations we explore the idea of almost-periodicity. We consider a non-stationary process with time-varying mean modeled by an almost periodic function of time, which is in turn represented by the Flexible Fourier Form. In a multivariate setting, deviations from the mean are assumed to be of a stationary VAR form. The approach allows for making inferences regarding comparison of amplitudes, phase shifts and frequencies of the individual components among countries. As an application we consider the issue of business cycle synchronization for a group of Central European countries. Usefulness of the approach is illustrated by identification of similarities and differences in business cycle fluctuations in the group of countries and by generating density forecasts that take into account all the information. We also consider certain extensions to the basic model, including generalizations of the stochastic assumptions and introducing additional non-linearities. Moreover, we discuss relationship of the approach to the alternatives that assume similar dynamics.

E913: **Almost-periodicity in TVP-GARCH models for long series of daily financial returns**
Presenter: **Mateusz Pipien**, Cracow University of Economics, Poland
Co-authors: Blazej Mazur
Bayesian inference techniques has been previously used to detect Almost Periodic (AP) time variation in volatility of long financial time series, with application to daily logarithmic returns of S&P500 for the period 1950-2012. We augment the standard volatility equation of the AR(1)-t-GJR GARCH(1,1) specification by introducing a Flexible Fourier Form component that represents an Almost Periodic function of time. The resulting model can be interpreted as a second order Almost Periodically Correlated (APC) process, or as a GARCH process with time varying parameters. The process is non-stationary with time-varying, bounded unconditional variance. Similar models dealing with the problem of time-heterogeneity have been proposed in the literature however, the alternative specifications do not explore the idea of regular, almost periodic pattern of time variation. The time varying volatility effect is shown to be significant for the dataset under consideration. The above framework is extend as a flexible conditional sampling distribution in the form of the generalized asymmetric Student-t. This allows for two sources of asymmetry: one arising from a skewing mechanism and the other resulting from differences in degrees of freedom in the tails of the distribution. Moreover, the shape parameters representing asymmetry are also assumed to be time-varying. As we examine the S&P500 daily returns, it turns out that no evidence in favor of the skewing mechanism is found, but there exist a significant asymmetry in the left and the right side degrees of freedom. Moreover, the time variation in the d-o-f parameters is significant and it is different for both tails. We illustrate some consequences of the results for inference on selected risk measures (VaR and ES).

Sunday 15.12.2013 14:40 - 16:20 Parallel Session L – CFE

CSI04 Room Beveridge THE DESIGN OF OPTIMAL POLICIES Chair: Michel Juillard

C1245: **Capital controls: A normative analysis**
Presenter: **Bianca de Paoli**, Federal reserve bank of New York, United States
Co-authors: Anna Lipinska
Countries' concerns with the value of their currency have been extensively studied and documented. Capital controls can be (and often are) used as a tool to manage exchange rate fluctuations. We investigate whether countries can benefit from using such a tool. We develop a welfare based analysis of whether (or, in fact, how) countries should tax international borrowing. Our results suggest that restricting international capital flows with the use of these taxes can be beneficial for individual countries although it would limit cross-border pooling of risk. This is because while consumption risk-pooling is important, individual countries also care about domestic output fluctuations. Moreover, the results show that countries decide to restrict the international flow of capital exactly when this flow is crucial to ensure cross-border risk-sharing. Our findings point to the possibility of costly "capital control wars" and, thus, significant gains from international policy coordination.

C1254: **Computation of LQ Approximations to Optimal Policy Problems**
Presenter: **Joseph Pearlman**, City University London, United Kingdom
Co-authors: Paul Levine
This paper describes a series of algorithms that are used to compute optimal policy under full and imperfect information. Firstly we describe how to obtain linear quadratic (LQ) approximations to a nonlinear optimal policy problem. We develop novel algorithms that are required as a result of having agents with forward-looking expectations, that go beyond the scope of those that are used when all equations are backward-looking; these are utilised to generate impulse response functions and second moments for the case of imperfect information. We describe algorithms for reducing a system to minimal form that are based on conventional approaches, and that are necessary to ensure that a solution for fully optimal policy can be computed. Finally we outline a computational algorithm that is used to generate solutions when there is a zero lower bound constraint for the nominal interest rate. Most importantly, we show how to use these calculations to check the second order conditions for a maximum.

C1265: **Optimal fiscal and monetary policy mix: A Ramsey approach**
Presenter: **Giovanni Di Bartolomeo**, Sapienza University of Rome, Italy
Co-authors: Patrizio Tirelli
The widely held belief that New-Keynesian models cannot predict optimal positive inflation rates is challenged. In fact these are justified by the Phelps argument that monetary financing can alleviate the burden of distortionary taxation. We obtain this result because, in contrast with previous contributions, our model accounts for public transfers as a component of fiscal outlays. We also contradict the view that the Ramsey policy should minimize inflation volatility and induce near-random walk dynamics of public debt in the long-run. In our model it should instead stabilize debt-to-GDP ratios in order to mitigate steady-state distortions. Our results thus provide theoretical support to policy-oriented analyses which call for a reversal of debt accumulated in the aftermath of the 2008 financial crisis.

CS114 Room Deller FINANCIAL APPLICATIONS II Chair: Panayiotis C. Andreou

C1083: **Intraday herding around stock splits**
Presenter: **Maria Chiara Iannino**, University of Vienna, Austria
Co-authors: Michela Verardo
Investors' herding behavior around stock splits is investigated. Using intraday trading data on a large sample of US stocks, we estimate several measures of intraday herding over a window that includes the day of the split announcement as well as the day of the adjustment of prices and shares. Our measures of herding capture both the clustering of trades at the same time and the intertemporal correlation of buys and sells by institutional and individual investors. Our results show that, across all the different measures, herding tends to increase on both announcement and adjustment days. Furthermore, we find that herding is consistently higher during the post-adjustment period. We then test whether the patterns in herding behavior around stock splits can be used to distinguish among potential explanations for the decision to split. In particular, we focus on the liquidity and signalling hypotheses. We find that herding is more intense on adjustment days than announcement days, suggesting that improvements in liquidity and marketability are the main reasons for stock splits. Finally, we explore the signalling hypothesis by analyzing the ability of herding behavior to explain market reactions to split announcements and future stock returns.

C982: **Commonality in liquidity dimensions: A generalized dynamic factor model approach**
Presenter: **Julia Reynolds**, Vienna Graduate School of Finance, Austria
The application of factor model methods to financial data has introduced key insights into asset pricing and risk analysis, particularly in the analysis of liquidity risk. Recent studies have made use of factor analysis to determine commonality in observable liquidity measures, typically corresponding to information about price, volume, and time, to find evidence for an unobservable market liquidity factor. This study builds on this growing literature by addressing two common limitations, and by contributing evidence on the variation in market liquidity commonality over time in response to changing liquidity conditions. First, by extending recent results obtained from the Generalized Dynamic Factor Model (GDFM), the limiting assumptions imposed by the use of static factor models are likewise avoided. Secondly, by modeling the time dimension of liquidity as a mean-reversion parameter in an Ornstein-Uhlenbeck (OU) process of changes in daily stock prices, high-frequency financial data is used to construct an empirical measure of this oft-overlooked dimension, thus reconciling all three liquidity dimensions in a single GDFM analysis. Lastly, the comparison of GDFM analyses for time periods before, during, and after the 2007-2008 financial crisis provides consistent evidence that commonality in market liquidity dimensions increases following periods of extreme liquidity conditions.

C933: **Imposing views on dynamic factor models**
Presenter: **Martin van der Schans**, Ortec-Finance, Netherlands
Dynamic factor models can be used to produce multivariate density forecasts for the future evolution of (a large number of) economic and financial variables such as interest rates, asset returns and inflation rates. Since the parameters in this type model are estimated on historical data, the fundamental assumption is that there is relevance in historical data for assessing future risk and return. In practical applications, however, it is indispensable to combine the model predictions with views on the economy. This is for example the case when an economic bureau or an investment committee of an institution holds a different expectation of the equity risk premium for the next three years than the models estimated on historical data foresee. A simple approach would be to adjust the forecast for one individual variable in the model. However, this ignores any correlation structure that exists between variables as imposed by the model structure. A computationally efficient and consistent methodology will be introduced that allows for imposing of such views on the model. Consistency is attained by exploiting factor structure to determine the impact of these views on other variables. Finally, the impact on investment decisions will be investigated by constructing optimal portfolios.

C1153: **The effects of market maturity and informed trading on cash-futures basis**
Presenter: **Emily Lin**, SJU, Taiwan

Cash-futures basis, a proxy for arbitrage opportunities, is examined. The dynamics of informed trading and the changing roles of speculators and arbitrageurs are analyzed under several interesting market characteristics of an emerging futures market. We demonstrate positive basis spreads shrink but negative ones expand caused by informed trading in the index futures market during the non-expiration period, while the pattern is opposite during the pre-expiration period. This difference seems resulting from informed traders behave diversely during these two periods. We suggest market frictions also account for some portion of the negative spreads. On the contrary, volatility and liquidity each significantly widen spread magnitude for the two periods, findings that support dominance of speculation over arbitrage in the market. Finally, we construct a market-maturity measure to track the basis interactions with the above variables throughout the course of the market's maturation.

C030: **Time consistent G-expectation and bid-ask dynamic pricing mechanisms for contingent claims under uncertainty**
Presenter: **Wei Chen**, Shandong University, China

Time consistent dynamic pricing mechanisms of European contingent claims under uncertainty are studied by using the G framework. We consider a financial market consisting of a riskless asset and a risky stock with price process modelled by a geometric generalized G-Brownian motion, which features the drift uncertainty and volatility uncertainty of the stock price process. A time consistent G-expectation is defined by the viscosity solution of the G-heat equation. Using the time consistent G-expectation we define the G dynamic pricing mechanism for the claim. We prove that the G dynamic pricing mechanism is the bid-ask Markovian dynamic pricing mechanism. A full nonlinear PDE is derived to describe the bid (resp. ask) price process of the claim. Monotone implicit characteristic finite difference schemes for the nonlinear PDE are given, nonlinear iterative schemes are constructed, and the simulations of the bid (resp. ask) prices of contingent claims under uncertainty are implemented.

CS06 Room Bloomsbury PERFORMANCE EVALUATION Chair: Juha Joenvaara

C286: **The economic value and statistical properties of manipulation-proof performance measures**
Presenter: **Juha Joenvaara**, University of Oulu, Finland
Co-authors: Robert Kosowski, Jussi Klemela

Recent research on portfolio performance measures that are less prone to manipulation often appeals to non-normalities in hedge fund returns for motivation, but no comprehensive analysis of the economic value of manipulation proof performance measures (MPPMs) for hedge fund selection exists. First, we contribute by analyzing the economic value and statistical properties of MPPMs using a large consolidated hedge fund database containing option holdings data. Second, we build on a previous work by proposing a novel nonparametric way to incorporate macroeconomic information into the estimation of a conditional version of their unconditional MPPM. We show that kernel regressions exploit more accurate conditional estimates of MPPMs than standard linear regressions. Finally, we find that using our conditional version of the MPPM to rank funds delivers superior out-of-the-sample risk-adjusted performance for investors compared to a range of other alternative performance measures.

C319: **Measuring hedge fund performance: A Markov regime switching with false discoveries approach**
Presenter: **Gulten Mero**, Universite de Cergy-Pontoise and THEMA, France

A Markov regime switching approach accounting for false discoveries is developed in order to measure hedge fund performance. This approach allows us to extract information from both time-series and cross-sectional dimensions of individual fund panels in order to distinguish between skilled, unskilled and zero-alpha funds for a given state of the economy. Applying our approach to individual hedge funds belonging to Long/Short Equity Hedge strategy, we find that their performance cannot be explained by luck alone, and that the proportion of zero-alpha funds in the population decreases when accounting for regime dependence with time-varying transition probabilities. However, the proportion of truly skilled funds is higher during expansion periods, while unskilled funds tend to be more numerous during recession periods. Moreover, sorting on regime-dependent alphas instead of unconditional alphas improves the ability to maximize the performance persistence effect of top performer fund portfolios.

C362: **On the (Ab)use of Omega**
Presenter: **Michele Costola**, University of Padua, Italy
Co-authors: Massimiliano Caporin, Gregory Jannin, Bertrand Maillet

Several recent finance articles employ the Omega measure- defined as a ratio of potential gains out of possible losses - for gauging the performance of funds or active strategies, in substitution of the traditional Sharpe ratio, with the arguments that return distributions are not Gaussian and volatility is not, always, the relevant risk metric. Other authors also use the same criterion for optimizing (non-linear) portfolios with important downside risk. However, the relevance of such approaches is wondered, based on the results of a basic illustration showing that the Omega measure is inconsistent with the Second-order Stochastic Dominance (SoSD) criterion. Moreover, it is observed that the ordering of two portfolios depends upon the chosen threshold and the trade-off between return and risk corresponding to the Omega measure seems to be strongly influenced by the first moment of returns. Finally, it is shown, using complementary simulations, that the Omega ratio may be misleading in some cases when financial decisions are involved.

C892: **Disclosures difficult to deal with: Robust hedge fund exposures and alphas**
Presenter: **Alexandru Popescu**, University of Lugano, Switzerland
Co-authors: Fabio Trojani, Lorenzo Camponovo

A new method to analyze hedge fund risk-adjusted performance is proposed based on voluntarily disclosed financial information. Our method is designed to offer a superior degree of resistance to anomalous disclosures of unknown form, such as reporting errors or fund return revisions. In a consolidated data-set of voluntary hedge fund disclosures, we detect a substantial fraction of funds with robust and persistent out-of-sample performance. Other widely-used approaches often fail to identify funds with profitable payoffs, especially during turbulent market periods, because of a small fraction of influential self-reported returns detected by our method. This evidence points to the need of more robust methods for interpreting publicly accessible financial information that is voluntarily disclosed to widely-used hedge fund databases.

CS07 Room Athlone FACTOR MODELS AND FORECASTING APPLICATIONS Chair: Scott A. Brave

C348: **Wavelet-based early warning composite indicators: An application to the US financial stress index**
Presenter: **Marco Gallegati**, Polytechnic University of Marche, Italy

The usefulness of wavelet multi resolution analysis for the construction of composite indicators with leading properties is exploited. The proposed "wavelet-based" methodology is applied to the construction of an early warning composite index of financial stress obtained by aggregating several scale-based sub-indexes whose components are selected on the basis of their cross-correlations properties at each frequency band. The benefits of the proposed methodology are assessed by evaluating its predictive power relative to the individual financial variables taken in isolation through an out-of-sample forecasting exercise. The findings indicate that the wavelet-based composite indicator largely outperforms any individual financial variable taken in isolation in early detecting financial stress at every horizon and that the gain tends to increase as the time horizon increases.

C512: **Core and crust: Consumer prices and the term structure of interest rates**
Presenter: **Andrea Ajello**, Board of Governors of the Federal Reserve System, United States
Co-authors: Luca Benzoni, Olena Chyruk
The aim is to propose a model for nominal and real term structures of interest rates that includes core, food, and energy inflation series. This framework captures different frequencies in inflation fluctuations: Shocks to core are more persistent and less volatile than shocks to food and, especially, energy (the 'crust'). The three inflation series are combined into a measure of total inflation to price nominal Treasuries. It is found that a common structure of latent factors determines and predicts the term structure of yields and inflation. The model outperforms popular benchmarks, e.g., ARMA, VAR, and random walk specifications, and it is at par with the Survey of Professional Forecasters in forecasting inflation. Real rates implied by the proposed model uncover the presence of a time-varying component in TIPS yields that is attributed to well-documented disruptions in the inflation-indexed bond market. Finally, a pronounced declining pattern in the inflation risk premium is found, that illustrates the changing nature of inflation risk in nominal Treasuries.

C681: **Evaluating nowcasts of GDP growth**
Presenter: **Robert Butters**, Northwestern University, United States
Co-authors: Scott Brave
The aim is to evaluate several alternative forecasting procedures for the current quarter's growth in Gross Domestic Product (GDP) as well as leading professional forecasts such as the Survey of Professional Forecasters and Macro Advisors. By leveraging a real-time archive, spanning from March 2001—May 2013, of the indicators used to create the Chicago Fed National Activity Index (CFNAI) the relative out of sample performance of several leading "nowcasting" models is investigated. The single factor coincident indicator framework is augmented by including GDP growth within the panel of monthly indicators creating a mixed frequency panel. Six different models are estimated GDP growth is forecast out of sample. Root mean squared error (RMSE) measures of the forecast performance are provided. The results implications on the recovery for the most recent recession and the long term trend of GDP are commented.

C882: **US bank holding companies and systemic risk**
Presenter: **Scott Brave**, Federal Reserve Bank of Chicago, United States
Co-authors: Hesna Genay
A hierarchical extension of the factor-augmented VAR (FAVAR) framework is developed. The key distinction of our model is that the dynamics of the level-one latent common factors are assumed to follow a fixed effects panel VAR (PVAR) structure which includes observable macroeconomic factors. Using our framework, we examine the systemic risk associated with shocks to two latent common factors extracted from a large number of variables describing the condition of US bank holding companies. These shocks are shown to produce impulse responses with properties resembling liquidity and capital shocks to the banking sector that produce significant effects on economic activity. We further demonstrate that it is possible to construct time-varying network analogues of connectedness through a variance decomposition of these shocks at the individual bank holding company level. Periods of financial stress are then shown to be correlated with increases in connectedness.

CS19 Room Court COPULAS IN FINANCE: NEW DEVELOPMENTS Chair: Jean-David Fermanian

C152: **Portfolio default models with lack-of-memory**
Presenter: **Matthias Scherer**, Technische Universitaet Muenchen, Germany
Co-authors: Jan-Frederik Mai, Steffen Schenk, Natalia Shenkman
The multivariate lack-of-memory property characterizes the Marshall-Olkin law (in the continuous case) and the wide-sense geometric distribution (in the discrete case). Both distributions have various appealing properties to model joint defaults or insurance claims. However, it is challenging to apply them in high dimensions, due to their exponentially increasing number of parameters. Moreover, in some cases it is reasonable to relax the lack-of-memory property to allow for richer stylized facts. We take both requirements into account and design high-dimensional portfolio default models that consider tractable subfamilies of the Marshall-Olkin law as starting point and enrich them to obtain new multivariate distributions that might also be interesting for other applications in reliability theory.

C163: **A note on intra day price spike dependence in the Australian power market**
Presenter: **Hans Manner**, University of Cologne, Germany
Co-authors: Oliver Grothe, Dennis Tuerk
Understanding the co-dependence of extreme observations, so called spikes, in real-time electricity prices between interconnected markets has a crucial role in risk-management. Furthermore, it is of great importance when pricing and hedging interregional spreads and/or valuing interconnectors between these markets. In this context the Australian power exchange provides an ideal framework. While univariate models for spikes have recently been proposed for intradaily data in these markets, the nature of co-spikes still remains largely unexplored in the existing literature. We address this topic and analyze the occurrence of such co-spikes and their dependence by using bivariate binary choice models with a general dependence structure. Furthermore, we provide a model for the joint distribution of co-spikes that is suitable for analyzing price spikes without the need to specify a model for moderate prices. We find evidence for strong linkages in the tails and non-exchangeable dependence structures.

C165: **Arbitrage detection in a multi-asset market: The edge of copulas**
Presenter: **Bertrand Tavin**, Pantheon Sorbonne University, France
The existence of arbitrage opportunities in a multi-asset market is studied when risk-neutral marginal distributions of asset prices are known. We first propose an intuitive characterization of the absence of arbitrage opportunities in terms of copula functions. We then address the problem of detecting the presence of arbitrage by formalizing its resolution in two distinct ways that are both suitable for the use of optimization algorithms. The first method is valid in the general multivariate case and is based on Bernstein copulas that are dense in the set of all copula functions. The second one is easier to work with, but is only valid in the bivariate case. It relies on recent results about improved Frechet-Hoeffding bounds in presence of additional information. For both methods, details of implementation steps and empirical applications are provided.

C226: **Approximation of tail dependence via a new family of copulas**
Presenter: **Fabrizio Durante**, Free University of Bozen-Bolzano, Italy
In order to estimate some risk measures derived from a set of observations of correlated variables, a widely used method consists in fitting some parametric joint model to the data and, hence, calculate the risk measure by using these fitted models. However, due to model uncertainty, this procedure is not conservative from the viewpoint of a risk manager, as the middle part of the distribution might influence the overall estimation more than the tail part. In order to overcome such difficulties, it could be convenient to develop dependence models for conditional joint extremes, along the lines of the Pickands-Balkema-De Haan Theorem. Starting with this idea, a new class of copulas has been recently characterized, which can be used to approximate the dependence structure of a random vector when one component of the vector is taking on extreme values. We present this new family of copulas, which can capture non-exchangeable dependence structures and can be easily simulated. Moreover, related inference methods will be presented, especially about a goodness-of-fit testing.

CS22 Room Russell MULTIPLE RISKS MANAGEMENT Chair: Christian Francq

C181: **Credit and liquidity in interbank rates: A quadratic approach**
Presenter: **Jean-Paul Renne**, Banque de France, France
Co-authors: Simon Dubecq, Alain Monfort, Guillaume Roussellet
A quadratic term-structure model of the EURIBOR-OIS spreads is proposed. These spreads are affected by both credit and liquidity risks, which we aim at disentangling. Liquidity effects reflect banks preferring a portfolio of cash and easy-to-liquidate swap contracts to interbank loans, to cope with potential future liquidity needs. Credit effects correspond to the premium required by the lender for the borrower's default risk compensation. Our approach allows us to identify credit and liquidity effects and to further decompose the whole term structure of spreads into credit- and liquidity-related parts. Our results shed new light on the effects of unconventional monetary policy carried out in the Eurosystem. In particular, our findings suggest that most of the recent easing in the euro interbank market is liquidity related.

C518: **Endogenous financial network formation**
Presenter: **Jean Cyprien Heam**, Autorite de Controle Prudentiel and CREST, France
Co-authors: Erwan Koch
A model of financial network formation is proposed where interconnections results from endogenous actions of financial institutions, taking into account the specificities of banking and insurance industries (maturity transformation, sinistrality...). The interconnections, based on both cross-stockholding and debt holding, appear are best protection against financial shocks such as nonperforming loans or against physical shock such as natural disasters. The influence on the financial stability and on the social welfare of the financial institution characteristics and of the regulation design is analyzed.

C607: **Identifying SIFIs: Toward the simpler approach**
Presenter: **Jeremy Dudek**, CREST - Paris Dauphine, France
Co-authors: Sylvain Benoit, Manizha Sharifova
Systemic risk measures generally aim to identify systemically important financial institutions (SIFIs) that would allow regulators to allocate macro-prudential capital requirements in order to reduce risk stemming from such institutions. Among widely-cited are the measures of tail dependence in financial institutions' equity returns, such as δCoVaR and Marginal Expected Shortfall (MES). The aim is to compare nonlinear and linear approaches to modeling return dependence in the estimation of the δCoVaR and MES. The results show that while the refined and complicated estimation techniques are able to produce more accurate value of institution's systemic risk contribution they do not greatly improve in terms of identifying SIFIs compared to simpler linear estimation method. Modeling dependence linearly sufficient to identify and rank SIFIs.

C662: **Multi-level conditional VaR estimation in dynamic models**
Presenter: **Jean-Michel Zakoian**, CREST, France
Co-authors: Christian Francq
Joint estimation of conditional Value-at-Risk (VaR) is considered at several levels, in the framework of general conditional heteroskedastic models. The volatility is estimated by Quasi-Maximum Likelihood (QML) in a first step, and the residuals are used to estimate the innovations quantiles in a second step. The joint limiting distribution of the volatility parameter and a vector of residual quantiles is derived. We deduce confidence intervals for general Distortion Risk Measures (DRM) which can be approximated by a finite number of VaR's. We also propose an alternative approach based on non Gaussian QML which, although numerically more cumbersome, has interest when the innovations distribution is fat tailed. An empirical study based on stock indices illustrates the theoretical findings.

CS27 Room Holden MIXTURE AND REGIME-SWITCHING MODELS IN EMPIRICAL FINANCE Chair: Markus Haas

C331: **Markov switching variance regimes for identification of contemporaneous causalities between real activity and financial markets**
Presenter: **Sebastian Mueller**, Christian-Albrechts-University Kiel, Germany
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 economic theory. 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, from assumptions on the long-run effects of shocks or by restricting parameter signs based on apriori knowledge or researchers judgement. Recent statistical alternatives make use of information from higher-order moments. The identification problem can be solved by exploiting Markov switching regimes in the structural variances for an exact identification. Thus, additionally imposed restrictions become testable. A Markov switching variance structural vector-autoregressive model (MSH-VAR) is employed to investigate the contemporaneous causal transmission pathways of structural shocks between real economic activity, stock and commodity markets. Subsequently, previous stated assumptions from economic theory are re-examined.

C585: **Independent component analysis meets maximum likelihood estimation**
Presenter: **Jochen Krause**, University of Zurich, Switzerland
Co-authors: Marc Paolella
A new maximum likelihood-based approach to independent component analysis (ICA) is introduced which facilitates dimension reduction and model estimation in a one-step procedure. Fast maximum likelihood estimators as well as a statistical test for the optimal number of independent components are derived. Besides the necessity of having non-Gaussian component models as is mandatory for ICA, the method is general and can be extended to include different models for the individual (independent) components. Based on augmented likelihood estimation (ALE) for the estimation of finite mixture distributions, an application to portfolio allocation using mixtures of normals as component models and with expected shortfall as the downside risk measure is given.

C612: **On the continuous limit of GARCH**
Presenter: **Emese Lazar**, University of Reading, United Kingdom
Co-authors: Carol Alexander
GARCH processes constitute the major area of time series variance analysis hence the limit of these processes is of considerable interest for continuous time volatility modelling. The limit of the GARCH(1,1) model is fundamental for limits of other GARCH processes yet it has been the subject of much debate. This limit has been previously derived as a stochastic volatility process that is uncorrelated with the price process but later the limit was derived as a deterministic volatility process and several other contradictory results followed. This continuous limit is reconsidered, arguing that because the strong GARCH model is not aggregating in time it is incorrect to examine its limit. Instead it is legitimate to use the weak definition of GARCH that is time aggregating.

C634: **A multivariate regime-switching GARCH model with an application to portfolio optimization during financial turmoil**
Presenter: **Markus Haas**, University of Kiel, Germany
Co-authors: Ji-Chun Liu
The aim is to consider a multivariate regime-switching GARCH model with regime-dependent but otherwise constant conditional correlation matrices. This model can be viewed as a generalization of the one suggested by Pelletier, which is an interesting alternative to DCC-type specifications. Stationarity conditions are derived and consistency of the MLE is established. A Lagrange multiplier test for within-regime correlation dynamics is devised, and a simple recursion for calculating multi-step ahead conditional covariance matrices is provided. An application to mean-variance portfolio optimization involving international stock and real estate equity markets shows that accounting for regime-dependent correlations may reduce portfolio risk as compared to both CCC- and DCC-type processes.

CS31 Room Chancellor's CO-MOVEMENTS IN MACROECONOMICS AND FINANCE Chair: Alain Hecq

C069: **Long memory through correlation**
Presenter: **Guillaume Chevillon**, ESSEC Business School, France
Co-authors: Alain Hecq, Sebastien Laurent
The aim is to analyze a novel source of long memory through multicollinearity. We consider a vector autoregression of order one, a VAR(1) of large dimension, and whose matrix coefficient is close to the identity, with small off diagonal elements. We use a final equation representation to show that as the VAR dimension tends to infinity, individual variables exhibit strong persistence akin to fractional integration of order strictly less than 1/2. Hence long memory can arise when considering the individual elements of a vector which nearly consists of uncorrelated random walks. We consider the implications of our findings for the volatility of asset returns.

C305: **Modelling volatility co-movements in international stock markets using single equation models**
Presenter: **Gianluca Cubadda**, University of Rome Tor Vergata, Italy
Co-authors: Barbara Guardabascio, Alain Hecq
The presence of common volatility in financial markets, a feature that is extensively documented in the contagion literature, is difficult to empirically assess when the number of assets increases. For instance, the pure variance model based on a canonical correlation framework seems to have severe size distortions when the number of series N is larger than 2 or 3. We determine in a simple way the number of common components transmitting the whole time varying feature of about 10 to 25 assets. Doing this, we are in the so-called medium N framework, where both the time series models suffer from the curse of dimensionality and the assumptions for double asymptotics are not fully met. We first use a vector autoregressive model to capture the dynamics of about 10 to 15 daily realized volatility series computed on intraday data. However, rather than determining the number of common volatility factors at the multivariate level, we elaborate on our previous results concerning the determination of common business cycle fluctuations in the conditional mean. We show that parsimonious single-equation models can be estimated on each realized volatility series to split a set of N markets in three groups. The first group comprises markets that share a synchronous common volatility feature, a non-synchronous common volatility component is present among the markets of the second group, and the third group collects markets that exhibit idiosyncratic volatilities. Moreover, we offer a method for constructing a portfolio without any time varying component, and hence a portfolio whose risk is much easier to forecast.

C355: **Generalised canonical correlation analysis revisited**
Presenter: **Michael Massmann**, VU Amsterdam, The Netherlands
The aim is to revisit the problem of generalising canonical correlation analysis to settings when the linear relationship between the covariates is disturbed by error terms that are not Normally and independently distributed (NID). Particular interest lies on inferential procedures for the number of generalised canonical variates, such as the likelihood ratio test or the approximations suggested previously. Several departures from the usual NID error specification are considered: stationary ARMA(p, q) errors, GARCH(p,q) errors and t-distributed errors. An empirical application of the methodology to forecasting Eurozone inflation is finally included, manufacturing output and employment using a large number of predictors.

C777: **A stochastic discount factor approach to asset pricing using panel data asymptotics**
Presenter: **Joao Victor Issler**, Getulio Vargas Foundation, Brazil
Using the Pricing Equation in a panel-data framework, we construct a novel consistent estimator of the stochastic discount factor (SDF) which relies on the fact that its logarithm is the "common feature" in every asset return of the economy. Our estimator is a simple function of asset returns and does not depend on any parametric function representing preferences.The techniques discussed were applied to two relevant issues in macroeconomics and finance: the first asks what type of parametric preference-representation could be validated by asset-return data, and the second asks whether or not our SDF estimator can price returns in an out-of-sample forecasting exercise. In formal testing, we cannot reject standard preference specifications used in the macro/finance literature. Estimates of the relative risk-aversion coefficient are between 1 and 2, and statistically equal to unity. We also show that our SDF proxy can price reasonably well the returns of stocks with a higher capitalization level, whereas it shows some difficulty in pricing stocks with a lower level of capitalization.

CS109 Room Bedford APPLIED ECONOMETRICS I Chair: Maral Kichian

C876: **Variable-addition tests for multivariate regression models**
Presenter: **Jose Murteira**, FEUC and CEMAPRE, Portugal
Specification tests for multivariate regression models that generalize goodness-of-link tests for univariate generalized linear models are presented. The proposed procedure leads to an approximation of the maintained model that involves the inclusion of additional covariates within the null conditional mean, which can then be assessed through a significance test of added regressors. The use of the proposed methodology is illustrated for several common regression models of discrete choice and multivariate fractional data, as well as for non-nested regression models.

C1110: **Carry trade activities: A multivariate threshold model analysis**
Presenter: **Matthias Gubler**, Swiss National Bank, Switzerland
The relationship between carry trade positions and some key financial as well as macroeconomic variables is analyzed using a multivariate threshold model. It is often stated that the Swiss franc serves as a funding currency. Therefore, carry trades based on the currency pairs US dollar/Swiss franc and euro/Swiss franc over the period from 1995 to mid-2008 are examined. Generalized impulse responses differ in magnitude and significance between periods with a large and small interest-rate differential (so-called carry). While a positive shock to the carry is followed by a rise in carry trade positions in times with a small carry, it triggers a decline in these positions in the other regime. The results suggest that the shock to the carry itself is not enough to compensate investors for the increased foreign exchange risk. Moreover, a positive stock market price shock is associated with a rise in carry trade positions since investors may use stock portfolios as collateral for liquidity. A sudden unwinding of carry trades significantly impacts the exchange rate, irrespective of the regime. Furthermore, carry trade activities Granger-cause the nominal exchange rate in periods with a small carry. The Granger causality test results further indicate feedback trading.

C1113: **A multifactor model of credit spread changes**
Presenter: **Carolina Castagnetti**, University of Pavia, Italy
Co-authors: Eduardo Rossi

The determinants of credit spread changes are investigated. We adopt a heterogeneous panel with a multifactor error model to investigate the determinants of credit spread changes. We verify in advance the presence of a factor structure in the error term by means of two recently proposed test procedures. Our data support the assumption of a one unobserved common factor. We make use of a two-step estimation procedure. The analysis is carried out with an unbalanced panel of monthly data on more than 500 European issues over the period 2000 to 2011. Our empirical results suggest that the CDS spread variation alone predicts large part of the delta credit spread and reduces the impact of the unobserved factor. Comparison with previous findings suggest that the CDS are able to account for a large part of the liquidity premium in the corporate bond market.

C1134: **A random walk stochastic volatility model for income inequality**
Presenter: **Haruhisa Nishino**, CHIBA University, Japan
Co-authors: Kazuhiko Kakamu

Income inequality is a great issue also in the Japanese economy likewise the world economy, and for example the Gini coefficient is a famous measure of economic inequalities. We estimate inequality, including the Gini coefficients, using a lognormal parametric model. Since microdata are difficult to access in Japan and also in other developing countries, we use grouped data. The asymptotic theory of selected order statistics enables us to construct a linear model based on grouped data. We extend the linear model to a dynamic model in the form of a stochastic volatility (SV) model. We consider a stochastic volatility model and a random walk stochastic volatility model. Using Japanese data, we estimate these SV models by the Markov Chain Monte Carlo (MCMC) method and compare the two models.

C367: **Persistence of informality in a developing country: An Ar (1) pseudo panel**
Presenter: **Jhon James Mora**, Icesi, Colombia
Co-authors: Juan Muro

Informality is a common phenomenon in developing countries and an unusual one in industrialized countries. The persistence of informal employment is indicative of the impossibility of moving out of this status for a certain period of time. Using pseudo panel data, empirical evidence is presented to show that this phenomenon occurs in a developing country like Colombia where education helps mitigate the persistence of informality. We model a pseudopanel Ar(1) of informality and discuss the permanent effects of minimum wage over informality.

CS40 Room Jessel LIQUIDITY RISK Chair: Gaelle Le Fol

C089: **Large tick assets: Implicit spread and optimal tick size**
Presenter: **Mathieu Rosenbaum**, University Pierre and Marie Curie, France
Co-authors: Khalil Dayri

A framework linking microstructural properties of an asset to the tick value of the exchange is provided. In particular, we bring to light a quantity, referred to as implicit spread, playing the role of spread for large tick assets, for which the effective spread is almost always equal to one tick. The relevance of this new parameter is shown both empirically and theoretically. This implicit spread allows us to quantify the tick sizes of large tick assets and to define a notion of optimal tick size. Moreover, our results open the possibility of forecasting the behavior of relevant market quantities after a change in the tick value and to give a way to modify it in order to reach an optimal tick size. Thus, we provide a crucial tool for regulators and trading platforms in the context of high frequency trading.

C268: **Liquidity risk estimation in conditional volatility models**
Presenter: **Serge Darolles**, Paris Dauphine, France
Co-authors: Christian Francq, Gaelle Le Fol, Jean Michel Zakoian

Until recently, the liquidity of financial assets has typically been viewed as a second-order consideration in the asset-management industry. Liquidity was frequently associated with simple transaction costs that impose little effect, temporary if any, on asset prices and whose shocks could be easily diversified away. Yet, the evidence, especially the recent liquidity crisis, suggests that liquidity is now a primary concern. The aim is to propose a static liquidity risk measure leading to a better evaluation of the latter risk by distinguishing the market volatility shocks with persistent effects from liquidity shocks with temporary effects. This approach will allow isolating the liquidity risk even in the case where volumes are not observed.

C969: **Funding liquidity risk from a regulatory perspective**
Presenter: **Christian Gourieroux**, University of Toronto and CREST, Canada
Co-authors: Jean-Cyprien Heam

In the Basel regulation the required capital of a financial institution is based on conditional measures of the risk of its future equity value such as Value-at-Risk, or Expected Shortfall. In Basel 2 the uncertainty on this equity value is captured by means of changes in asset prices (market risk), and default of borrowers (credit risk). This analysis is extended by taking also into account the funding and market liquidity risks. These latter risks are consequences of changes in customers or investors' behaviors and usually concern the liability component of the balance sheets. In this respect our analysis is in the spirit of the most recent Basel 3 and Solvency 2 regulations. Our analysis highlights the role of the different types of risks in the total required capital. It also leads to clearly distinguish defaults due to liquidity shortage and defaults due to a lack of solvency.

C183: **Liquidity and credit risk premia in government bond yields**
Presenter: **Magdalena Grothe**, European Central Bank, Germany
Co-authors: Jacob Ejsing, Oliver Grothe

The aim is to quantify liquidity and credit premia in German and French government bond yields. For this purpose, we estimate term structures of government-guaranteed agency bonds and exploit the fact that any difference in their yields vs. government bonds can be attributed to differences in liquidity premia. Adding the information on risk-free rates, we obtain model-free and model-based gauges of sovereign credit premia, which are an important alternative to the information based on CDS markets. The results allow us to quantify the price impact of so-called safe haven flows, which strongly affected bond markets in late 2008/early 2009 and again during some phases of the sovereign debt crisis. Thus, we show to what extent these effects disguised the increase of sovereign credit premia in the government yields of core euro area countries.

CS91 Room Senate SIMULATION-BASED INFERENCE IN ECONOMETRICS Chair: Gael Martin

C980: **Numerically accelerated importance sampling for nonlinear non-Gaussian state space models**
Presenter: **Siem Jan Koopman**, VU Amsterdam, The Netherlands
Co-authors: Andre Lucas, Marcel Scharth
A general and efficient likelihood evaluation method for nonlinear non-Gaussian state space models is proposed via the combination of numerical and Monte Carlo integration methods. The method minimizes the simulation effort required for efficient importance sampling (EIS) to construct a global approximation to the likelihood function by replacing a key part of the simulations by numerical integrations only. We refer to our method as numerically accelerated importance sampling (NAIS). Using the new method, we can evaluate efficiently the likelihood function for models with high-dimensional state vectors and low-dimensional signals. We find large efficiency gains in an extensive Monte Carlo study as well as in an empirical application using a stochastic volatility model for U.S. stock returns with multiple volatility factors.

C603: **Efficient method of moments estimators for integer time series models**
Presenter: **Andrew Tremayne**, University of Liverpool, United Kingdom
Co-authors: Vance Martin, JUng Robert
The parameters of Poisson integer autregressive models can be estimated by maximum likelihood where the prediction error decomposition, together with convolution methods, is used to write down the likelihood function. When a moving average component is introduced this is not the case. The use of efficient method of moment techniques is considered as a means of obtaining practical estimators of relevant parameters using simulation methods. Under appropriate regularity conditions, the resultant estimators are consistent, asymptotically normal and under certain conditions achieve the same efficiency as maximum likelihood estimators. Simulation evidence on the efficacy of the approach is provided and it is seen that the method can yield serviceable estimates, even with relatively small samples. Estimated standard errors for parameters are obtained using subsampling methods. Applications are in short supply with these models, though the range is increasing. An example is provided using a well-known data set in the branching process literature that has hitherto proved difficult to model satisfactorily; this requires use of a mixed specification with special features.

C890: **Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator**
Presenter: **Michael Pitt**, University of Warwick, United Kingdom
When an unbiased estimator of the likelihood is used within a Metropolis-Hastings scheme, it is necessary to tradeoff 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 specific assumptions about the likelihood estimator, we provide guidelines on the number of samples to select for a general Metropolis-Hastings proposal. We give elements of theory which justifies the use of these assumptions for a large class of models. We illustrate on a variety of models that these assumptions are approximately satisfied experimentally and that the theoretical insights with regards to inefficiency and computational time hold true.

C545: **Approximate Bayesian computation in state space models**
Presenter: **Gael Martin**, Monash University, Australia
Co-authors: Brendan McCabe, Christian Robert, Ole Maneesoonthorn
Exploiting the likelihood-free techniques of approximate Bayesian computation, a new approach to inference in state space models is proposed. Building on the concept of the asymptotic sufficiency of the maximum likelihood estimator (MLE), the MLE of the parameters of an auxiliary, or approximating, model forms the basis for the summary statistic in an ABC matching criterion. The augmented unscented Kalman filter (AUKF) is used for this purpose, with the speed and simplicity of the AUKF calculations being critical to the success of the ABC scheme. Consideration is given to the very concept of sufficiency in the state space context, and to the effectiveness of ad-hoc summary measures in this setting. The application of the proposed method in multiple parameter settings is an important consideration, with a separate treatment of scalar (or lower-dimensional blocks of) parameters proposed, based on marginal likelihood methods. Particular emphasis is given to the situation where the state variable (and possibly the observed) is driven by a continuous time process, in which case exact Bayesian inference is typically infeasible as a result of intractable transitions. A stochastic volatility model for financial returns, based on a square root specification for volatility, is used for illustration, with the ABC method seen to produce a very accurate estimate of the exact posterior, which is accessible in this case. Extension to a jump diffusion model for volatility is also explored.

CS45 Room Torrington ADVANCES IN DSGE MODELS Chair: Alexander Meyer-Gohde

C714: **Nominal wage rigidity and forward guidance at the zero lower bound**
Presenter: **Stephane Moyen**, Deutsche Bundesbank, Germany
Co-authors: Xavier Fairise, Michael Krause
In a standard new Keynesian model we assess the role of nominal wage rigidity for monetary policy effectiveness in a liquidity trap. It is well known that the availability of a commitment technology greatly facilitates the stabilisation of output and inflation, relative to the outcomes under discretion. Commitment matters even more at the ZLB, as higher expected future inflation reduces ex ante real interest rates today and thus exerts upward pressure in a slump. By keeping inflation more stable, wage rigidity reduces the probability of deflation and thus makes the ZLB constraint on nominal interest rates less binding. In this sense, wage rigidity reduces the importance of commitment relative to discretionary policy. We show that downward nominal wage rigidity further strengthens this result.

C055: **Monetary policy and the term structure of interest rates**
Presenter: **Martin Kliem**, Deutsche Bundesbank, Germany
Co-authors: Alexander Meyer-Gohde
The influence of monetary policy on the slope and curvature of the yield curve is investigated. Therefore, we analyze a reliable medium DSGE model as commonly used for monetary policy analysis nowadays. In particular, the model is a New-Keyensian model with nominal and real friction. The analysis focuses on the risk introduced by the central bank given a specific monetary policy stance. We show how that an inflation targeting central bank affects the precautionary motives of the agents and therefore affects nominal real risk in the economy. In particular, a stronger reaction to inflation reduces nominal risk, which reduces the precautionary motive, finally, pushes the risk-free rate up towards its deterministic value. Otherwise, increasing the reaction coefficient on the output gap shifts the yield curve downward, increases the slope due to shifting real into nominal risk in the economy.

C068: **Currency risk in currency unions**
Presenter: **Alexander Kriwoluzky**, University of Bonn, Germany
Sovereign yield spreads within currency unions may reflect the risk of outright default. Yet, if exit from the currency union is possible, spreads may also reflect currency risk. We develop a New Keynesian model of a small member country of a currency union, allowing both for default within and exit from the union. Initially, the government runs excessive deficits as a result of which it lacks the resources to service the outstanding debt at given prices. We establish two results. First, the initial policy regime is feasible only if market participants expect a regime change to take place at some point, giving rise to default and currency risk. Second, the macroeconomic implications of both sources of risk differ fundamentally. We also analyze the 2009-2012 Greek crisis, using the model to identify the beliefs of market participants regarding regime change. We find that currency risk accounts for about a quarter of Greek yield spreads.

C025: **Risk-adjusted linear approximation**
Presenter: **Alexander Meyer-Gohde**, Humboldt University Berlin, Germany
The purpose is to develop linear approximations of a canonical DSGE around approximations of the stochastic steady state and ergodic mean. The algorithm makes use only of a standard perturbation output to construct the approximations, which can be used to study properties of risk sensitivity using standard tools from linear approximations. The method is evaluated using Euler equation errors in a model with stochastic volatility and recursive preferences. It is found that it significantly improves the accuracy compared to standard linear approximations along the dimension of the volatility state while demonstrating only marginal gains in the dimension of the capital stock.

CS47 Room Montague VOLATILITY MODELS AND THEIR APPLICATIONS Chair: Yasuhiro Omori

C129: **A dynamic factor stochastic volatility model with leverage effect and its application**
Presenter: **Tsunehiro Ishihara**, Hitotsubashi University, Japan
Co-authors: Yasuhiro Omori
A multivariate factor stochastic volatility model with the factor dynamics and the leverage effect is proposed and applied to the portfolio optimization problem. Efficient Bayesian estimation and simulation methods using Markov chain Monte Carlo and particle filter are proposed. The model and the techniques are illustrated with 33 stock market index of Japan.

C145: **Estimating stochastic volatility models using realized measures**
Presenter: **Jeremias Bekierman**, University of Cologne, Germany
Co-authors: Bastian Gribisch
The basic stochastic volatility (SV) model is extended in order to incorporate the realized variance (RV) as an additional measure for the latent daily volatility. The particular model we use accounts for overnight, microstructure and leverage effects. We estimate the model using simulated maximum likelihood based on Efficient Importance Sampling (EIS), producing numerically accurate parameter estimates and filtered state sequences. The model is applied to daily asset returns and realized variances of New York Stock Exchange (NYSE) traded stocks. To capture the long memory property of asset volatility, we explore different autoregressive dynamics for the latent volatility process, including heterogeneous autoregressive (HAR) dynamics. Estimation results indicate that the model with HAR-dynamics significantly improves the model fit relative to the other proposed dynamics.

C732: **Particle learning for stochastic volatility models with nonlinear leverage effects**
Presenter: **Teruo Nakatsuma**, Keio University, Japan
Co-authors: Kenichiro McAlinn, Hiroaki Katsura
Negative correlation between the previous day's stock return and today's volatility is observed in many stock markets. This relationship is known as the leverage effect in the literature and has been regarded as an indisputable stylized fact among researchers. Recently, however, some studies found that the leverage effect in the ARCH-type model might be nonlinear and the impact of the past return on the present volatility could be better pproximated by a higher-order Hermite polynomial. We attempt to extend this nonlinear leverage approach to the stochastic volatility model, instead of the ARCH-type model. Since it is difficult to estimate a stochastic volatility model with a higher-order Hermite polynomial leverage even in the Bayesian MCMC approach, we apply particle learning for Bayesian estimation of this nonlinear time series model. For stock index returns in the major markets, we found some supportive evidence for the existence of the non-linear leverage effect in the stochastic volatility model.

C874: **Skew exponential power stochastic volatility model for analysis of skewness, heavy tailness, quantiles and expectiles**
Presenter: **Genya Kobayashi**, University of Tokyo, Japan
A stochastic volatility model using skew exponential power (SEP) distribution is considered. SEP distribution has two tail parameters that control left and right tails, and can express flexible tail behaviour. Instead of the original parameterisation of SEP where there are two scale parameters, we introduce a skewness parameter which takes values between 0 and 1, and controls the overall asymmetry of the return distribution. This eases the interpretation of the model, since the scale of the return distribution is already modelled through stochastic volatility. Furthermore, since our model specification includes asymmetric Laplace distribution and asymmetric normal distribution as special cases, our model is related to quantile regression and expectile regression for stochastic volatility models. Therefore, the proposed model provides a unified framework for analysis of return skewness and tail heaviness, quantiles, and expectiles. The model is estimated by a simple Markov chain Monte Carlo method using the uniform scale mixture representation of SEP distribution. The proposed model is demonstrated using simulated datasets and real datasets.

CS86 Room Gordon NEW TRENDS IN TIME SERIES ECONOMETRICS AND SPATIAL ECONOMETRICS Chair: Jean-Yves Pitarakis

C349: **Looking for a trend in distribution characteristics: The case of global warming**
Presenter: **Jesus Gonzalo**, University Carlos III de Madrid, Spain
Co-authors: M. Dolores Gadea
The issue of Global Warming (GW) raises an important chain of questions about: (i) The existence of GW; (ii) Whether GW is caused by human activities or not; (iii) The economic effects of GW; (iv) Economic policies to mitigate the effects of GW, etc. Without a clear answer to the first question, it makes no sense to ask the others. The goal of this paper is to offer a complete answer to the first question. We start by defining GW as an increasing trend in global temperatures. Trend is understood in a much broader sense than in the literature so far. We look for trends in many of the characteristics of the temperature distribution and not only in the average. We estimate them by what we denominate realized quantiles. They form a time series and offer a picture of the historical evolution of the temperature distribution. After estimating a given set of distribution characteristics Ci(t) (based on realized quantiles) we test for the existence of a trend in each of them. If trend is not rejected we model it and produce forecasts of the trending behavior of the whole temperature distribution.

C543: **Indirect inference in spatial autoregression**
Presenter: **Francesca Rossi**, University of Southampton, United Kingdom
Co-authors: Peter C.B. Phillips, Maria Kyriacou
It is well established that the Ordinary Least Squares (OLS) estimator of the spatial parameter in pure Spatial Autoregression (SAR) is inconsistent. The aim is to explore the use of indirect inference estimation. Under broad conditions, indirect inference based on OLS is shown to be consistent and asymptotically normal as the sample size tends to infinity. The estimator has the advantage of being computationally straightforward compared with standard Gaussian Maximum Likelihood (ML). Monte Carlo experiments based on various specifications of the weight matrices confirm that the indirect inference estimator displays little bias even in very small samples. Interestingly, its performance is comparable to that of the Gaussian ML.

C633: **Properties of the maximum likelihood estimator in spatial autoregressive models**
Presenter: **Federico Martellosio**, University of Surrey, United Kingdom
Co-authors: Grant Hillier
The (quasi-) maximum likelihood estimator (MLE) for the autoregressive parameter in a spatial autoregressive model cannot in general be written explicitly in terms of the data. The only known properties of the estimator have hitherto been its first-order asymptotic properties, derived under specific assumptions on the evolution of the spatial weights matrix involved. It is shown that the exact cumulative distribution function of the estimator can, under mild assumptions, be written down explicitly. All exact properties of the MLE can, in principle, be derived from this result. Several examples of theoretical and practical interest are analyzed in detail. These are of interest in their own right, but also illustrate that asymptotic normality can be a spectacularly bad approximation to the distribution of the MLE. As an extreme example, it is shown that the support of the MLE may not be the entire parameter space, whatever the sample size.

C762: **Identifying observed factor models in high dimensional factor models**
Presenter: **Liang Chen**, Universidad Carlos III de Madrid, Spain
Despite their growing popularity, factor models have been often criticized for lack of identification of the factors since they can only be consistently estimated up to a rotation. We try to identify the orthogonal factors estimated using principal components by associating them to a relevant subset of observed variables. We first propose a selection procedure to choose such a subset, and then test the hypothesis that true factors are exact linear combinations of the selected variables. Not only our estimation procedure is able to correctly identify the true observed factors even in the presence of mild measurement errors, but also we show that out test statistics are more general than those previously proposed. The good performance of our method in finite samples and its advantages relative to the other available procedures are confirmed through simulations. Empirical applications include the identification of the underlying risk factors in large datasets of returns of stocks and portfolios as well as interpreting the factors in a large panel of macroeconomic time series. In both cases, it is shown that the underlying factors can be closely approximated by a few observed variables.

CS64 Room Woburn CREDIT RISK Chair: Simona Sanfelici

C083: **Counterparty risk and funding costs in a multiple-curve framework**
Presenter: **Andrea Pallavicini**, Banca IMI, Italy
The market practice of extrapolating different term structures from different instruments lacks a rigorous justification in terms of cash flows structure and market observables. We derive a multiple term structure dynamics starting from a consistent theory for pricing under credit, collateral and funding risks, by considering realistic assumptions on collateralization, wrong-way risk, gap risk, credit valuation adjustments and funding effects, including the treasury operational model. A collateralized version of the HJM framework is presented and applied to price derivatives in the presence of different collateralization policies.

C338: **Capturing volatility in correlated default risk**
Presenter: **Pawel Janus**, UBS, Switzerland
A novel approach to modeling conditional loss distribution for global credit portfolio is presented. The major modeling determinant comes from an observation that defaults of single names will cluster if there are common factors that affect individual default risk. To model correlated default risk we propose a Merton-type model with asset value changes modeled as a linear multifactor model. Statistical factors are extracted from global equity universe. Factor loadings are homogenized for region-industry-cap buckets leading to a parsimonious correlation structure which determines likelihood of joint defaults or downgrades. The heterogeneity of default thresholds is captured by credit ratings. The key novelty is the stochastic correlation introduced by means of the singular Wishart distribution imposed on systematic covariation. The scale matrix of the Wishart density governs the level and volatility of correlated default risk, and can be straightforwardly extracted from equity market with any desired estimation frequency. The proposed modeling framework with stochastic correlated default risk facilitates a thorough impact study of credit cycles on loss distribution which is a key edge over standard factor models. The merits of the new model are discussed in the simulation study. The empirical illustration is based on a global issuer risk portfolio with broad coverage.

C724: **Market-driven securitizations**
Presenter: **Damiano Bruno Silipo**, University of Calabria, Italy
Co-authors: Luca Giordano, Leone Leonida
The literature has focused on four main motivations for securitization. Banks are believed to securitize loans for reasons associated to liquidity and funding - essentially because of shortage or high costs of retail deposits. In addition to this, they are known to use securitization to transfer credit risk to the market, to adjust their capital ratio in order to meet economic and regulatory capital requirements and, finally, to resort to loan securitization to improve profitability. All the above mentioned incentives for securitization the literature has been focusing on are balance-sheet related. The aim is to test the hypothesis that stock market conditions play an important role in securitization incentives, and banks recur to securitization to grasp new profits opportunities provided by favorable conditions in the stock market. Results based on a large sample of banks over the period 1999-2011 suggest that the higher the return on the stock price of the bank and the market index, the higher the amount of securitization. Hence, the paper adds to the literature a new motivation for securitization.

C693: **A boundary element PDE approach to corporate debt**
Presenter: **Chiara Guardasoni**, University of Parma, Italy
Co-authors: Simona Sanfelici
A Boundary Element Method (BEM) is proposed for the analytic PDE-based evaluation of a defaultable zero-coupon bond in a two-factor first-passage-time model. The default triggering barrier and the recovery payoff associated with the default event can be specified in a large variety of ways according to the real-life bond covenants. The advantages of BEM, when compared to domain methods, such as Finite Element Methods or Finite Difference Methods, are well known: only the boundary of the domain needs to be discretized; particularly, exterior problems with unbounded domains but bounded boundaries are handled as easily as interior problems; the solution in the interior of the domain is approximated with a rather high convergence rate and can be evaluated in particular points of the domain and not necessarily everywhere; infinity conditions

are implicitly satisfied. The method is based on the explicit knowledge of a fundamental solution for the pricing PDE. However, when only the characteristic function of the underlying process is known, it is possible to recover the transition density, i.e. the fundamental solution, by Fourier inverse transform.

CS101 Room 349 COMPUTATIONAL ECONOMETRICS II Chair: Matthew Dixon

C910: **Particle efficient importance sampling**
Presenter: **Marcel Scharth**, University of New South Wales, Australia
Co-authors: Robert Kohn
The efficient importance sampling (EIS) method is a general principle for the numerical evaluation of high-dimensional integrals that uses the sequential structure of target integrands to build variance minimising importance samplers. Despite a number of successful applications in high dimensions, it is well known that importance sampling strategies are subject to an exponential growth in variance as the dimension of the integration increases. We solve this problem by recognising that the EIS framework has an offline sequential Monte Carlo interpretation. The particle EIS method is based on non-standard resampling weights that take into account the look-ahead construction of the importance sampler. We apply the method for a range of univariate and bivariate stochastic volatility specifications. We also develop a new application of the EIS approach to state space models with Student's t state innovations. Our results show that the particle EIS method strongly outperforms both the standard EIS method and particle filters for likelihood evaluation in high dimensions. Moreover, the ratio between the variances of the particle EIS and particle filter methods remains stable as the time series dimension increases. We illustrate the efficiency of the method for Bayesian inference using the particle marginal Metropolis-Hastings and importance sampling squared algorithms.

C922: **Estimating the Markov-switching GARCH model with a deterministic particle filter**
Presenter: **Maciej Augustyniak**, University of Montreal, Canada
Co-authors: Mathieu Boudreault, Manuel Morales
The Markov-switching GARCH model allows for a GARCH structure with time-varying parameters. Its introduction in the 1990s was motivated by the inability of the standard GARCH model to reflect structural changes in financial time series which are usually found to occur over many years of data. The combination of Markov-switching and GARCH processes complicates estimation as exact computation of the log-likelihood is infeasible due to a path dependence problem. This difficulty led to computationally intensive estimation methods and to simpler techniques based on an approximation of the model, known as collapsing procedures. An original framework is developed to conduct maximum likelihood inference in the Markov-switching GARCH model, generalizing and improving previously proposed collapsing approaches. This method is demonstrated to correspond to a deterministic particle filter, establishing a novel relationship between particle filtering and collapsing procedures. The discovery of this missing link justifies the validity of the proposed estimation framework. Simulation and empirical studies show that it allows for a fast and accurate estimation of the Markov-switching GARCH model.

C1074: **A class-free information criteria**
Presenter: **Sylvain Barde**, University of Kent, United Kingdom
A class-free information criterion (CIC) is developed that is analogous to the Akaike information criterion (AIC) in its theoretical derivation and yet can be applied to any model able to generate simulated or predicted data. Both the AIC and the proposed CIC rely on the Kullback-Leibler (KL) distance between the prediction and the real data as a measure of prediction accuracy. In the case of the AIC, this is estimated via the maximum likelihood approach. The proposed CIC relies instead on the literal interpretation of the KL distance as the inefficiency of compressing real data using modeled probabilities, and therefore uses the output of a compression algorithm to obtain an estimate of the KL distance. In practice, the algorithm is first trained on the predicted data from the model before being used to compress the real data, the efficiency of the compression providing the measure of the model's accuracy. Two Monte Carlo tests are provided as an illustration, one on an ARMA process with ARCH errors and one on an agent-based model with recruitment. In both cases the CIC is able to recover the true model from a set of alternatives, which confirms the good performance of the criterion.

C1193: **Strategies for recursively estimating the simultaneous equations model**
Presenter: **Stella Hadjiantoni**, Queen Mary University of London, United Kingdom
Co-authors: Erricos Kontoghiorghes
The recursive estimation of the simultaneous equations model (SEM) has been investigated. A new method to derive the three stage least squares (3SLS) estimator recursively when an out-of-core algorithm needs to be developed or when new data become available is presented. The SEM to be estimated has been reformulated to an equivalent model where the initial model is updated so that the new observations are added and with endogeneity being eliminated. Previous computations are efficiently utilized. It has been proved that the generalized least squares solution of the proposed model gives the 3SLS estimator of the SEM when it is estimated from scratch. The proposed model has been reformulated as a generalized linear least squares problem for the development of computationally efficient algorithms. The main computational tool is the generalized QR decomposition which is employed based on hyperbolic Householder transformations. The special structure of the matrices and properties of the SEM are exploited in order to reduce the computational burden of the estimation algorithm. In addition, the computation of the iterative 3SLS estimator of the updated observations SEM has been considered. The method is also extended to the downdating problem of deleting observations from the SEM after the 3SLS estimator has been computed.

C1226: **Equity indexes analysis and synthesis by using wavelet transforms**
Presenter: **Andrejs Puckovs**, Riga Technical University, Latvia
Light to equity indexes analysis is given using wavelet transforms (Direct Continuous Wavelet transform and Inverse Continuous Wavelet transform as well as Direct Discrete Wavelet Transform and Inverse Discrete Wavelet Transform). Direct Continuous Wavelet Transform has been used as a tool to describe some properties of equity indexes such as Hurts exponent, wavelet coefficients probability distribution, pseudo frequencies on a certain scaling parameters. Most of them are essential to find significant changes in equity index behavior and predict instability. Some variations of Fractal Brownian Motion are proposed in current article; as a result the stochastic process with "dynamic" Hurts exponent is defined. Some algorithms for process simulation by using Direct and Inverse Continuous Wavelet Transforms via "Morlet" mother wavelet function are provided within article. Proposed variations of Fractal Brownian Motion are beneficial for equity index simulations with variant return indicators on various investment horizons.

Sunday 15.12.2013 14:40 - 16:20 Parallel Session L – ERCIM

ES19 Room B33 BIOSTATISTICS AND BIOINFORMATICS Chair: Maria De Iorio

E653: **Bayesian models for cost-effectiveness analysis in the presence of structural zeros**
Presenter: **Gianluca Baio**, University College London, United Kingdom
Bayesian modelling for cost-effectiveness data has received much attention in both the health economics and the statistical literature in recent years. Cost-effectiveness data are characterised by a relatively complex structure of relationships linking the suitable measure of clinical benefit (QALYs) and the associated costs. Simplifying assumptions, such as (bivariate) normality of the underlying distributions are usually not granted, particularly for the cost variable, which is characterised by markedly skewed distributions. In addition, individual-level datasets are often characterised by the presence of structural zeros in the cost variable. Hurdle models can be used to account for the presence of excess zeros in a distribution and have been applied in the context of cost data. Their application is extended to cost- effectiveness data, defining a full Bayesian model which consists of a selection model for the subjects with null costs, a marginal model for the costs and a conditional model for the measure of effectiveness (conditionally on the observed costs). The model is presented using a working example to describe its main features.

E1001: **Bayesian deconvolution and quantitation of NMR metabolic profiles with BATMAN**
Presenter: **Timothy Ebbels**, Imperial College London, United Kingdom
Metabolomics is the study of metabolites - the small molecules which provide the energy and building blocks of life. Levels of metabolites in biofluids and tissues are a sensitive barometer of biological status, and can reveal a great deal about the health and disease. Nuclear Magnetic Resonance (NMR) spectroscopy is widely used in metabolomics to assay the metabolite levels and produces complex spectra, typically consisting of hundreds of overlapped peaks. Peak annotation, deconvolution and quantitation are key bottlenecks in the current metabolomics pipeline and thus key targets for new computational approaches. Current solutions either require intensive manual effort, or simply defer these problems to later stages of the workflow. To tackle this, we developed BATMAN – the Bayesian AuTomated Metabolite Analyser for NMR spectra which decomposes spectra of complex mixtures into their constituent pure components. For known metabolites, BATMAN uses resonance patterns from the Human Metabolome Database. This extensive prior information also allows us to deconvolve overlapped resonances and obtain concentration estimates for the corresponding metabolites. Spectral signals without prior information are modelled using wavelets. In performance tests, BATMAN enabled semi-automated quantification of metabolites from biofluids with similar accuracy to current manual approaches, thus alleviating one of the major bottlenecks in NMR metabolomics.

E1169: **A Bayesian model for population pharmacogenetics**
Presenter: **Gary Rosner**, Johns Hopkins University, United States
Co-authors: Masanao Yajima, Donatello Telesca, Peter Mueller
A general framework is proposed for the exploration of interactions between population pharmacokinetics (PK) and inherited genetic configurations (Gx) in a coherent Bayesian population pharmacogenetics model (PKGx). Complex dependence patterns are represented using the Markov structure of chain graphs. Model choice and computation of model parameters are based on MCMC simulation. We discuss posterior screening of significant interactions through a decision-theoretic argument and apply the proposed method to the study of a chemotherapeutic agent commonly used in the treatment of solid tumors.

E1237: **Subgroup reporting using nonparametric Bayesian inference**
Presenter: **Peter Mueller**, UT Austin, United States
Bayesian inference for subgroups in clinical trials is discussed. The key feature of the proposed approach is that we separate the decision problem of selecting subgroups for reporting and the probability model for prediction. For the latter we use a flexible nonparmetric Bayesian model, while the decision problem is based on a parsimonious description of the subgroups and a stylized utility function.

ES21 Room B35 ADVANCES IN LATENT VARIABLE EXTRACTION FROM MULTI-BLOCK DATA STRUCTURES Chair: Laura Trinchera

E256: **A heuristic procedure for finite structural equation mixture models**
Presenter: **George Marcoulides**, University of California at Riverside, United States
Researchers often face the difficult task of determining whether considered data have been sampled from a single population or from a finite set of homogeneous populations. Mixture modeling is one popular technique that can be used to address this task. A new heuristic procedure for the analysis of finite mixture models is introduced. Results show that the new procedure can help in the appropriate analysis and interpretation of sampled data.

E302: **A multiway discriminant analysis: STATIS-LDA**
Presenter: **Robert Sabatier**, Montpellier University, France
Co-authors: Christelle Reynes
Most of multiway methods consist in an unsupervised analysis which makes their utility quite limited in data analysis. In particular, supervised classification is a very frequent task in data analysis, especially in the case of meta-analyses that should be processed through multiway methods. Obviously, identical groups through all tables are considered. Yet, among multiway methods proposed in literature, a few of them deal with discrimination (excluding some analyses whose applicability is limited to low dimensional tensors), hence, those methods are too rarely used in chemometrics, image analysis... STATIS-LDA is a multiway discriminant analysis, which provides a methodology allowing us to perform supervised classification by proposing a new generalization of STATIS procedure. The used criterion can be seen as a weighted sum of the criteria of usual discriminant analyses of each table. Especially, this method allows us to select the most interesting tables with regards to discrimination. Some applications on simulated and real datasets (including high dimensional ones) will be proposed in order to demonstrate its simple use and to introduce the associated interpretation tools.

E470: **Supervised component generalized linear regression**
Presenter: **Xavier Bry**, Montpellier, France
Co-authors: Catherine Trottier, Frederic Mortier
The problem of component-based regularization of a multivariate generalized linear model (MGLM) is addressed. In the current estimation of a MGLM, the correlation structure of regressors is not used as the basis on which to lean strong predictive dimensions. Looking for linear combinations of regressors that merely maximize the likelihood of the MGLM has two major consequences: 1) collinearity of regressors is a factor of estimation instability, and 2) as predictive dimensions may lean on noise, both predictive and explanatory powers of the model are jeopardized. For a single dependent variable, attempts have been made to adapt PLS Regression, which solves this problem in the classical linear model, to GLM estimation. A technique is proposed: Supervised Component Generalized Linear Regression (SCGLR), that extends the Fisher scoring algorithm so as to combine PLS regression with MGLM estimation. SCGLR is designed to deal with mixtures of numeric and nominal regressors. It is

eventually extended so as to take an offset into account, for count data models such as Poisson or Binomial. It is tested on both simulated and real data.

E763: **Beyond the measurement scale: The non-metric partial least squares approach**
Presenter: **Giorgio Russolillo**, CNAM, France
The Partial Least Squares approach to latent variable path modeling (PLS-PM) is a component-based method that allows investigating the relationships among several sets of variables connected by a path. This method is implemented through a flexible PLS algorithm that, depending on the chosen option, can optimize covariance or correlation based criteria. PLS-PM only applies to quantitative (metric) data, although in many real applications users are interested in analyzing data observed on ordinal or nominal measurement scales. To handle measurement scale heterogeneity in PLS framework, the Non-Metric PLS (NM-PLS) approach has been recently proposed. It extends PLS methods optimizing covariance-based criteria to the treatment of non-metric variables and non-linearity. NM-PLS methods are implemented through PLS-type algorithms that work as optimal scaling procedure: each different value or category is considered as a scaling parameter, which is estimated through a numerical value. NM-PLS algorithms estimate both model and scaling parameters in order to satisfy PLS model criteria. The main features and properties of the NM-PLS algorithms are presented. The standard NM-PLS is extended to PLS criteria related to correlation. Finally, the effects of the parametric expansion on the variability of the estimates in the NM-PLS approach to multi-block analysis are discussed.

ES23 Room G16 FUZZY CLUSTERING Chair: M. Brigida Ferraro

E248: **Robust constrained fuzzy clustering**
Presenter: **Luis Angel Garcia-Escudero**, Universidad de Valladolid, Spain
Co-authors: Heinrich Fritz, Agustin Mayo-Iscar
It is known that outliers and noise can be very harmful when applying (fuzzy) clustering methods. This is the reason why several methods have been proposed in the fuzzy clustering literature trying to address robustness issues. A new robust fuzzy clustering approach called F-TCLUST will be presented. The approach is based on trimming a fixed proportion of observations self-determined by the data set. The proposed approach also considers an eigenvalue ratio constraint that makes it a mathematically well-defined problem and serves to control the allowed differences among cluster scatters. A computationally feasible algorithm is proposed for its implementation in practice. Some guidelines are given about how to choose the parameters involved in this robust fuzzy clustering approach. A comparison with other robust fuzzy clustering approaches will be presented.

E400: **Fuzzy clustering induced correlation for feature extraction in lower dimension spaces**
Presenter: **Mika Ilic**, University of Tsukuba, Japan
The idea of explaining similarity of objects by exploiting the classification structures has been used in many methodologies such as weighted scaling. Fuzzy clustering is one weighted scaling, in which the weight is explained as a fuzzy grade which shows degree of belongingness of an object to a cluster. This degree usually satisfies conditions such as sum of the degrees of an object with respect to clusters being 1 and the values of degree being in a range from 0 to 1. Using these conditions, a new similarity of objects is defined when some conditions satisfy the relationship between similarity of fuzzy classification structures and similarity of objects. In addition, a correlation based on fuzzy classification structures is defined, which is formally shown by a sum of correlation of objects and correlation of fuzzy classification structures; however, it measures similarity of two kinds of correlations and dissimilarity of objects included weights as dissimilarity of classification structures. This correlation is useful for the analysis of high dimension and small sample data to obtain the features of data in a lower dimensional space. Several numerical examples are shown to show a better performance of the proposed methods.

E1021: **Multi-sample fuzzy clustering approach for random fuzzy sets**
Presenter: **Ana Belen Ramos-Guajardo**, University of Oviedo, Spain
Co-authors: Paolo Giordani
A fuzzy clustering method for grouping random fuzzy sets is proposed. The procedure is based on the computation of a $p-$value matrix. The $p-$values are obtained by comparing the expected values of the random fuzzy sets by means of a multi-sample bootstrap test already proposed in the literature. In fact, the $p-$value matrix can be viewed as a relational data matrix since the $p-$values represents a kind of dissimilarity (or similarity) between the expected values. For this reason, fuzzy clustering techniques for relational data can be applied. In this context, a generalization of the FANNY algorithm, called fuzzy relational data clustering (FRC) algorithm, is considered. One of the most important advantages of the FRC with respect to other procedures is that the relational data could not be derived from Euclidean distance as the measure of dissimilarity between the objects. Some simulations are presented to show the behaviour of the procedure and an application to a real-life situation is also included.

E727: **Ensemble detection in parallel spike train data using protoclusters**
Presenter: **Christian Borgelt**, European Centre for Soft Computing, Spain
Co-authors: Christian Braune, Christoph Doell, Rudolf Kruse
In the context of point processes we investigate a method inspired by the DBSCAN algorithm for detecting related subgroups in a set of point processes, especially neural spike trains. Such spike trains capture the firing of neurons, and synchronous firing may indicate coordinated processing of information. Previous work showed how to define a metric on point process data and how to calculate a mean/average process, thus making clustering approaches applicable. One of the major remaining problems of identifying groups of neurons exhibiting synchronous activity is that – in contrast to conventional clustering – they may overlap on an arbitrary number of elements. To identify such groups we first detect so-called protoclusters (similar to DBSCAN's ε-neighbourhood) by an iterative procedure that finds small groups of highly similar spike trains and removes them from the data set. The detected protoclusters are then organized into a graph structure. Cliques of this graph form the actual clusters and allow us to cope with overlapping clusters as well as to identify conventional clusters with high accuracy. Modelling spike trains as e.g. Poisson processes enables us to evaluate our method on large artificial data sets containing several interesting characteristics.

ES34 Room B18 CLUSTERING METHODS FOR FUNCTIONAL DATA Chair: Aurea Grane

E040: **Model-based clustering for multivariate functional data**
Presenter: **Julien Jacques**, University Lille I and INRIA, France
Co-authors: Cristian Preda
The first model-based clustering algorithm for multivariate functional data is proposed. After introducing multivariate functional principal components analysis, a parametric mixture model, based on the assumption of normality of the principal component scores, is defined and estimated by an EM-like algorithm. The main advantage of the proposed model is its ability to take into account the dependence among curves. Results on real datasets illustrate the efficiency of the proposed method.

E072: **Quantization and clustering with Bregman divergences**
Presenter: **Aurelie Fischer**, Universite Paris Diderot, France

The problem of quantization of a random variable X taking values in a separable and reflexive Banach space, and the related question of clustering independent random observations distributed as X, will be addressed. To this end, we use a quantization scheme with a class of distortion measures, called Bregman divergences, and provide conditions ensuring the existence of an optimal quantizer and an empirically optimal quantizer. We will also discuss rates of convergence.

E166: **Sparse clustering of functional data**
Presenter: **Valeria Vitelli**, University of Oslo, Norway
Co-authors: Davide Floriello, Piercesare Secchi

When faced with a clustering problem, we do not expect the real underlying groups to differ in all features considered. Most likely only a limited number of variables are relevant to detect differences among groups. Moreover, this problem further complicates as much as the number of features p is larger than the sample size n. For this reason many statistical methods, denominated 'sparse', have been proposed, aimed at clustering data while jointly selecting the most relevant features for clustering. Our aim is to consider the same problem in a functional data context, where sparsity means selecting Borel subsets of the domain where the clusters distinguish the most. First, the sparse multivariate clustering problem, formulated via a hard thresholding strategy, is proven to have unique solution. Then, the problem of classifying functional data and jointly selecting relevant data features is analytically defined as a constrained optimization of a functional, over the set of possible cluster partition and over a set of admissible weighting functions, responsible for feature selection. An implementing algorithm is also proposed. Improvements on standard functional clustering techniques are discussed in the light of simulation studies. Finally, an application to the Berkeley Growth Study is also presented.

E456: **Supervised classification for Gaussian processes**
Presenter: **Jose Berrendero**, Universidad Autonoma de Madrid, Spain
Co-authors: Antonio Cuevas, Jose Luis Torrecilla

Given training samples of trajectories corresponding to two different Gaussian processes, the problem of classifying a new trajectory whose membership is unknown in one of the two models is considered. When both Gaussian measures are equivalent, the optimal Bayes classification rule can be obtained from the Radon-Nikodym derivative of one measure with respect to the other. Classic results (due to Cameron, Martin and Shepp, among others) provide expressions for the Radon-Nikodym derivatives of measures equivalent to Wiener measure. From these expressions it is then possible to derive the optimal Bayes rules. In some cases, an expression for the corresponding Bayes error can also be obtained. Although theoretically insightful, the optimal rules are unfeasible because they depend on the mean function and covariance structure of the processes, which are usually unknown. It is shown how to build feasible rules by replacing the unknown elements in the optimal classifier with suitable estimators computed from the training samples.

ES37 Room B29 ALGORITHMS FOR ROBUST STATISTICAL PROCEDURES Chair: Mia Hubert

E374: **Robust maximum association estimators and an extension for sparsity**
Presenter: **Andreas Alfons**, Erasmus University Rotterdam, Netherlands
Co-authors: Christophe Croux, Peter Filzmoser

The maximum association between two multivariate random variables is defined as the maximal value that a bivariate association measure between respective one-dimensional projections of each random variable can attain. It is proposed to use the Spearman or Kendall rank correlation as projection index as more robust alternatives to the Pearson correlation. Theoretical and numerical results show that maximum rank correlation estimators are favorable due to good efficiency and robustness properties. Concerning computation, a projection pursuit algorithm is proposed based on alternating series of grid searches in two-dimensional subspaces of each data set, and present a fast implementation of the algorithm for the statistical computing environment R. In addition, an extension of the algorithm is presented, that allows for sparse estimation of the projection directions to increase the interpretability of the results in higher dimensions.

E497: **The forward search for very large datasets**
Presenter: **Domenico Perrotta**, European Commission - JRC, Italy
Co-authors: Marco Riani, Andrea Cerioli

The identification of atypical observations and the immunization of data analysis against both outliers and failures of modelling are important aspects of modern statistics. The forward search is a graphics rich approach that leads to the formal detection of outliers and to the detection of model inadequacy combined with suggestions for model enhancement. The key idea is to monitor quantities of interest, such as parameter estimates and test statistics, as the model is fitted to data subsets of increasing size. Some computational improvements of the forward search algorithm are proposed and a recursive implementation of the procedure is provided which exploits the information of the previous step. The output is a set of efficient routines for fast updating of the model parameter estimates, which do not require any data sorting, and fast computation of likelihood contributions, which do not require any inverse matrix or QR decomposition. It is shown that the new algorithms enable a reduction of the computation time by more than 80%. Furthemore, the running time now increases almost linearly with the sample size. All the routines presented are included in the FSDA toolbox for MATLAB, which is freely downloadable from Internet.

E530: **A robust procedure for shift detection in time series**
Presenter: **Roland Fried**, TU Dortmund University, Germany
Co-authors: Herold Dehling, Martin Wendler

A robust nonparametric test for a shift in a time series of weakly dependent observations is proposed. The test is a robust alternative to the classical CUSUM test. It is based on the median of all pairwise differences, maximized over all splits of the data after any time point. Proper standardization of the test statistics is crucial: For both tests, one needs to estimate the corresponding asymptotical variance, which is an infinite sum of covariances at all time lags. For the proposed test, one needs to estimate additionally the density of the difference between an independent pair of observations. The experience with different estimators of these quantities is discussed. The performance of the tests is compared via simulations, which indicate that the new test offers higher robustness and better power in the presence of outliers and in case of skewed and heavy tailed distributions. Applications to real data sets are provided as well.

E654: **A fast algorithm for robust SUR**
Presenter: **Tim Verdonck**, KU Leuven, Belgium
Co-authors: Mia Hubert, Ozlem Yorulmaz

The seemingly unrelated regression (SUR) model is a generalization of a linear regression model consisting of more than one equation, where the disturbance terms of these equations are contemporaneously correlated. Ignoring this contemporaneous correlation and estimating these equations separately leads to inefficient results. SUR estimates all equations simultaneously and takes into account the covariance structure of the residuals. Since the traditional estimator is very sensitive to outliers, a robust version using S-estimators has already been proposed in literature. S-estimators

are very robust, affine equivariant and asymptotically normal, but are also difficult to calculate. Therefore a fast algorithm is presented, denoted FastSUR, to compute robust SUR models and its good performance is shown in a simulation study. SUR models are often used in statistics and econometrics. The focus is on the General Multivariate Chain Ladder (GMCL) model that employs SUR to estimate its parameters. Consequently, this multivariate stochastic reserving method takes into account the contemporaneous correlations among run-off triangles and allows structural connections between these triangles. The FastSUR algorithm is plugged in into the GMCL model to obtain a robust version.

ES41 Room G15 MATRIX ALGORITHMS AND HPC FOR LARGE SCALE DATA ANALYSIS I **Chair: Costas Bekas**

E822: **Leveraging as a paradigm for statistically informed large-scale computation**
Presenter: **Michael Mahoney**, Stanford, United States
Statistical leverage has historically been used in regression diagnostics to flag outliers, but recently it has emerged as a design principle to obtain faster (both in theory and in practice) algorithms for large-scale matrix and regression problems. Interestingly, these algorithms work for arbitrary, i.e., worst-case, input, but implicitly they use traditional statistical ideas. As an example, by approximating or preconditioning to be uniform the leverage scores of an arbitrary tall matrix, we can obtain algorithms for very overconstrained least-squares approximation that beat Lapack subroutines in terms of clock time for arbitrary matrices of size as small as thousands by hundreds. This approach of using the empirical statistical structure of the input to obtain better algorithms for arbitrary large-scale problems, as well as statistical questions raised by this paradigm, will be discussed.

E794: **Fast SDD solvers via sampling by approximate leverage scores**
Presenter: **Ioannis Koutis**, University of Puerto Rico-Rio Piedras, United States
Recent progress in algorithms for solving symmetric diagonally dominant (SDD) linear systems is reviewed. This type of systems can now be solved in time near-proportional to the number of non-zero entries in the system matrix, thus enabling the solution of very large systems. The new solvers are based on fast computations of approximate leverage scores for the rows of the underlying incidence (tall-and-thin) matrix. We also discuss the challenges of extending the approach to solvers for other types of linear systems and least squares problems.

E833: **Using ILU to estimate the diagonal of the inverse of a matrix**
Presenter: **Andreas Stathopoulos**, College of William and Mary, United States
Co-authors: Lingfei Wu, Jesse Laeuchli, Vasilis Kalantzis, Stratis Gallopoulos
For very large, sparse matrices, the calculation of the diagonal or even the trace of their inverse is a computationally intensive task that involves the use of Monte Carlo (MC). Hundreds or thousands of quadratures ($x^TA^{-1}x$) are averaged to obtain 1-2 digits of accuracy in the trace. An ILU preconditioner can speed up the computation of the quadratures, but it can also be used to reduce the MC variance (e.g., by estimating the diagonal of $A^{-1}-(LU)^{-1}$). Such variance reduction may not be sufficient, however. We take a different approach based on two observations. First, there is an old, but not much used algorithm to obtain the diagonal of $M=(LU)^{-1}$ when L,U are sparse, in time linear to the number of nonzero elements of L,U. For some matrices, the factors from the ILU preconditioner provide a surprisingly accurate estimate for the trace of A^{-1}. Second, we observed that, in general, the diag(M) captures the pattern (if not the individual values) of the diag(A^{-1}) very well. Therefore, if we solved for the exact $A^{-1}_{jj}=e_j^TA^{-1}e_j$, for $j=1,\ldots,k$ anchor points, we could find an interpolating function $f(M_{jj},j)\approx A^{-1}_{jj}$, for all j. Preliminary experiments have shown that $k=100$ is often enough to achieve $O(10^{-2})$ accuracy in the trace estimation, much faster than classical MC.

E869: **Accelerating parallel data uncertainty quatification by solving linear systems with multiple right-hand sides**
Presenter: **Costas Bekas**, IBM Research, Greece
Co-authors: Efstratios Gallopoulos, Vasileios Kalantzis, Alessandro Curioni
Measuring the uncertainties in data collections is a useful preprocessing step for many tasks in data analysis. Much information can be derived from the elements of the inverse of the sample covariance matrix (precision matrix), especially those along the diagonal. We are interested in estimating a large number or all of the diagonal elements. Traditional methods for computing the diagonal elements have cubic cost. As an alternative, approximations of the diagonal via stochastic estimation has been proposed. Stochastic estimation depends on the rapid multiplication of the precision matrix with a set of carefully selected vectors. Some of us managed to reduce previously the cubic complexity by using iterative methods and massive parallelism. More recently, we exploited certain properties of the problem and developed a numerical method based on block Krylov methods in conjunction with projection techniques for solving linear systems with multiple right-hand sides. We focus now on parallel implementation for the methods presented therein. What we propose to show is that our algorithms succeed not only in accelerating the convergence of the kernel iterative solver but are scalable and fast on modern architectures. We will demonstrate with numerical experiments that in this manner it becomes practical to perform such uncertainty quantification on problems of very large scale. We will also show that the proposed techniques can also be used in network analysis.

ES48 Room B34 GRAPHICAL MODELING **Chair: Giovanni Marchetti**

E351: **MCMC strategies for the analysis of discrete graphical models of marginal independence**
Presenter: **Claudia Tarantola**, University of Pavia, Italy
Co-authors: Ioannis Ntzoufras
Different MCMC strategies for the analysis of marginal log-linear models with a bi-directed graph representation are presented. Two alternative parameterisations are considered: one corresponding to cell-probabilities (pi-parameterisation), and the other corresponding to the marginal log-linear representation (lambda-parameterisation). Working with the pi-parameterisation it is possible to perform a conjugate and conditional conjugate analysis based on product of Dirichlet priors. On the other hand, the lambda-parameterisation a conjugate analysis is not feasible. In this case two approaches are presented and compared. The first approach consists of Metropolis-Hastings algorithms with proposals based on the pi-parameterisation. The corresponding proposal distribution on the marginal log-linear parameters is then calculated using the determinant of Jacobian matrix of the function linking the two parameterisations. The second approach is based on directly simulating the parameters of each marginal association log-linear parameters. Unfortunately, at each iteration of the MCMC sampler it is necessary to implement iterative methods to calculate the cell probabilities and thus the model likelihood, and this will slow down the speed of the algorithm.

E1178: **Traceable regressions: Their graphs, structures and implications**
Presenter: **Nanny Wermuth**, Chalmers University of Technology, Sweden
Assumptions and advantages of using graphs of traceable regressions are stressed and illustrated. In contrast to the formulation of traditional models for many variables, the regression graphs for traceable regression permit to derive important consequences of a given dependence structure and to heck these against knowledge available in a substantive context.

E610: **Log-mean linear parameterization of discrete multivariate regression chain graph models**
Presenter: **Alberto Roverato**, Universty of Bologna, Italy
Co-authors: Monia Lupparelli
Chain graph models are families of models that represent dependencies in systems where the variables are grouped in blocks as responses, intermediate responses and purely explanatory variables. Four different interpretations of a chain graph as independence model are available, and the focus is on the so-called Multivariate Regression Chain Graphs (MRCGs). This family of models allows one to represent the dependence of each response marginally on covariates but also, more generally, to model the joint dependence of arbitrary subsets of responses on covariates. Alternative parameterizations of MRCGs for discrete variables are available. These are equivalent for the specification of the independence model, but differ with respect to the kind of interaction among responses they represent. Consequently, also the values of the regression coefficients depend on the parameterization adopted, and their interpretation is typically non-straightforward. The aim is to extend the Log-Mean Linear (LML) parameterization previously introduced for discrete graphical models of marginal independence to MRCGs and show that this generalization preserves the appealing features of the LML parameterization. Furthermore, in this approach, the regression coefficients have a straightforward interpretation in terms of relative risks both for regression corresponding to single responses and for regression corresponding the interactions among responses.

ES56 Room B36 NON-GAUSSIAN MIXTURE MODEL-BASED CLASSIFICATION Chair: Paul McNicholas

E043: **Flexible mixture modelling using the multivariate skew-Student's $t-$normal distribution**
Presenter: **Tsung-I Lin**, National Chung Hsing University, Taiwan
Co-authors: Hsiu J. Ho, Chia-Rong Lee
The aim is to present a robust probabilistic mixture model based on the multivariate skew-Student's $t-$normal distribution, a skew extension of the multivariate Student's t distribution with more powerful abilities in modelling data whose distribution seriously deviates from normality. The proposed model includes mixtures of normal, t and skew-normal distributions as special cases and provides a flexible alternative to recently proposed skew t mixtures. We develop two analytically tractable EM-type algorithms for computing maximum likelihood estimates of model parameters in which the skewness parameters and degrees of freedom are asymptotically uncorrelated. We also present a procedure of merging mixture components to automatically identify the number of clusters by fitting piecewise linear regression to the rescaled entropy plot. The effectiveness and performance of the proposed methodology are illustrated by two real examples.

E328: **On mixtures of skew factor models**
Presenter: **Geoffrey McLachlan**, University of Queensland, Australia
Co-authors: Sharon Lee
Finite mixtures of distributions are being increasingly used in statistical analyses, in order to classify and cluster data. In many applications with modern day data sets, the dimension of the observation vector can be quite large, so that normal mixture models can be highly parameterized. One way of proceeding with the application of normal mixtures to high-dimensional data sets is to adopt mixtures of factor models. Here we extend the use of such mixtures by allowing the factors and the observations to have non-normal distributions with the focus on the use of skew normal and t-distributions. Examples are given in the context of the supervised classification of data.

E546: **Mixtures of common skew-t factor analyzers**
Presenter: **Paula Murray**, University of Guelph, Canada
Co-authors: Ryan Browne, Paul McNicholas
Building on their symmetric analogues, mixtures of common skew-t factor analyzers are introduced for model-based clustering of high-dimensional data. Parameter estimation is carried out using the alternating expectation-conditional maximization algorithm, and the Bayesian information criterion is used for model selection. The mixtures of common skew-t factor analyzers model is applied to simulated data as well as real gene expression microarray data. It concludes with a discussion of the results as well as suggestions for future work.

E1007: **A mixture of generalized hyperbolic distributions**
Presenter: **Ryan Browne**, University of Guelph, Canada
Co-authors: Paul McNicholas
A mixture of generalized hyperbolic distributions is introduced as an alternative to the ubiquitous mixture of Gaussian distributions as well as their near relatives of which the mixture of multivariate t and skew-t distributions are predominant. The mathematical development of our mixture of generalized hyperbolic distributions model relies on its relationship with the generalized inverse Gaussian distribution. The latter is reviewed before our mixture models are presented along with details of the aforesaid reliance. Parameter estimation is outlined within the expectation-maximization framework before the performance of our mixture models is illustrated in clustering applications on simulated and real data. In particular, the ability of our models to recover parameters for data from underlying Gaussian, and skew$-t$ distributions is demonstrated. Finally, the role of Generalized hyperbolic mixtures as a superclass as well as the anticipated impact of these models on the model-based clustering, classification and density estimation literature is discussed with special focus on the role of Gaussian mixtures.

ES78 Room B30 METHODS IN STATISTICAL MODELLING Chair: Heather Turner

E063: **A hybrid symbolic-numerical method for determining model structure**
Presenter: **Diana Cole**, University of Kent, United Kingdom
Co-authors: Remi Choquet, Ben Hubbard
In some models it is not possible to estimate all the parameters regardless of the amount of data collected. Such a problem stems from the inherent structure of a model; for example, two parameters could be confounded and only ever appear as a product. This is known as parameter redundancy or the parameters are termed unidentifiable. In complex models the nature of the confounding may not be obvious. Numerical and Symbolic methods exist to detect parameter redundancy. Numeric methods can be inaccurate and lead to the wrong conclusions, but symbolic methods may be difficult to use or infeasible to use in some cases. An alternative is a hybrid-symbolic numeric method. This method combines symbolic and numeric methods to create an algorithm that is extremely accurate compared to other numeric methods and is computationally inexpensive. A series of generic computational steps are developed to create a method that is ideal for practitioners to use. This method is illustrated using ecological models, such as multi-site capture-recapture models and occupancy models.

E312: **The Laplace approximation powered by automatic differentiation**
Presenter: **Hans Skaug**, University of Bergen, Norway
The Laplace approximation is used in software packages, as well as by individual researchers, to calculate the marginal likelihood in hierarchical models. In combination with sparse matrix technology it provides a generic approach to model fitting, covering spatial and state-space models among other model classes. We will explain how the Laplace approximation can be made transparent via a numerical technique known as automatic differentiation (AD). The use of AD greatly reduces the cost of building and modifying models, and we will describe how these ideas have been

implemented in the open source software package AD Model Builder (ADMB). The flexible interface of ADMB encourages the user to explore models beyond those that can be fitted in R and other statistical packages. We will focus on spatial hierarchical models.

E358: **A general framework for parametric survival analysis**
Presenter: **Michael Crowther**, University of Leicester, United Kingdom
Co-authors: Paul Lambert
Parametric survival models are being increasingly used as an alternative to the Cox model in biomedical research. Through direct modelling of the baseline hazard function we can gain greater understanding of the risk profile of patients over time, obtaining absolute measures of patient risk. Commonly used parametric survival models, such as the Weibull, make restrictive assumptions of the baseline hazard function, such as monotonicity, which is often violated in clinical datasets. We propose a general framework of survival models, where any well-defined baseline hazard function can be specified, with the model estimated using maximum likelihood utilising numerical quadrature. Special cases of this framework are described, using restricted cubic splines or fractional polynomials to model the baseline, with extension to complex time-dependent effects following naturally. Time-varying covariates can be incorporated and the extension to relative survival models is described. User friendly Stata software is provided which significantly extends parametric survival models available in standard software.

E424: **Data visualization (and beyond) with local principal curves and manifolds**
Presenter: **Jochen Einbeck**, Durham University, United Kingdom
Co-authors: Ludger Evers
Principal curves are descriptively defined as smooth curves through the "middle" of a multivariate data could. They can be considered as a nonparametric version of linear principal component analysis (PCA). Among various other techniques proposed for their estimation, principal curves can be constructed by iterating local PCA and mean shift steps. The R package LPCM is firstly introduced, which implements this technique. It will be illustrated how the fitted principal curve can not only be used for data visualization, but also for data compression, regression, or density estimation. The concept of principal curves can be generalized towards principal surfaces (2-dim smooth data summaries) or even principal manifolds (smooth objects of arbitrary dimension). Local principal curves can be extended to local principal manifolds by building a mesh of appropriately constrained triangles (simplices, tetrahedrons, etc.). The approach has been implemented in an experimental R package, which forms a powerful visualization tool and also supports multivariate (penalized) regression based on the fitted manifold.

ES124 Room B20 BAYESIAN SEMI- AND NONPARAMETRIC MODELLING III Chair: Igor Pruenster

E457: **Coherent states stochastic expansion for quantum homodyne tomography**
Presenter: **Eric Barat**, CEA, France
Co-authors: Zacharie Naulet, Judith Rousseau
Quantum homodyne tomography is considered as an inverse statistical problem in which the aim is to estimate the quantum state of a light beam. The quantum state is characterized by the Wigner distribution W of the electric and magnetic fields q and p. Observations consist of independent samples from the distribution of $y = (x, \theta)$ where $x = q\cos\theta + p\sin\theta$, where θ is independent of (q,p). This situation is related to classical emission tomography problems, except that though observations distributions remain *bona fide*, W may in contrast exhibit patches of negative values. A random Wigner distribution may be derived from random wave functions. Thus, a nonparametric Bayesian kernel model for the wave function prior is proposed. For the kernel's choice it supports the canonical coherent state of the system corresponding to the linear representation of the Weyl-Heisenberg group. This choice leads to a neat interpretation in quantum physics as in approximation theory. The stochastic expansion is based on a random probability measure. For the special case of Dirichlet process probability measure, a Gibbs sampler is developed relying on a particle approximation emerging from a recent study of measure-valued Markov chains.

E804: **Bayesian analysis for monotone hazard ratio**
Presenter: **Yongdai Kim**, Seoul National University, South Korea
A Bayesian approach is proposed for estimating the hazard functions under the constraint of a monotone hazard ratio. We construct a model for the monotone hazard ratio utilizing Cox's proportional hazards model with a monotone time-dependent coefficient. To reduce computational complexity, we use a signed gamma process prior for the time-dependent coefficient and the Bayesian bootstrap prior for the baseline hazard function. We develop an efficient MCMC algorithm and illustrate the proposed method on simulated and real data sets. We discuss asymptotic properties.

E649: **Asymptotic estimators for discovery probabilities**
Presenter: **Bernardo Nipoti**, University of Turin and Collegio Carlo Alberto, Italy
Co-authors: Stefano Favaro
Species sampling problems have recently emerged in genomic applications. This context gives rise to new challenges since one has to deal with very large genomic libraries containing a huge number of distinct genes with only a small portion of the library being sequenced. Recent works have shown that a Bayesian nonparametric approach can be successfully undertaken to estimate quantities of interest. With large samples, the exact evaluation of the devised estimators becomes impracticable and asymptotic approximations can be sought. The study focuses on the asymptotic behavior of discovery probability and coverages: this allows us to provide both asymptotic estimators and a measure of their uncertainty. As a byproduct, an intriguing structural analogy, between the devised estimators and the well-known frequentist counterparts, is highlighted.

E525: **Large scale nonparametric Bayesian manifold regression**
Presenter: **Rajarshi Guhaniyogi**, Duke University, United States
Co-authors: David Dunson
Nonparametric regression for high-dimensional predictors is a common problem in applications, but computational and statistical efficiency are major concerns. Fortunately, in a number of real life applications predictors lie on a noisy lower dimensional manifold. Manifold regression attempts to estimate such a manifold. Motivated from the predictor compression idea, a compressed Gaussian process regression model is proposed, which randomly projects high dimensional predictors to lower dimension, bypassing computationally prohibitive estimation of manifold. Conditionally on the projection matrix and covariance parameters, the posterior distribution for model parameters and posterior predictive distributions for out-of-sample observations are available analytically. Running the analysis in parallel for many random projections and covariance parameters, model averaging is used to combine the results. The algorithm is extremely fast to implement and assumes easy extensions to massive p and n. Proposed model is found to yield superior predictive performance in massive n, p settings compared to the state of the art competitors.

EP02 Room Macmillan POSTER SESSION II Chair: Francisco Torres-Ruiz

E1018: **Recursive filtering algorithm in systems with correlated random measurement matrices**
Presenter: **Irene Garcia-Garrido**, Universidad de Granada, Spain
Co-authors: Josefa Linares-Perez, Raquel Caballero-Aguila, Aurora Hermoso-Carazo
The recursive least-squares linear filtering problem is addressed in a class of discrete-time stochastic systems with independent random state transition matrices and auto-correlated and cross-correlated random measurement matrices. Usually, the recursive optimal state estimation in systems with random matrices is addressed by transforming the original system into a system with deterministic parameter matrices and state-dependent process and measurement noises, to which the optimal Kalman filter is applied. In contrast, the proposed optimal recursive filtering algorithm, obtained by an innovation approach, does not require any system transformation. The proposed algorithm is applied to multi-sensor systems with correlated randomly delayed measurements, which constitute a current topic of growing interest due to its wide applications in networked systems. The Bernoulli random variables characterizing the measurement delays in each sensor are assumed to be correlated at consecutive sampling times; this correlation model covers situations where two successive observations cannot be delayed, and hence it can be applied to networked systems with stand-by sensors for the immediate replacement of a failed unit. A system with state-dependent multiplicative noise, and delayed measurements coming from two sensors with different delay characteristics is considered to illustrate the effectiveness of the proposed recursive filtering algorithm.

E1024: **On a generalized class of logistic diffusion processes**
Presenter: **Francisco Torres-Ruiz**, Granada, Spain
Co-authors: Antonio Barrera-Garcia, Patricia Roman-Roman
Mathematical models are widely used in several research fields to explain a variety of dynamical phenomena. Internal and external conditions can produce random influences that must be taken into account. For this reason stochastic modeling is usually required. In particular, growth phenomena can be well described by the theory of stochastic diffusion processes. Logistic models, in combination with stochastic theory, allow researchers to analyze phenomena with specific characteristics such as a sigmoidal growth. Several growth models based on logistic curves are commonly used. For instance, the hyperbolastic growth model of type I, which is an extension of the logistic one, has been successfully applied to biological processes like stem cell growth dynamics, as well as to epidemiological phenomena. Dealing with all these logistic-type models from a unique, generalized point of view, can help determine a rigorous way to approach dynamical. We aim to establish the theoretical framework to define a functional generalization of the logistic model, extending it to a diffusion process which could explain logistic-type phenomena under unknown random influences. This leads to a characterization of several models (hyperbolastic included) as particular cases of a general theory.

E1025: **A diffusion process related to a logistic curve with multiple inflexion points**
Presenter: **Patricia Roman-Roman**, Universidad de Granada, Spain
Co-authors: Eva Poza-Cruz, Francisco Torres-Ruiz
Sigmoidal growth models (either deterministic or stochastic) are commonly developed to study the behaviour of variables exhibiting a slow initial growth followed by a quick exponential growth which then slows down until reaching their carrying capacity. Nevertheless, in many real-life situations, maximum growth is reached after several stages, each of them showing a deceleration followed by explosive, exponential growth. This commonly implies the usual sigmoidal curve being replaced by a curve with more than one inflexion point. Some examples can be found in the growth of several species of fruit-bearing trees, as well as in the fatigue profiling of the muscles of mice. We present a logistic-type diffusion model whose mean is a sigmoidal curve with several inflexion points. The maximum likelihood estimation of its parameters is carried out and several examples are provided

E1069: **An alternative estimation approach to fit a heterogeneity linear mixed model**
Presenter: **Marie-Jose Martinez**, University of Grenoble, France
Co-authors: Emma Holian
An alternative estimation approach is proposed to fit a linear mixed effects model where the random effects follow a finite mixture of normal distributions. This model, called a heterogeneity linear mixed model, is an interesting tool since it relaxes the classical normality assumption and is also perfectly suitable for classification purposes, based on longitudinal profiles. Instead of fitting directly the heterogeneity linear mixed model, we propose to fit an equivalent mixture of linear mixed models under some restrictions which is computationally simpler. Indeed, unlike the former model, the latter can be maximized analytically using an EM-algorithm and the obtained parameter estimates can be easily used to compute the parameter estimates of interest. We study and compare the behaviour of our approach on simulations. Finally, the use of our approach is illustrated on a real data set.

E1070: **Fitting over and under-dispersed count data with the hyper-Poisson regression model**
Presenter: **Antonio Jose Saez-Castillo**, University of Jaen, Spain
Co-authors: Antonio Conde-Sanchez, Seyed Hadi Khazraee
The hyper-Poisson regression model is formulated as a generalization of the Poisson regression model that permits to handle overdispersion and underdispersion in the same dataset, depending on the value of the covariates. In contrast to other models, the covariates enter the mean function at the same time that they influence the dispersion of the distribution, in such a way that the effect of each variable on the expected mean is easily interpretable. We discuss its capability of fitting in a simulation context and also with several real datasets, and compare it with with other common regression models.

E1072: **Linear and quadratic estimators based on covariances with multiple fading measurements**
Presenter: **Josefa Linares-Perez**, Universidad de Granada, Spain
Co-authors: Raquel Caballero-Aguila, Aurora Hermoso-Carazo, Irene Garcia-Garrido
Stochastic uncertainties usually arise in data transmission problems involving communication networks. Missing measurements have been widely treated due to their applicability to model many real problems and, usually, Bernoulli variables are used to model the fact that the measured signal is either completely lost or successfully transmitted. Recently, this missing measurement model has been generalized to cover some practical applications where only partial information is missing (fading measurements). The linear and quadratic least-squares signal estimation problems from fading measurements coming from multiple sensors are addressed when, at each sensor, the fading probability is described by a scalar random variable with arbitrary discrete probability distribution over [0,1]. At each sensor, this model considers the possibility of observations containing only partial information about the state, or even only noise, and the different sensors may have different fading probabilities. Linear and quadratic recursive filtering algorithms are designed using an innovation approach; they do not require full knowledge of the state-space model for the signal process, but just the moments up to the fourth-order of the signal and the observation noise. Recursive expressions for the estimation error covariance matrices are also derived, and the performance of estimators is illustrated by a numerical example.

E1080: **Comparing generalized Pareto models fitted to extreme observations: An application to the largest temperatures in Spain**
Presenter: **Andres M Alonso**, Universidad Carlos III de Madrid, Spain
Co-authors: Patricia de Zea Bermudez, Manuel G Scotto

A subsampling-based testing procedure for the comparison of the exceedance distributions of stationary time series is proposed. The proposed testing procedure has a number of advantages including the fact that the assumption of stationary can be relaxed for some specific forms of non-stationary and also that the two time series are not required to be be independently-generated. The performance of the testing procedure is illustrated through a simulation study and with an empirical application to a set of data concerning daily maximum temperature in the seventeen autonomous communities of Spain for the period 1990-2004. The autonomous communities were clustered according to the similarities of the fitted Generalized Pareto models and then mapped. The cluster analysis reveals a clear distinction between the four north-east communities on the shores of the Bay of Biscay and the remaining regions. A second cluster corresponds to the southern Mediterranean area and the central region which corresponds to the communities with highest temperatures.

E1270: **Bayesian analysis of the log-linear Birnbaum-Saunders model**
Presenter: **Heleno Bolfarine**, University of Sao Paulo, Brazil
Co-authors: Aldo Garay

The Birnbaum-Saunders (BS) model is a two-parameter, unimodal distribution with positive skewness and non-negative support. This model has attracted considerable attention in the statistical literature during the past decade. The interest for the BS model is due to its physical and theoretical arguments, its attractive properties and its relationship with the normal distribution. In this paper, we study log-linearextensions of the BS model using different parameterizations. Specifically, we analyze several possible parameterizations, use the Bayesian approach for model fitting and compare their performance using the deviance information criterion (DIC). To implement the Bayesian approach we make use of Markov chain Monte Carlo (MCMC) methodology. To study influence of observations, we consider statistics based on theKulback-Lieber divergence index. The approach is implemented in two real data sets previously analyzed in the literature.

EP03 Room Macmillan POSTER SESSION III Chair: Francisco Torres-Ruiz

E1091: **Influence diagnostics for censored data**
Presenter: **Inmaculada Barranco-Chamorro**, University of Sevilla, Spain
Co-authors: Florencia Osorio-Baeza

Influence analysis deals with the study and assessment of the variation caused in statistical conclusions by perturbations in data and/or model assumptions. Influence diagnostics are necessary to identify individuals and combination of cases that may have an important effect on conclusions of a statistical study. We first review methods handling this topic for censored data. We focus on log-location-scale regression models with censored data. We study in depth the second order approach to local influence in these models. We show that this approach provides interesting results. In fact, these techniques can be considered as a useful, complementary tool to other diagnostics in the literature. Examples and simulations are included to illustrate the effectiveness of these diagnostics.

E1116: **Building the structure function of reliability systems using statistical regression tools**
Presenter: **Antonio Jesus Lopez-Montoya**, University of Granada, Spain
Co-authors: Maria Luz Gamiz-Perez, Maria Dolores Martinez-Miranda

Nonparametric kernel methods are used in regression models to derive the structure function of a reliability system. The nonparametric approach delivers estimators that represent better the given data and help to prevent the problems that result from inappropriate parametric assumptions. Previously, the structure function is formulated as the conditional mean in a regression model, that is, the relationship between the components and the system itself. From these settings, the structure function can be estimated using modern regression techniques such as the local linear kernel approach. The only issue is that the provided function should meet a number of restrictions to ensure the common structure function properties: monotonicity, proper extrema and component relevance. An estimation strategy that consists of two steps is proposed. Firstly, local linear kernel regression is used to estimate the structure function. Secondly, and to ensure the monotonicity of the estimated function, a monotonization procedure based on a simple projection of the convex polyhedral cone is performed. The cone projection algorithm is more accurate than classical Pool-Adjacent-Violators Algorithm (PAVA). While PAVA breaks down to dealing with complex systems with more than two components, the cone projection algorithm is simpler to generalize to higher dimensions.

E1143: **A simulation study for testing equality of unbalanced several gamma means**
Presenter: **Evren Ozkip**, Anadolu University, Turkey
Co-authors: Berna Yazici, Ahmet Sezer

Multiple comparison methods (MCMs) are widely used to investigate differences between population means or, more generally, between subsets of population means using sample data. The gamma distribution is a generalization of the exponential distribution which has been widely used as a model in reliability studies and life testing experiments. A simulation study for testing equality the mean of unbalanced several gamma populations is presented. The methods are based on the concepts of generalized F-test and Computational approach test (CAT). The purpose of this study is to compare these tests according to type one error and powers in different combinations of parameters. An application is shown using rainfall data.

E998: **On choosing an appropriate regression model in presence of multicolinearity**
Presenter: **Maruf Raheem**, University of Uyo, Nigeria
Co-authors: Kayode Samson

The aim is to illustrate how to detect and solve the problem of multicolinearity in regression analysis. As such, Variance Inflation Factor (VIF) and the Condition Index (CI) were used as measures of such detection. Ridge Regression (RR) and the Principal Component Regression (PCR) were the two other approaches used in modeling apart from the conventional simple linear regression. For the purpose of comparing the two methods, simulated data were used. Our task is to ascertain the effectiveness of each of the methods based on their respective mean square errors. From the result, we found that Ridge Regression (RR) functions better than principal component regression when multicolinearity exists among the predictors.

E1185: **Variance function estimation in nonparametric regression models with multiple covariates**
Presenter: **Kee-Hoon Kang**, Hankuk University of Foreign Studies, Korea, South
Co-authors: Seok-Oh Jeong

The estimation problem of variance function in nonparametric heteroscedastic regression models with multiple predictors is considered. A nonparametric estimator using the differences between observations is proposed. This is an extension of previous related works. Some theoretical properties including bias-variance calculation and asymptotic normalities are derived. Numerical performances are explored via some simulated data and real example.

E1196: **Multiple correspondence analysis for educational test items**

Presenter: **Sayaka Arai**, The national center for university entrance examinations, Japan

Multiple correspondence analysis (MCA) is a method for quantifying categorical multivariate data. It represents data as clouds of points in a low-dimensional space. Application of MCA to academic examination data is considered. Academic examination data is usually analyzed using item response theory (IRT). IRT, which is also called latent trait theory, is a method for estimating latent traits (e.g., academic proficiency) and is used in many educational testing applications. IRT characterizes each test item by its item parameters such as difficulty and discrimination ability. On the other hand, MCA represents the pattern of relationships of categorical variables (i.e., test items) and it also represents the relationships within each test item. Therefore, MCA may provide useful information for item analysis. An empirical study using a linguistic performance test data is done. The analysis using MCA is compared to that of other methods.

E484: **An attitude scaling application of fuzzy sets to profile analysis**

Presenter: **Giorgos Stamou**, National Technical University of Athens, Greece

Co-authors: Catherine Michalopoulou, Maria Symeonaki

Profile analysis is usually concerned with scaling individuals. In psychological and educational applied testing, the term “profile” comes from the practice of plotting individuals’scores on a battery of tests’standardized variables as a profile. However, this type of analysis requires normative data and therefore is but rarely used in attitude scaling.A methodology for developing individuals’profiles based on fuzzy Likert scale scores is presented using percentile norms. The method is empirically demonstrated on a Likert scale measuring xenophobia that was used in a large-scale sample survey conducted in Northern Greece by the National Centre for Social Research. The performance of the proposed method is assessed by using single indicators of xenophobia.

Sunday 15.12.2013 16:50 - 18:30 Parallel Session M – CFE

CS89 Room B35 MARKET MICROSTRUCTURE Chair: Liam Cheung

C1031: **News, liquidity dynamics and intraday jumps: Evidence from the HUF/ EUR market**
Presenter: **Xing Han**, Ghent University, Belgium
Co-authors: Frederick van Gysegem, Michael Froemmel
Intraday jumps on a pure limit order FX market are studied by linking them to news announcements and liquidity shocks. First, we show that jumps are frequent and contribute greatly to the return volatility. Nearly half of the jumps can be linked with scheduled and unscheduled news announcements. Furthermore, we show that jumps are information based, whether they are linked with news announcements or not. Prior to jumps, liquidity does not deviate from its normal level, nor do liquidity shocks offer any predictive power for jump occurrence. Jumps emerge not as a result of unusually low liquidity but rather as a result of an unusually high demand for immediacy concentrated on one side of the book. During and after the jump, a dynamic order placement process emerges: some participants endogenously become liquidity providers and absorb the increased demand for immediacy. We detect an interesting asymmetry and find the liquidity providers to be more reluctant to add liquidity when confronted with a news announcement around the jump. Further evidence shows that participants submit more limit orders relative to market orders after a jump. Consequently, the informational role of order flow becomes less pronounced in the thick order book after the jump.

C1004: **On the aggressiveness of high frequency traders**
Presenter: **Dong Zhang**, Stockholm University, Sweden
Co-authors: Bjorn Hagstromer, Lars Norden
Order aggressiveness of market-making high-frequency traders (HFTs), opportunistic HFTs, and non-HFTs is studied. We find that market-making HFTs follow their own group's previous order submissions more than they follow other traders' orders. Opportunistic HFTs and non-HFTs tend to split market orders into small portions submitted in sequence. HFTs submit more (less) aggressive orders when the same-side (opposite-side) depth is large, and supply liquidity when the bid-ask spread is wide. Thus, HFTs adhere strongly to the trade-off between waiting cost and the cost of immediate execution. Non-HFTs care less about this trade-off, but react somewhat stronger than HFTs to volatility.

C787: **Fragmentation and market quality in the Canadian equity markets**
Presenter: **Liam Cheung**, McGill University, Canada
The relation between market quality and fragmentation over multiple heterogeneous marketplaces is studied in the Canadian equity market using a proprietary database of recorded market data from all Alternative Trading Systems (ATSs). First, we confirm that fragmentation is associated with higher market quality on an aggregated stock basis. Second, based on segmented data from each individual marketplace, we determine that the larger alternative marketplaces are the primary sources of the positive relation to market quality on an aggregate basis. Third, after adjusting for selection bias, we find that total ATS market share has a negative relation to effective spread and a positive relation to market impact. While certain ATSs still have a positive relation between effective spread and market share, the size of the ATS is no longer a determinant in the association of marketplaces and improved market quality. Subsequent tests on individual stocks reveal that days with higher fragmentation have a positive (negative) relation to market quality for high (low) volume stocks. Our findings indicate therefore that the increased fragmentation of equity markets may not be a uniformly beneficial practice for equity market quality.

CS15 Room B34 APPROXIMATE BAYESIAN COMPUTING Chair: Veronika Czellar

C344: **Approximate Bayesian computation with indirect summary statistics**
Presenter: **Alexander Gleim**, University of Bonn, Germany
Co-authors: Christian Pigorsch
Approximate Bayesian Computation (ABC) has become a popular estimation method for situations where the likelihood function of a model is unavailable. In contrast to classical Bayesian inference this method does not build on the availability of the likelihood function but instead relies on a distance function which measures the distance between some empirical summary statistics and their simulation based counterparts. An open question is the selection of these statistics, particularly regarding their sufficiency properties. We propose an indirect approach with summary statistics based on statistics of a suitably chosen auxiliary model. We show sufficiency of these statistics for Indirect ABC methods based on parameter estimates (ABC-IP), likelihood functions (ABC-IL) and scores (ABC-IS) of the auxiliary model. In the case of score based summary statistics we provide an efficient way of weighting them. This weighting scheme allows us to appropriately assess the distance between the true posterior distribution and the approximation based on the ABC-IS method. We illustrate the performance of these Indirect ABC methods in a simulation study and compare them to a standard ABC approach. We also apply the ABC-IS method to the problem of estimating a continuous-time stochastic volatility model based on non-Gaussian OU-processes.

C551: **Approximate Bayesian computation and sequential Monte Carlo methods for risk management and insurance applications**
Presenter: **Gareth Peters**, University College London, United Kingdom
The development of Approximate Bayesian Computation methodology for risk management and insurance applications is discussed. In particular, it focuses on Bayesian modelling under the Loss Distributional Approach for heavy tailed and flexible skew-kurtosis loss processes. We demonstrate recent methodology developed for estimation of such models in the class of Sequential Monte Carlo methods, which utilise Partial Rejection Control to manage the ABC context efficiently.

C639: **Methods to improve ABC inference of state space models**
Presenter: **Dennis Prangle**, Reading University, United Kingdom
Co-authors: Paul Fearnhead
State space models have many applications in economics and finance. However, inference for many models of interest is difficult as their likelihood cannot be numerically evaluated, for example when the model involves an alpha stable distribution. Approximate Bayesian computation methods can avoid the difficulty at the cost of providing only approximate inference results. The aim is to present work in progress on methods to reduce approximation error for inference of state space models.

C571: **Accurate methods for approximate Bayesian computation filtering**
Presenter: **Veronika Czellar**, HEC Paris, France
Co-authors: Laurent Calvet
The Approximate Bayesian Computation (ABC) filter extends the particle filtering methodology to general state-space models in which the density of the observation conditional on the state is intractable. The aim is to provide an exact upper bound for the mean squared error of the ABC filter and show that under appropriate bandwidth and kernel specifications, ABC converges to the target distribution as the number of particles goes to infinity. The optimal convergence rate decreases with the dimension of the observation space but is invariant to the complexity of the state space. It is also shown that the usual adaptive bandwidth used in the ABC literature leads to an inconsistent filter. A plug-in rule for the bandwidth is

developed and the good accuracy of the resulting filter is demonstrated in a variety of cases including α-stable stochastic volatility models and multifractal asset pricing models with investor learning.

CS32 Room B30 SOME DEVELOPMENT IN RISK MODELS Chair: Javier Hidalgo

C817: **Risk aversion of market makers and asymmetric information**
Presenter: **Umut Cetin**, LSE, United Kingdom
Co-authors: Albina Danilova

The equilibrium impact of market makers' risk aversion on the equilibrium in a speculative market consisting of a risk neutral informed trader and noise traders is analyzed. The unwillingness of market makers to bear risk causes the informed trader to absorb large shocks in their inventories. The informed trader's optimal strategy is to drive the market price to its fundamental value while disguising her trades as the ones of an uninformed strategic trader. This results in a mean reverting demand, price reversal, and systematic changes in the market depth. We also find that an increase in risk aversion leads to lower market depth, less efficient prices, stronger price reversal and slower convergence to fundamental value. The endogenous value of private information, however, is non-monotonic in risk aversion.

C837: **Testing for equality of an increasing number of spectral density functions**
Presenter: **Pedro Souza**, London School of Economics, United Kingdom
Co-authors: Javier Hidalgo

The aim is to examine a nonparametric test for the equality among an increasing number of spectral density functions. One example of interest is when we wish to test the constancy of the covariance structure of a time series sequence. A second example of relevance is with spatiotemporal data or in a panel data framework we wish to decide if the dynamic structure along time does not depend on the location in space or on the individual. A third example can correspond to the problem of whether the spectral density is separable. A final example is in a panel data context, if we wish to test whether the dynamics is invariant across individuals. Two features of the test are that (a) despite the fact that we do not assume a particular dynamic specification under the null hypothesis, there is no need to choose any bandwidth parameter, by taking advantage of the fact we can make use of the fact that the dimension increases to infinity, and (b) the asymptotic distribution of the test is pivotal and under some mild conditions it converges to a normal random variable. Finally, we present a Monte-Carlo experiment to illustrate the finite sample performance of the test as well as a valid bootstrap procedure.

C838: **The scale of predictability**
Presenter: **Andrea Tamoni**, London School of Economics, United Kingdom
Co-authors: Federico Bandi, Benoit Perron, Claudio Tebaldi

It is argued that predictive relations that are found to be elusive when using raw data may hold for different layers or details in the cascade of shocks affecting financial time series. This observation leads to a notion of detail–specific predictability. Said differently, there may be components of the return process with cycles of a certain frequency which are predictable but this predictability may be hidden by higher-frequency components leading the raw returns for which predictability is, as generally found, uncertain. The aim is to propose direct extraction of the time-series details – and regressions on the details - as well as indirect extraction by means of two-way aggregation of the raw series - and regressions on forward/backward aggregates of the raw series. The mapping between the two methods is established theoretically and their close relation is exploited empirically. While the direct method allows one to identify the data generating process (i.e., the details and, upon reconstruction, the original series), the indirect method provides us with a way to evaluate the frequency at which layers in the information flow are connected across economic variables and employ this information for prediction. By re-writing the way in which one implements long-run predictability (aggregated regressand on past aggregated regressors, rather than on past regressor over one period), two-way aggregation provides a natural way to exploit scale-specific predictability in asset allocation for the long run. For different asset classes, we show that two-way aggregation yields predictability relations that are considerably stronger than in the extant work.

C778: **Heavy stress-testing in non-normal markets via entropy pooling**
Presenter: **David Ardia**, Laval University, Canada
Co-authors: Attilio Meucci

The issue of parametric entropy pooling with non-normal risk drivers is addressed. Our approach rests fundamentally on two pillars: a copula-marginal decomposition, which allows us to model skew and fat tails via marginals, and parallel computing to optimize our search over a large number of parameters. We illustrate the new approach with the heavy stress-testing of a portfolio of options.

CS103 Room B29 APPLIED ECONOMETRICS II Chair: James Mitchell

C967: **How efficient is the global oil market: A multifractal perspective**
Presenter: **Marc Gronwald**, University of Aberdeen, United Kingdom
Co-authors: Cristina Sattarhoff

The efficiency of global oil markets is studied using a newly developed test for the type of financial market efficiency as well as a proxy for the degree of financial market efficiency, based on the multifractal random walk model. This procedure allows for differentiating between random walk and multifractal martingale price processes as well as between different degrees of martingale efficiency. Applying this test using roll-ing sample windows allows one to study the development of this efficiency measure over time can be studied. From this exercise interesting insights with respect to the fierce debate on potentially undesirable influences of the "financialization of oil futures markets" observed after 2002 on crude oil prices ("Masters hypothesis"). The argument that higher market liquidity is beneficial for the market efficiency is often invoked in this context. We contribute to this discussion by studying whether oil market efficiency actually changed/improved after 2002.

C1106: **Asymmetric relationship between financial integration and economic growth**
Presenter: **Duygu Yolcu Karadam**, Middle East Technical University, Turkey
Co-authors: Nadir Ocal

Financial globalization and its implications for growth have been discussed intensively as also witnessed during global financial crisis. The question of whether benefits or costs of financial integration outweigh has not clearly answered yet. It is argued that the growth effects of international financial integration depend on some structural and economic characteristics of the economies. Countries need to satisfy some initial conditions in order to benefit from capital flows and reduce the risks associated with them. We examine whether the growth effects of financial integration display asymmetric features due to some structural and economic characteristics such as financial sector development, institutional quality, trade openness and macroeconomic policies. To this end, we employ a class of nonlinear models called Panel Smooth Transition Regression Models (PSTR), to analyze the effect of financial integration on growth for a panel data set of around 80 countries over the period of 1970-2010. Empirical results show significant nonlinearities in the relationship between financial openness and economic growth. International financial integration has asymmetric effects on growth due to the levels of financial development, institutional quality, trade openness, inflation volatility and government budget balance. We also find that the signs and the magnitudes of these effects can differ for emerging, advanced and other developing countries.

C1102: **Nowcasting Swiss GDP in real-time squared**
Presenter: **Boriss Siliverstovs**, KOF ETHZ, Switzerland

The aim is to put a large-scale factor model comprising of 567 economic indicators to a forecasting exercise in real-time squared. That is, we strictly use only the information that was available to us at a time of making forecasts and we announce our forecasts at the same time we made those. The fact that we produce forecasts of quarterly GDP growth at the weekly frequency allows us to continuously monitor the state of the economy and trace the influence of newly released patches of economic data on our forecasts. On this basis, we were able to capture a phase of economic recovery after the Great Recession, a phase of relatively high growth during the year 2010, and a phase of declining confidence caused by escalating European debt crisis and growing fears of entering a new recession during the evaluation period from 2010Q1 until 2012Q4. According to our results, the latter phase started in May 2011 and continued until November 2011, when pessimistic sentiments dominated economic outlook. For Switzerland, this period is marked by rapid appreciation of the Swiss Franc, prompting the Swiss National Bank to introduce the minimum bound on the CHF/EUR exchange rate in September 2011. In the year 2012 we observe a slight improvement in the outlook during the first half. In the latest period the Swiss economy entered a phase of a rather low but steadily positive growth.

C1109: **Business tendency surveys and macroeconomic fluctuations**
Presenter: **Rolf Scheufele**, Swiss National Bank, Switzerland
Co-authors: Daniel Kaufmann

The information content of business tendency surveys for key macroeconomic variables in Switzerland is investigated. Therefore, we consider a large data set of the KOF – the leading business cycle research institute in Switzerland – which collects information of firms of various sectors conducted at monthly and quarterly frequency (about 300 time series). The obvious advantage of business tendency surveys is their prompt availability and that they are not subject to revisions. In the sequel we show how one can optimally relate the qualitative response data to macroeconomic aggregates of the official statistic. Our aim is to learn which economic factors, as reported by individual firms, are mostly explaining the current state of the economy. In a further step we conduct an out-of-sample study to utilize the survey information for nowcasting and short-term forecasting. To summarize the information content of the business tendency surveys we propose a large-scale factor model for the unbalanced data set which is able to deal with mixed frequencies and a large fraction of missing observations. It turns out that the predictive ability of our survey-based factor approach is quite encouraging. This implies that qualitative survey data are extremely helpful for monitoring the current state of the economy.

C855: **External finance, internal finance and innovation: Evidence from the Italian manufacturing firms**
Presenter: **Marianna Succurro**, University of Calabria, Italy
Co-authors: G. Damiana Costanzo, Damiano B. Silipo

The impact of different sources of finance on innovation is studied by focusing on Italian manufacturing firms over the 2003-2010 period. By using a dynamic econometric model based on an original measure of R&D expenditure, we get surprising results compared to previous findings. The effects of finance on innovation can be heterogeneous depending on firms' relative size, geographic location and technological intensity of the sector. In the Centre-North, where access to financial markets is easier, firms rely on external funding to undertake R&D. By contrast, in Southern regions, where access to external finance is harder, firms do R&D by substituting external funds with internal resources. The different role played by financial institutions in Mezzogiorno and Centre-North of Italy also emerges by comparing large companies and SMEs. Empirical evidence shows that external finance plays a crucial role in spurring innovation of high-tech firms, independently from their size. For patent probability, on the contrary, internal funding seems economically more relevant than external funding.

CS82 Room B33 THE SIMULATION OF NON-LINEAR MODELS Chair: Michel Juillard

C346: **The pruned state-space system for non-linear DSGE models: Theory and empirical applications**
Presenter: **Martin M Andreasen**, Aarhus University, Denmark

The pruned state-space system for higher-order approximations to the solutions of DSGE models is studied. For second- and third-order approximations, we derive the statistical properties of this system and provide closed-form expressions for first and second unconditional moments and impulse response functions. Thus, our analysis introduces GMM estimation for DSGE models approximated up to third-order and provides the foundation for indirect inference and SMM when simulation is required. We illustrate the usefulness of our approach by estimating a New Keynesian model with habits and Epstein-Zin preferences by GMM when using first and second unconditional moments of macroeconomic and financial data and by SMM when using additional third and fourth unconditional moments and non-Gaussian innovations.

C371: **Terms of trade uncertainty and business cycle fluctuations**
Presenter: **Johannes Pfeifer**, University of Tuebingen, Germany
Co-authors: Benjamin Born, Gernot Muller

The recent commodity price boom and the financial crisis have been accompanied by a large increase in price volatility, resulting in a significant increase in the uncertainty associated with the terms of trade for many countries. The aim is to study the effects of terms of trade uncertainty on business cycles through the lens of a small open economy DSGE model. Sequential Monte Carlo Methods are used to estimate a stochastic volatility model to deal with the latent state "uncertainty". Analyzing the case of Chile, the findings are fourfold. First, there is considerable evidence for time-varying terms of trade uncertainty in the data. Second, the ex-ante and ex-post effects of increased terms of trade uncertainty can account for about one fifth of Chilean output fluctuations at business cycle frequencies. Third, a two-standard deviation terms of trade risk shock, i.e. a 54 percent increase in uncertainty, leads to a 0.1 percent drop in output. The fact that terms of trade uncertainty more than doubled during the recent commodities boom suggests that the contribution of terms of trade risk during this more recent period may have been substantial. Finally, it is shown that both the cautionary savings motive of the representative household and the expansionary response of the central bank mitigate the drop in GDP.

C704: **OccBin: A toolkit for solving dynamic models with occasionally binding constraints**
Presenter: **Luca Guerrieri**, Federal Reserve Board, United States
Co-authors: Matteo Iacoviello

The purpose is to describe how to adapt a first-order perturbation approach and apply it in a piecewise fashion to handle occasionally binding constraints in dynamic models. Our examples include a real business cycle model with a constraint on the level of investment and a model of optimal consumption choice in the presence of liquidity constraints. In each case, we compare the piecewise linear perturbation solution with a high-quality numerical solution that can be taken to be virtually exact. The piecewise linear perturbation method can adequately capture key properties of the models we consider. A key advantage of this method is its applicability to models with a large number of state variables.

C929: **Nonlinear and stable perturbation-based approximations**
Presenter: **Joris de Wind**, CPB Netherlands Bureau for Economic Policy Analysis, Netherlands
Co-authors: Wouter den Haan

Users of regular higher-order perturbation approximations can face two problems: policy functions with odd oscillations and simulated data that explode. We propose a perturbation-based approximation that (i) does not have odd shapes, (ii) generates stable time paths, and (iii) avoids the

drawbacks that hamper the pruned perturbation approach. For models with nontrivial nonlinearities, we find that our alternative and the pruned perturbation approximations give a good qualitative insight in the nonlinear aspects of the true solution, but can differ from the true solution in some quantitative aspects, especially during severe peaks and troughs.

CS69 Room B18 CONTRIBUTIONS TO RISK MODELING AND MANAGEMENT I Chair: Stefan Mittnik

C1010: **A new set of improved value-at-risk backtests**
Presenter: **Gregor Weiss**, TU Dortmund University, Germany
Co-authors: Daniel Ziggel, Tobias Berens, Dominik Wied
A new set of formal backtests for VaR-forecasts is proposed. They significantly improve upon existing backtesting procedures. Our new test of unconditional coverage can be used for both directional and non-directional testing and is thus able to test separately whether a VaR-model is too conservative or underestimates the actual risk exposure. Second, we stress the importance of testing the property of independent and identically distributed (i.i.d.) VaR-exceedances and propose a simple approach that explicitly tests for the presence of clusters in VaR-violation processes. Results from a simulation study indicate that our tests significantly outperform competing backtests in several distinct settings. In addition, the empirical analysis of a unique data set consisting of asset returns of an asset manager's portfolios underline the usefulness of our new backtests especially in times of market turmoil.

C1229: **Financial crisis and country risk rating: A value-at-risk approximation**
Presenter: **Pilar Grau**, Universidad Rey Juan Carlos, Spain
Co-authors: Luis Miguel Doncel, Jorge Sainz
The country risk ratings elaborated by various agencies such as Moody 's and Standard and Poors have been of singular importance and have been reflected almost immediately in the price of debt. Thus, the results of country risk analysis have been used as tools for decision making both before a loan, for example, to decide how much to pay or what should be the risk premium, and after the loan, in an attempt to monitor the ability to repay the debt. It is therefore clear that a better analysis of the time series of both country risk ratings and financial data, can be useful both for policy makers and investor. This research proposes an adaptation of the Value-at - Risk (VaR) to the prediction of the conditional variance of the exchange rate risk ratings for some countries representative of four different geographical areas. VaR is a widely used measure for market risk, and is defined as the maximum potential loss over a period of time within a stated confidence level. We propose an adaptation of VaR to calculate, not only for the losses, ie the negative tail, but also for profits as the size of the potential positive changes in risk ratings also relevant countries. To evaluate the predictions we use backtesting. The results show that country risk returns have an asymmetric distribution. This approach can be useful, not only for countries wishing to attract foreign investment, but also for those who are contemplating carrying out such investments.

C917: **Benchmarking of unconditional VaR and ES calculation methods: A simulation analysis with truncated stable distribution**
Presenter: **Takashi Isogai**, Bank of Japan, Japan
Value at Risk (VaR) and Expected Shortfall (ES) calculation methods are analyzed in terms of bias and stability against benchmarks computed from a fat-tailed parametric distribution. The daily log returns of the Nikkei-225 stock index are modeled by a truncated stable distribution. The VaR and ES values of the fitted distribution are regarded as benchmarks. The fitted distribution is also used as a sampling distribution; sample returns with different sizes are generated for the numerical simulations of the VaR and ES calculations. Two parametric methods assuming normal distribution and generalized Pareto distribution and two non-parametric methods, namely, historical simulation and kernel smoothing, are selected as the targets of this analysis. A comparison of the simulated VaR, ES, and the ES/VaR ratio with the benchmarks at multiple confidence levels reveals that the normal distribution approximation has a significant downward bias, especially in the ES calculation. The estimates by the other three methods are much closer to the benchmarks on average, although some of them become unstable with smaller sample sizes and at higher confidence levels. Specifically, ES tends to be more biased and unstable than VaR at higher confidence levels.

C901: **Comparison of copula-GARCH models within market risk estimation**
Presenter: **Ales Kresta**, VSB-TU Ostrava, Czech Republic
Co-authors: Tomas Tichy
Modeling of financial time series and subsequent risk estimation is a challenging activity of any entity active at financial markets. For example, financial time series are characterized by volatility clustering and heavy-tailed distributions of returns. Both these characteristics have a great influence for risk estimation. Especially when modeling multi-dimensional probability distribution is needed or the shocks in terms of extreme losses (or returns) in particular risk drivers are considered. In this paper we focus on the copula-GARCH models. The copula functions are the tool which allows one to model the dependence among individual risk drivers without specifying their marginal distributions. Various probability distributions (and copula functions) are assumed. On the other hand, GARCH models allow depicting of the volatility clustering. These joined models are examined on several datasets and the results are evaluated due to standard tests.

C1133: **Component value at risk models with countercyclical adjustments for improved economic performance**
Presenter: **Jaideep Oberoi**, University of Kent, United Kingdom
Co-authors: Evangelia Mitrodima
A family of dynamic quantile models of the CAViaR type is proposed that account for both low and high frequency changes in Value at Risk (VaR). Using daily returns on an index and a stock, we produce rolling one-day-ahead VaR forecasts out-of-sample for several benchmark models. We find that the proposed models are superior to or comparable with existing models in terms of exception ratios (actual number / expected number of exceedances). As both the new and several existing models are not rejected by standard statistical tests such as the Dynamic Quantile test, we conduct comparisons based on reasonable economic criteria. We find that the proposed models exhibit a superior trade-off between the mean level and the variance of the time series of VaR forecasts. The motivation for this comparison is to reduce the daily transaction cost from changes in required equity while keeping the overall capital cost low. Despite this, the average loss in excess of VaR on exceedance days is lower in our models than existing ones. This implies that firms need to sell less of their protfolios to meet risk management requirements following negative returns, hence improving their overall performance.

CS04 Room G15 CONTRIBUTIONS TO MULTIVARIATE GARCH MODELS Chair: Uwe Hassler

C948: **Latent factor models with conditional heteroskedasticity: Estimation and forecast**
Presenter: **Giorgio Calzolari**, University of Firenze, Italy
Co-authors: Gabriele Fiorentini, Gian Piero Aielli

A large latent factor model, in which the volatilities of common and idiosyncratic factors are conditionally heteroskedastic, is considered. Exact likelihood-based estimation methods are not available when conditional variances depend on past values of the latent factors (common or specific). Indirect estimation procedures are available in the literature. Some of them calibrate the score of a Kalman filter approximation with inequality constraints on the auxiliary model parameters. Others deal sequentially with univariate linear regressions and univariate Garch-type models; they proved to be almost as accurate as the previous, with a remarkable computational simplification and speed. Volatility forecasts require the application of a Metropolis algorithm. This can also be done with two different approaches. (1) Multiple-move approach, simpler and faster, with the possible disadvantage of higher rejection probability even with a good proposal distribution. (2) Single-move approach, more complicated and computationally slower, with the possible advantage of a higher efficiency (in statistical terms) due to a very low rejection probability. We fully discuss computational details of the methods and we report the results of an extensive Monte Carlo experiment with empirically realistic designs to evaluate the performance of estimators and forecasts. Finally, we present an application on a moderately large number of weekly and daily returns from the stock market.

C1052: **Granger-causal analysis of VARMA-GARCH models**
Presenter: **Tomasz Wozniak**, University of Melbourne, Australia

A VARMA-GARCH model is used to investigate Granger causality in conditional mean and conditional variances of time series. Parametric restrictions for the hypothesis of noncausality in conditional variances between two groups of variables, when there are other variables in the system as well, are derived. These novel conditions are convenient for the analysis of potentially large systems of economic variables. To evaluate hypotheses of noncausality, a Bayesian testing procedure is proposed. It avoids the singularity problem that may appear in the Wald test and it enables the assumption of the existence of higher-order moments of the residuals required in classical tests to be relaxed.

C1058: **Biofuels or financialization: Explaining the increased correlation between grains and crude oil prices**
Presenter: **Harriet Mugera**, University of Trento, Italy
Co-authors: Christopher Gilbert

The correlation between grains and crude oil price returns has increased dramatically over the past decade. Alternative explanations are the use of food commodities as biofuels feedstocks, the financialization of agricultural futures markets and the claimed predominance of common demand over idiosyncratic supply side shocks in the post-Lehman period. We use estimate and implement a modified version of the standard Dynamic Conditional Correlation (DCC) multivariate GARCH model to examine these alternatives. This model proves superior to both the Constant Conditional Correlation (CCC) and the standard DCC models. We conclude that both biofuels and financialization contributed to the observed increased correlations increased correlations between food commodity and energy markets.

C1082: **A multiple bi-dimensional BEKK model for portfolio volatility matrix estimation of the Italian stock market**
Presenter: **Andrea Pierini**, Roma Tre, Italy
Co-authors: Alessia Naccarato

A combination of bi-dimensional BEKK models is proposed for the estimation of the volatility matrix for the Markowitz stock portfolio. Each diagonal element of the volatility is estimated taking the variance given by a bi-dimensional BEKK model with variables stock i and market index. Each off-diagonal element of the volatility is estimated taking the covariance given by a bi-dimensional BEKK model with variables stock i and j. The model is applied to a subset of promising universe among the series of data regarding the prices of the best capitalized 150 shares traded on the Italian stock market between 1 January 1975 and 31 August 2011. To select an optimal stock portfolio, the volatility matrix of the returns is needed. Its diagonal and off-diagonal elements are estimated separately as stated above and then the efficient frontier given by the solution of the estimated Markowitz problem is simulated. So we find the optimal number of stocks, fractions and return in order to obtain the minimum risk portfolio. This approach gives a time dependent overall estimation of the stock variances-covariances solving the computational burden by the decomposition of the original problem without losing the strength of the BEKK.

C1192: **Credit spread volatility: Findings from the U.S. corporate bond market**
Presenter: **Guiyun Cheng**, Lund University, Sweden

The volatility of credit spreads with investment-grade ratings and maturities of 2, 5, and 10 years on the U.S. corporate bond market is examined by using the newly developed GARCH-MIDAS model with estimation based on simulated annealing. The study reveals that over the past ten years the U.S. credit spread volatility is clustering and time-varying with two spiky volatile periods and moves countercyclically. We find that the volatility degree and persistence vary with ratings and maturities and most volatility persistence is above 0.9. We also find that the impact of main economic indicators on credit spread changes is not all significant and indicators related to risk-free rate have more explanatory power than equity return. Besides, in the long-run the effects of business cycle indicators on credit spread volatility are more powerful than those of individual macroeconomic indicators. Our findings provide empirical evidence that credit spread volatility should be taken into account in corporate bond pricing and credit risk models.

CS111 Room B36 TIME SERIES ECONOMETRICS I Chair: Helmut Herwartz

C964: **Model averaging in predictive regressions**
Presenter: **Chu-An Liu**, National University of Singapore, Singapore
Co-authors: Biing-Shen Kuo

Model combination in a predictive regression is investigated. We construct the forecast combination by averaging across over all candidate models. The goal is to select weights to minimize the one-step-ahead mean squared forecast error (MSFE) over the set of all possible forecast combinations. We derive the MSFE of the model averaging estimator in a local asymptotic framework. We show that the optimal model weights which minimize the MSFE depend on the local parameters and the covariance matrix of the predictive regression. We propose a plug-in estimator of the optimal weights and use these estimated weights to construct the forecast combination. Monte Carlo simulations show that our averaging estimator has much lower MSFE than the weighted AIC, the weighted BIC, the Mallows model averaging and the Jackknife model averaging estimators. As an empirical illustration, the proposed methodology is applied to the predictability of stock returns.

C669: **Penalized likelihood for a non-Gaussian state space model**
Presenter: **Frank Magalhaes de Pinho**, IBMEC, Brazil
Co-authors: Glaura Conceicao Franco

A new family of non-Gaussian spaces models was recently proposed and it is interesting because a significant amount of probability distributions are contained in it and the likelihood can be analytically computed. Previous paper showed that inference procedures for this family work satisfactorily

well, one of its parameters, ω, which impacts the variability of the model, is generally overestimated, regardless of the estimation method used. It proposes a penalized likelihood function to reduce empirically the bias of the maximum likelihood estimator of ω. Monte Carlo simulation studies are performed to measure the reduction of bias and mean square error of the obtained estimators.

C1139: **A new methodological approach for the selection of the error distribution in finance**
Presenter: **Julien Hambuckers**, University of Liege, Belgium
Co-authors: Cedric Heuchenne
Since 2008 and its financial crisis, an increasing attention has been devoted to the selection of an adequate error distribution in risk models, in particular for Value-at-Risk (VaR) predictions. We propose a robust methodology to select the most appropriate error distribution candidate, in a classical multiplicative heteroscedastic model. In a first step, unlike to the traditional approach, we do not use any GARCH-type estimation of the conditional variance. Instead, we propose to use a recently developed nonparametric procedure: the Local Adaptive Volatility Estimation (LAVE). The motivation for using this method is to avoid a possible model misspecification for the conditional variance. In a second step, we suggest a set of estimation and model selection procedures tests based on the so-obtained residuals. These methods enable to assess the global fit of a given distribution as well as to focus on its behaviour in the tails. Finally, we illustrate our methodology on three time series (UBS stock returns, BOVESPA returns and EUR/USD exchange rates).

C885: **A CLT For martingale transforms with slowly varying second moments and the limit theory of the QMLE for conditionally heteroskedastic models**
Presenter: **Stelios Arvanitis**, AUEB, Greece
Co-authors: Alexandros Louka
A limit theorem to a normal limit is provided for the standardized sum of a martingale transform that holds even in cases where the second moments diverge at an appropriately slow rate. This extends relevant results with stable but non normal limits to the case of asymptotic normality, as well as results of asymptotic normality by allowing domains of non-normal attraction. In those cases the rate is slower that $\sqrt{n}$ and it contains information for the rate of divergence of the truncated second moments. A major application concerns the characterization of the rate and the limiting distribution of the the Gaussian QMLE in the case of GARCH type models. By extending the relevant framework we accommodate for the case of slowly varying and potentially diverging fourth moments for the innovation process as well as the possibility that the parameter lies on the boundary. The results are of potential interest to financial econometrics in view of the remaining leptokurtosis of the empirical conditional distributions of asset returns.

C896: **Test of martingale difference hypothesis using RBFNN approximations**
Presenter: **Jinu Lee**, Queen Mary University of London, United Kingdom
A test of martingale difference hypothesis based on radial basis function neural network (RBFNN) approximations is presented. In a related literature, a formal theoretical justification for the validity of the test has been verified and the test appears to have a superior power performance to some competing tests. Yet, there is a neglected point with regard to finding the parameter values of radial basis function for the test specification. In the usual manner, ad hoc data-dependent values are considered for the parameters of the activation function under a certain assumption. Hence, a new testing procedure is proposed to include estimating the parameters using an approximate moving least squares method. Since the method is not dependent on the properties of sample data, it can make the test more robust when applied to one realised economic data. A Monte Carlo study examines the finite sample performance of the new test in terms of size and power, and an empirical application is followed to demonstrate the usefulness of the new test.

CS97 Room B20 FINANCIAL ECONOMETRICS II Chair: Manfred Gilli

C891: **Risk choices of portfolio managers in an emerging market economy**
Presenter: **Belma Ozturkkal**, Kadir Has University, Turkey
Co-authors: Ali Akkemik
Asset markets in emerging markets exhibit high volatility and low liquidity. Under such circumstances, portfolio management becomes difficult. Among emerging markets, Turkey in particular has high turnover, even when compared to developed markets, and this makes it an appealing case to study the risk choices of portfolio managers. The determinants of risk choices and speculative preferences of portfolio managers for their own investments are analyzed. We use a unique data set from a survey of portfolio managers in fall of 2012. The sample consists of 72 employees of portfolio management companies. In the empirical section, we run simple OLS and probit models.

C991: **The impact of ECB macro announcements on bid-ask spreads of European blue chips**
Presenter: **Tobias Ruehl**, University of Duisburg-Essen, Germany
The impact of macroeconomic announcements by the ECB on the transaction costs of European stock markets measured by the bid-ask spread is investigated using intraday data for the period between July 2012 and May 2013. The event study type analysis reveals a significant impact of announcements by the ECB on spreads. These effects are strongest for announcements that comprise unexpected information or a change in interest rates. The spreads rise sharply during the minutes surrounding the interest rate or other macroeconomic announcements by the ECB. For the Euro area analysis the stocks of the German Dax 30 and the French CAC 40 have been used. The stocks of the FTSE 100 have been used to compare the results of the Euro area to non Euro area stock reactions. Several robustness checks have been applied to verify whether other factors such as normal daytime-dependent frictions or other macroeconomic announcements are driving the results.

C806: **Liquidity shocks and stock bubbles**
Presenter: **Ogonna Nneji**, University of Reading, United Kingdom
This study presents and empirically tests a simple framework that examines the effects of asset liquidity (the ease with which stocks are traded) and funding liquidity (the ease with which market participants can obtain funding) shocks on stock bubbles. Three key findings emerge from this research. First, negative liquidity shocks increase the probability of stock bubbles collapsing, with shocks to funding liquidity having a more prevalent effect on bubbles than asset liquidity shocks. Second, the effects of these two genres of liquidity shocks on bubbles have become more significant after the subprime mortgage meltdown-induced crisis. Last, liquidity shocks can provide warning signals of an impending bubble collapse.

C897: **Structural models for intraday liquidity resilience in the limit order book**
Presenter: **Efstathios Panayi**, UCL, United Kingdom
Co-authors: Gareth Peters, Jon Danielsson, Jean-Pierre Zigrand
Limit Order Book modelling poses an important Big Data challenge for econometricians. A key component of this study is understanding liquidity in the LOB. Liquidity measures in financial markets commonly reflect instantaneous snapshots of a market's liquidity and inherently lack the notion of resilience, i.e. the speed with which prices and/or volumes recover after a particular market event. We generalise the class of static liquidity measures to redefine them as dynamic measures, in order to also account for the notion of resilience in LOB liquidity. This is achieved through the development of a multivariate survival time model for a new liquidity resilience measure that we define. The flexible modelling framework we have

developed allows for the use of different liquidity measures, and the resulting liquidity resilience measure may thus capture different aspects of the stochastic structure of the LOB. In this model, we incorporate appropriate censoring of each marginal survival process of the liquidity measures intra-daily as well as common factors from endogenous variables affecting each asset. We also employ a parametric quantile regression framework, and estimate confidence intervals for the duration of liquidity droughts. We illustrate our new dynamic liquidity duration measures on several datasets.

CS113 Room G16 FINANCIAL APPLICATIONS III Chair: Herman van Dijk

C1089: **An empirical analysis of FX volatility surface dynamics through the global financial crisis**
Presenter: **David Peel**, Curtin University, Australia
Co-authors: Hiroaki Suenaga, Kurt Smith

Implied volatility surfaces have been studied extensively for various option markets including equities, FX and commodities. It is widely acknowledged that the surface is non-constant across moneyness and maturity dimensions. Specifically, empirical studies find that the implied volatility depends non-linearly on the moneyness and time to maturity. However, once the level is controlled for, the relative shape of the volatility surface is stable, even over the period of the 1987 market crash. We examine the volatility surface of FX options, using more recent data from 2005 to 2013. Our model specifies the implied volatility as a quadratic function of ATM volatility, spot rate, and forward swap points. The three factors together account for 65 percent of the daily variations in the implied volatilities across five deltas, five maturities, and over eight years of daily observations. The estimated volatility surface however is not stable over time; rather it exhibits statistically significant changes around the period of Global Financial Crisis. This finding provides a stark contrast to previous studies which report the relative shape of the surface is stable over time.

C699: **Calibrating low latency option analytics on multi-core CPUs**
Presenter: **Matthew Dixon**, University of San Francisco, United States
Co-authors: Mohammad Zubair

Low-latency real-time analytics feeds provide tick-by-tick implied volatilities and greeks from exchange data. In order for the implied volatility surface to exhibit the empirically observed leverage effect, a Heston model can be used to price the options. Because the solution to the Heston model exists in semi-analytic form, frequent robust calibration of the model is computationally prohibitive. This paper explores three parallelization approaches for the Heston Model calibration procedure deployed on a multi-core CPU cluster. The contributions are (i) reporting of benchmarks demonstrating hybrid shared and distributed memory parallelization techniques using python packages for robust calibration of the Heston model and (ii) demonstration of the model error reduction in being able to frequently recalibrate a Heston model to an OPRA data feed of US equity options. We evaluated the performance for our implementation on a cluster of 32 dual socket Dell PowerEdge R410 nodes providing 256 cores in total. We obtain a speedup of 65x against the sequential version of the calibration error function evaluation. Overall, we are able to reduce the entire calibration process time of the sequential implementation from 230 to under 10 seconds.

C860: **Diversification gains of the integrated Latin American market**
Presenter: **Carlos Castro**, Universidad del Rosario, Colombia
Co-authors: Nini Johana Marin

Financial integration has been pursued aggressively across the globe in the last fifty years; however, there is no conclusive evidence on the diversification gains (or losses) of such efforts. We quantify the effects of financial integration on diversification strategies for emerging market investors. In particular, we look at the case of the Integrated Latin American Market (MILA). We use dynamic conditional correlation models (DCC, DECO) as well as factor models to quantify co-movements across securities. Furthermore, we review different methodologies to quantify diversification that take into account the dynamic patterns of co-movements across securities. Our preliminary results suggest; first, that market integration has not had a significant impact on dependence across the securities at the country or industry level and second, that in order to exploit diversification gains in these emerging markets, portfolio allocation strategies need to consider jointly the country and industry effect.

C879: **Oil – stocks diversification: A new evidence from a dynamic copulas and high frequency data**
Presenter: **Krenar Avdulaj**, Academy of Sciences of the Czech Republic-UTIA, Czech Republic
Co-authors: Jozef Barunik

Oil is widely perceived as a good diversification tool for stock markets. To fully understand the potential, we propose a new empirical methodology which combines generalized autoregressive score copula functions with high frequency data, and allows us to capture and forecast the conditional time-varying joint distribution of the oil – stocks pair accurately. Our realized GARCH with time-varying copula yields statistically better forecasts of the dependence as well as quantiles of the distribution when compared to competing models. Using recently proposed conditional diversification benefits measure which take into account higher-order moments and nonlinear dependence, we document reducing benefits from diversification over the past ten years. Diversification benefits implied by our empirical model are moreover strongly varying over time. These findings have important implications for portfolio management.

C1154: **Estimation of the stressed securitisation rating transition matrices: A multi-state Markov approach**
Presenter: **Dario Cziraky**, Barclays Investment Bank, United Kingdom
Co-authors: Pim Sadleir

Regulatory capital requirements for ECAI-rated structured finance exposures are based on agency ratings of individual tranches. Consequently, stress testing of securitisation regulatory capital requires rating transition matrices of tranche ratings. The challenge is in the estimation of the stressed transition matrices that are used in calculation of the stressed capital. A common approach is to calculate transition matrix using data from a stress period such as the credit crisis. Such an approach, however, fails to align the stressed transition probabilities with specific macroeconomic scenarios, which determine the required level of stress. We propose a multi-state Markov (MSM) framework for stress testing securitisation rating transition matrices that allows conditioning of rating migration probabilities on multiple macroeconomic factors such as house price index, inflation, unemployment, and stock market indices. The model is estimated by maximum likelihood and we show how specific constraints can be imposed while testing for significance of included variables, estimated parameters, and model fit. The MSM model is then applied to estimation of conditional transition matrices for US RMBS tranches where we demonstrate how transition matrices can be statistically conditioned to specific economic scenarios including a mild stress and a severe stress scenarios.

Sunday 15.12.2013 16:50 - 18:55 Parallel Session N – ERCIM

ESI01 Room Beveridge STATISTICAL LEARNING Chair: Stefan Van Aelst

E415: **Structured sparsity in large scale vector autoregressive models**
Presenter: **George Michailidis**, University of Michigan, United States
The analysis of large vector autoregressive (VAR) models has received considerable attention in both the statistics and econometrics literature in recent years. Such models have been shown to improve forecasting performance over small VARs and provide deeper understanding of structural relationships among macroeconomic variables. The methods for analyzing large scale VAR models rely primarily on incorporating structural assumptions or prior beliefs in the modeling framework to tackle the "curse of dimensionality". The aim is to study the problem of forecasting and system identification in large dimensional VAR models under different structural assumptions on the system matrices. An M-estimation framework is adopted that can accommodate both (group) sparsity and/or low rank structure, propose a fast estimation algorithm and provide non-asymptotic upper bounds on the estimation and forecast errors. The performance of the proposed methodology is evaluated on simulated data and illustrated on a macroeconomic data set.

E605: **On consistency, robustness, and bootstrap of some regularized kernel methods**
Presenter: **Andreas Christmann**, University of Bayreuth, Germany
Regularized kernel methods including support vector machines play an important role in statistical machine learning theory. They have good statistical and numerical properties and have demonstrated their good generalization properties in many applications. One nice fact is that the overall error of regularized kernel methods can often be split up into the approximation error and the statistical error, such that both error terms can then be dealt with different mathematical methods, e.g. minimization over dense reproducing kernel Hilbert spaces, concentration inequalities, and oracle inequalities. After a short overview, some recent results on consistency and robustness properties of such regularized kernel methods are given. Nonparametric estimation of quantile functions and scale functions is considered. Furthermore, some asymptotical results on the empirical bootstrap of support vector machines are given. These bootstrap results can be helpful to draw statistical decisions based on such regularized kernel methods.

E1078: **BigData: Statistical techniques for processing massive, high-dimensional data**
Presenter: **Ping Li**, Rutgers University, United States
BigData has become a popular topic of research in recent years, largely owing to the ever increasing capability of data generation/collection, and the increasing complexity of the engineering systems. A prominent example is Web search and Information Retrieval. Practitioners have been using statistical machine learning tools for building models from billions of observations (for example, click-through data). Ultra-high-dimensional data representations are commonly adopted in practice. Novel algorithms for achieving effective dimensionality reduction and data reductions are in urgent demand. Parallel processing, cloud computing, data streaming are great research topics with fascinating applications. This talk will cover a series of work on probabilistic methods for solving important BigData problems, with applications in machine learning, information retrieval, data compression, and sparse recovery.

ES03 Room Deller RESAMPLING PROCEDURES FOR DEPENDENT DATA Chair: Andres M. Alonso

E057: **Local bootstrap for robust spectral analysis**
Presenter: **Valderio Anselmo Reisen**, DEST-CCE-UFES, Brazil
Co-authors: Glaura Franco, Celine Levy-Leduc
A bootstrap procedure on the periodogram ordinates is employed in conjunction with robust spectral analysis to improve the performance of periodogram based estimators. It is well known that the ordinary periodogram is heavily affected by series contaminated by outliers or presenting large variance. Thus the ordinary periodogram is replaced by a robust periodogram, obtained by minimizing an L_p-norm criterion when calculating the Fourier coefficients of a least square regression. Bootstrap replicates of this robust periodogram are calculated by sampling with replacement from a small neighborhood around each frequency point considered. Asymptotic properties of the proposed procedure are studied and finite-sample properties are analyzed through an extensive simulation study.

E065: **Subsampling inference for the autocorrelations of GARCH processes**
Presenter: **Agnieszka Jach**, Universidad Carlos III de Madrid, Spain
Co-authors: Tucker McElroy
A self-normalization for the sample autocorrelations of GARCH(1,1) process (and of the square-/absolute-value processes) is provided. In the asymptotic distribution of the sample autocorrelations of GARCH(1,1) there are three rates of convergence that depend on the index of the regular variation of GARCH(1,1) process, $\kappa > 0$. Consequently, there are three different limit distributions: when $\kappa > 4$ the limit is Gaussian with the standard $\sqrt{n}$-rate, but the asymptotic var-cov matrix is given by the generalized, instead of the standard, Bartlett formula; when $\kappa \in (2,4)$ the limit is stable, κ-dependent, with nonanalytic cdf, and the rate is roughly $n^{1-2/\kappa}$; when $\kappa < 2$, the sample autocorrelations converge in distribution. We introduce a self-normalized sample autocorrelation statistic, which is computable without knowledge of κ (in particular, we need not assume the finiteness of the fourth moment), and which converges to a nondegenerate distribution. The sampling distribution can then be approximated nonparametrically by subsampling, as the corresponding asymptotic distribution is still parameter-dependent. The subsampling-based confidence intervals for the process autocorrelations are shown to have satisfactory empirical coverage rates in a simulation study. The impact of subsampling block size on the coverage is assessed. The methodology is further applied to daily returns of CAC40 and FTSA100 indices, their squares and absolute values.

E263: **A goodness-of-fit test for stochastic volatility models**
Presenter: **Wenceslao Gonzalez-Manteiga**, University of Santiago de Compostela, Spain
Co-authors: Jorge P. Zubelli, Abelardo Monsalve-Cobis
A goodness-of-fit test based on empirical processes is proposed as a model diagnostic method for continuous time stochastic volatility models. More specifically, a marked empirical process is constructed from the representation in the state-space model form associated with a discretized version of the underlying process. Distributions of these processes are approximated using bootstrap techniques. Finally, an empirical application to the EURIBOR (Euro Interbank Offered Rate) data set is conducted for illustration.

E455: **Identification and estimation of general ARMA models**
Presenter: **Carlos Velasco**, Universidad Carlos III de Madrid, Spain
Co-authors: Ignacio Lobato
The aim is to introduce frequency domain procedures for performing inference in general time series linear models. They allow for possibly noninvertible and/or noncausal processes in the absence on information of these potential nonfundamentalness properties. Information from

higher order moments is used to achieve identification on the location of the roots of the AR and MA polynomials for non-Gaussian time series. A minimum distance estimator is proposed that optimally combines the information contained in second, third, and fourth moments. Contrary to existing estimators, the proposed estimator is consistent under general assumptions, and can be computed in one single step. For the standard causal and invertible ARMA model with non-Gaussian innovations, the estimator can be asymptotically more efficient than Gaussian-based procedures, such as the Whittle estimator. For cases where Gaussian-based procedures are inconsistent, such as noncausal or noninvertible ARMA models, the proposed estimator is consistent under general assumptions. To perform inference in finite samples a bootstrap algorithm is proposed that preserves higher order dependence. The proposed procedures also overcome the need of using tests for causality or invertibility.

E568: **Resampling copula under stochastic monotonicity**
Presenter: **Dominik Sznajder**, KU Leuven, Belgium
Co-authors: Irene Gijbels
Stochastic monotonicity concept defines very strong dependence of a bivariate random vector. In particular, it implies regression monotonicity and tail monotonicity, which further implies quadrant dependence. It is also called complete regression dependence, which refers to the underlying idea that all conditional quantiles are monotonic functions (in the same direction). The stochastic monotonic relation between random variables is of particular interest in financial, insurance and econometric studies. Therefore, a testing procedure is proposed which explores the fact that stochastic monotonicity is a feature of the underlying copula function of a random vector. The test statistic is a violation measure based on the empirical copula estimator and the proposed statistical inference relies on a resampling procedure from a smooth constrained copula estimator. It corresponds to a copula estimator under the null hypothesis and is based on local polynomial smoothing of the initial constrained estimator and on transforming its partial derivatives by a rearrangement operator.

ES97 Room Bloomsbury ESTIMATING AND MODELING FUNCTIONALS IN MULTISTATE MODELS Chair: Hein Putter

E103: **A competing risks approach for nonparametric estimation of transition probabilities in a non-Markov illness-death model**
Presenter: **Jan Beyersmann**, U Ulm, Germany
Competing risks model time to first event and type of first event. An example from hospital epidemiology is the incidence of hospital-acquired infection, which has to account for hospital discharge of non-infected patients as a competing risk. An illness-death model would allow to further study hospital outcomes of infected patients. Such a model typically relies on a Markov assumption. However, it is conceivable that the future course of an infected patient does not only depend on the time since hospital admission and current infection status, but also on the time since infection. We demonstrate how a modified competing risks model can be used for nonparametric estimation of transition probabilities when the Markov assumption is violated.

E276: **Computationally simple state occupancy probability estimates for multi-state models under interval censoring**
Presenter: **Andrew Titman**, Lancaster University, United Kingdom
Multi-state models for interval censored data are typically fitted under parametric assumptions. A desirable way of assessing the appropriateness of the model is to compare estimates of state occupancy probabilities with some non-parametric estimate. However, existing non-parametric estimates of state occupancy either make crude assumptions, leading to bias, or else are more computationally intensive than fitting the original parametric model. A computationally simple method for obtaining non-parametric estimates of the state occupation probabilities is proposed for progressive multi-state models where transition times between intermediate states are subject to interval censoring but the time of absorption is known up to right censoring. The method separates estimation of overall survival, using standard methods for survival data, and estimation of the conditional cumulative incidence of progression to a series of subsets of the state space performed using methods for current status competing risks data. The resulting estimates of state occupancy are unbiased, without requiring a Markov assumption, when the disease process and examination times are independent. An inverse visit-intensity weighed estimator is proposed for cases where the time to next examination depends on the last observed state.

E289: **Boosting multi state models**
Presenter: **Holger Reulen**, University of Goettingen, Germany
Co-authors: Thomas Kneib
Functional gradient descent boosting is a convenient estimation approach and well-established answer to variable selection and model choice in a variety of regression settings, e.g. linear, generalized linear, structured (geo)additive, survival or quantile regression models. We develop a boosting approach for multi state models where the negative log partial likelihood serves as the loss function to minimize. We derive a suitable working response for the functional negative gradient descent algorithms by determining the negative partial first derivative of the loss function. Ideas behind boosting as well as the link to multi-state models will be discussed, software implementation by R package 'gamboostMSM' will be introduced and motivational illustrations on simulation study results will be presented.

E477: **Methods for testing the Markov condition in the illness-death model**
Presenter: **Mar Rodriguez-Girondo**, University of Vigo, Spain
Co-authors: Jacobo de Una-Alvarez
The illness-death model is a particular multi-state model that provides a suitable representation of the individual's history when a unique intermediate event can be experienced before the main event of interest. Nonparametric estimation of transition probabilities in this and other multi-state models is usually performed through the Aalen-Johansen estimator under a Markov assumption. A process is Markov if, conditionally to the present, its future evolution is independent of the past. The aim is to provide nonparametric methods for checking the Markov condition in the illness-death model based on measuring the future-past association along time. Firstly, a local test of significance is proposed for zero future-past association at each time point t, measured through the Kendall's tau. Besides, two global methods are presented, resulting from summarizing the performance of the local test along a grid of t values. Firstly, a supremum-type test statistic is proposed, considering the Kendall's tau as a process indexed by t. Secondly, data-driven algorithms are explored to select the "best" t point with respect to a given criterion, specifically, maximizing the power. The finite sample performance of the new methods is investigated through simulations and the application to biomedical datasets.

E770: **Regression models for expected length of stay**
Presenter: **Hein Putter**, Leiden University Medical Center, Netherlands
Co-authors: Mia Klinten Grand
In an ageing society, it is crucial to be able to estimate healthy life expectancy and expected life years spent in disability, also conditional on the current age and health/disability status of an individual. Moreover, it is important to be able to assess the effects of individual characteristics, like gender and socio-economic status, and of behavioral characteristics, like dietary habits and smoking, on these quantities. A simple and effective regression method, based on pseudo-observations obtained from a multi-state model, is proposed to address this question. An illustration based on the AHEAD study is provided.

ES15 Room Senate ROBUSTNESS ISSUES IN COMPLEX STRUCTURED MODELS Chair: Andrea Cerioli

E385: **Monitoring methods of robust regression**
Presenter: **Anthony Atkinson**, London School of Economics, United Kingdom
Co-authors: Marco Riani, Andrea Cerioli
Methods for robust regression are often compared using the breakdown point. Very robust estimators, such as those from Forward Search (FS) and Least Trimmed Squares (LTS) algorithms, have an asymptotic breakdown point of 50% for very remote contamination. The properties of these estimators depend on the distance between the regression data and the outliers. Evaluation over a parametric path in the space of models provides a systematic study of estimators as the outliers and central data move from being far apart to close together. Properties of interest are the variance and squared bias of the estimators, together with power when used in the detection of outliers. For an example with 30% contamination the FS outperforms LTS. It will be argued that the seeming superior performance of FS is due to the flexibility of choice of the subset of observations used for parameter estimation; for 30% contamination around 70% of the observations are used for fitting, whereas the LTS uses half the observations. It is proposed to compare algorithms by monitoring their properties as their parameters are adjusted to allow for varying levels of contamination. For LTS we look at a series of fits as the trimming parameter decreases from 0.5 to 0. Likewise, for S and MM estimators, we vary the efficiency of the estimators and so the breakdown points. The procedure parallels the use of graphical output during the progress of the Forward Search.

E170: **Asymptotic analysis of the forward search**
Presenter: **Bent Nielsen**, University of Oxford, United Kingdom
Co-authors: Soren Johansen
The Forward Search is an iterative algorithm concerned with detection of outliers and other unsuspected structures in data. This approach has previously been suggested, analysed and applied for regression models. An asymptotic analysis of the Forward Search is made. The argument involves theory for a new class of weighted and marked empirical processes, quantile process theory, and a fixed point argument to describe the iterative element of the procedure.

E502: **On consistency factors and efficiency of robust S-estimators**
Presenter: **Francesca Torti**, Joint Research Center, European Commission, Italy
Co-authors: Marco Riani, Andrea Cerioli
The aim is to tackle the problem of obtaining the consistency factors of robust S-estimators of location and scale both in regression and multivariate analysis. Theoretical results are provided, proving new formulae for their calculation and shedding light on the relationship between these factors and important statistical properties of S-estimators. Computational advances are also given, suggesting efficient procedures which simply require the repeated computation of the inverse of the cdf of the Chi2 distribution so that hardly any time is lost to this calculation when computing S-estimates. In addition, when the purpose is to fix the efficiency of the scale estimator, it is possible to quantify to what extent the approximate algorithms which are currently available provide an acceptable solution, and when it is necessary to resort to the exact formulae. Finally, even if the work concentrates on S-estimates and Tukey's biweight loss function, the main results can be easily extended to calculate the tuning consistency factors for other popular loss function and other robust estimators.

E679: **A general framework for robust fitting of hidden Markov models**
Presenter: **Antonello Maruotti**, University of Southampton, United Kingdom
Hidden Markov models (HMMs) have been widely used to model a wide range of time-dependent heterogeneous data. The areas of application of HMMs range from biology and medicine to physics, economics and marketing. On the one hand these models can be applied to data where observations originate from heterogeneous groups and group membership is not known, and on the other hand to provide approximations for multimodal distributions. In most applications the HMMs parameters are estimated by maximum likelihood via the well-known Baum-Welch algorithm. It is well known, however, that the maximum likelihood estimator can be very sensitive to outliers in the data. In fact, even a single outlier can completely ruin the maximum likelihood estimator. Suggested solutions for robustification against outliers (in a general latent class framework) are of various kinds. Among others, we include to add a noise component or an improper constant uniform, trimming or down-weighting observations. We compare robust methods for the analysis of time dependent data. Several methods based on an HMMs specification, assuming the existence of a latent process which follows a finite-state first-order Markov chain and affects the distribution of the response variable, are specified. We show that robust approaches work comparably well to the maximum likelihood estimator when there are no outliers and outperform it under outlier-contaminated simulation schemes.

E791: **A test for approximate stochastic order based on trimming**
Presenter: **Eustasio del Barrio**, Universidad de Valladolid, Spain
Co-authors: Pedro C. Alvarez Esteban, Juan A. Cuesta-Albertos, Carlos Matran
Stochastic ordering among distributions has been considered in a variety of scenarios. Economic studies often involve research about the ordering of investment strategies or social welfare. However, as noted in the literature, stochastic orderings are often a too strong assumption which is not supported by the data even in cases in which the researcher tends to believe that a certain variable is somehow smaller than other. Instead of considering this rigid model of stochastic order we propose to look at a more flexible version in which two distributions are said to satisfy an approximate stochastic order relation if they are slightly contaminated versions of distributions which do satisfy the stochastic ordering. The minimal level of contamination that makes this approximate model hold can be used as a measure of the deviation of the original distributions from the exact stochastic order model. Our approach is based on the use of trimmings of probabiliy measures. We discuss the connection between them and the approximate stochastic order model and provide theoretical support for its use in data analysis. We also provide simulation results and a case study for illustration.

ES89 Room Montague STATISTICS AND DATA ANALYSIS FOR INDUSTRY AND ENGINEERING Chair: Antonio D'Ambrosio

E611: **Modeling of increasing age- and state-dependent degradation processes: a new closed form Markovian solution**
Presenter: **Massimiliano Giorgio**, Second University of Naples, Italy
Co-authors: Maurizio Guida, Gianpaolo Pulcini
A new continuous Markovian model for increasing degradation processes is presented, in which the increments of degradation over disjoint time intervals are non-independent, but depend both on the current age and on the current state of the unit under study. Unlike other increasing Markovian processes with non-stationary and non-independent increments recently proposed in the literature, the proposed process is mathematically tractable and does not introduce undesirable approximations due to time and/or state discretization. Indeed, for such a new process, the conditional distribution of the degradation growth over a generic time interval, given the current state and age of the unit, can be formulated in closed form. The main properties of the proposed process are discussed and maximum likelihood estimators of the model parameters are developed. Finally, an application relative to a real set of degradation data is used to illustrate the estimation procedure and the capability of the proposed process in a real applicative context.

E819: **Penalized decomposition of Kα1, Kα2 X-ray diffraction profiles**
Presenter: **Johan de Rooi**, Erasmus Medical Center, Netherlands
Co-authors: Amarante Bottger, Paul Eilers
X-ray diffraction is a method used to identify and characterize crystalline materials. A typical diffraction pattern consists of intensities (photon counts) versus the diffraction angle 2θ. The data are a composition of two identical, but slightly shifted patterns of different intensity. The shift is caused by the X-ray source and can be calculated. The proposed procedure combines decomposition of the two patterns and noise removal in one step, where the noise obeys Poisson statistics. The procedure relies on the composite link model, which models so-called 'indirect observations'. The expected counts $E(y)$ are the sum of two components: $E(y) = \mu = \breve{\gamma} + \bar{\gamma}\tau = \exp(\breve{B}\beta) + \exp(\bar{B}\beta)\tau$ with τ defining the relative intensity of the second component. The vector of coefficients β is the same for both components, the B-spline basis $\bar{B}$ realizes the shift of the second component. Because the system is ill-conditioned, regularization is required. By penalizing the differences between adjacent spline coefficients the model becomes estimable and the estimated profiles are smooth. Using P-splines, results in the penalized composite link model. Estimation is performed using IRWLS, and the model is optimized using the AIC.

E1234: **Synthetic sampling technique for categorical database**
Presenter: **Roberta Siciliano**, University of Naples Federico II, Italy
Co-authors: Massimo Aria, Filomena Mauriello
The aim is to study a new synthetic sampling technique for categorical data. The problem of imbalance has got more and more emphasis in recent years. Imbalanced data sets exist in many real-world domains, such as oil-spill detection, network intrusion detection, fraud detection. In these domains, what we are really interested in is the minority class rather than the majority class. Thus, we need a fairly high prediction for the minority class. It has been largely reported that, whatever standard classification method is chosen, such a failure occurs in non-trivial learning problems. The focus is at the data level. These solutions alter the original class distribution, driving the bias towards the minority or positive class. In this way these strategies can be applied in any learning system, since they act as a pre-processing phase, allowing the learning system to receive the training instances as if they belonged to a well-balanced data set. We propose a novel Synthetic Sampling Technique for Categorical database (SSTC). SSTC applied to several highly and moderately imbalanced data sets shows improvement in prediction performance on the minority class and overall improved F-values.

E550: **Stochastic modelling of degradation in civil structures**
Presenter: **Iunio Iervolino**, Univ of Naples Federico II, Italy
Co-authors: Massimiliano Giorgio, Eugenio Choccarelli
The study deals with the stochastic modeling of degrading civil structures. The two, typical, categories of degradation phenomena considered are: (i) cumulative deterioration due to damage produced by earthquakes, that are modeled as point overloads, and (ii) continuous, age-related, deterioration of structural characteristics. The cumulative degradation process due to earthquakes is described via a compound Poisson process. Degradation increments are assumed to be independent and identically distributed non-negative variables. Aging is modeled via a continuous stochastic process with non-negative, independent, and stationary increments. The deterioration process is finally formulated as the sum of the two considered contributions. Two couples of modeling solutions are implemented and compared: (1) Gamma distribution for the degradation increments and Gamma process for the continuous deterioration; (2) Inverse Gaussian distribution for the degradation increments and Inverse Gaussian process for the continuous deterioration. In both the cases earthquakes occurrence is modeled via a Homogeneous Poisson Process. In this framework, closed-form approximations for life-cycle structural reliability assessment are derived. Examples referring to a simple structural system illustrate potential applicability of the proposed approach within the performance-based earthquake engineering framework.

E1232: **Investigating sensitivity analysis methods for application to an offshore wind operations and maintenance cost model**
Presenter: **Rebecca Martin**, Industrial Doctorate Centre for Offshore Renewable Energy, United Kingdom
Co-authors: Iraklis Lazakis, Sami Barbouchi
A tool has been developed by EDF R&D to calculate the cost of operations and maintenance of offshore wind farms. The issue addressed in this investigation is the uncertainty of factors that affect the cost of O&M. Key drivers of the cost of O&M and availability (the aggregated percentage of time that the wind farm is available to produce energy when the wind resource is present), such as weather windows and patterns of failures in turbines is still unknown. Whilst simplistic, locally based sensitivity analysis has been conducted for other O&M models, none have been applied to this tool. To identify the most appropriate sensitivity analysis method to use, an extended Morris method and a variance based method have been applied. This was done using MATLAB and sensitivity analysis framework software SimLab to generate a sample file of inputs and to conduct the Monte Carlo analysis on the result files. The results can be used to direct the efforts of further investigations into the uncertainties of aspects offshore wind farms.

ES24 Room Athlone STATISTICAL QUALITY CONTROL: THEORY AND PRACTICE Chair: Fernanda Figueiredo

E422: **Robust methods in acceptance sampling – PART II**
Presenter: **Elisabete Carolino**, ESTeSL-IPL, Portugal
Co-authors: Isabel Barao
In the statistical quality control of a production process the focus is either on the process itself with application of Statistical Process Control or on its frontiers, with application of Acceptance Sampling (AS) and Experimental Design. AS is used to inspect either the process output (final product) or the process input (raw material). The purpose of the design of a sampling plan is to determine a course of action that, if applied to a series of lots of a given quality, and based on sampling information, leads to a specified risk of accepting/rejecting them. Thus AS yields quality assurance. The classic AS by variables is based on the hypothesis that the observed quality characteristics follow the Gaussian distribution (treated in classical standards). This is sometimes an abusive assumption that leads to wrong decisions. AS for non- Gaussian variables, mainly for variables with asymmetric and/or heavy tailed distributions, is a relevant topic. When we have a known non-Gaussian distribution we can build specific AS plans associated with that distribution. Alternatively, we can use the Gaussian classical plans with robust estimators of location and scale. The problem of determining AS plans by variables for Extreme Value distributions (Weibull and Frechet) with unknown shape parameter is addressed. Classical plans, specific plans and plans using the robust estimates for location and scale are determined and compared.

E498: **Strategies to reduce the probability of a misleading signal**
Presenter: **Patricia Ramos**, CEMAT IST Universidade Tecnica de Lisboa, Portugal
Co-authors: Manuel Cabral Morais, Antonio Pacheco
Standard practice in statistical process control is to run two individual charts, one for the process mean and another one for the process variance. The resulting scheme is known as a simultaneous scheme and it provides a way to satisfy Shewhart's dictum that proper process control implies monitoring both location and dispersion. When we use a simultaneous scheme, the quality characteristic is deemed to be out-of-control whenever a signal is triggered by either individual chart. As a consequence, the misidentification of the parameter that has changed can occur meaning that a shift in the process mean can be misinterpreted as a shift in the process variance and viceversa. These two events are known as misleading signals

(MS) and can occur quite frequently. We discuss sufficient conditions to achieve values of PMS smaller than 0.5 and explore alternative control statistics to decrease the PMS. Informative examples concerning output processes in the univariate/multivariate settings are supplied.

E767: **Monitoring non-industrial processes**
Presenter: **Sotiris Bersimis**, University of Piraeus, Greece
Statistical Process Control techniques have been developed and used successfully in manufacturing for many decades. During the last decade, the basic techniques of Statistical Process Control have been gradually used effectively in non-manufacturing processes. We study the problem of process monitoring of non-industrial processes. Specifically, we propose techniques for monitoring the performance of a service process. The new techniques are applied to data from a commercial firm.

E861: **The STATIS methodology in statistical quality control**
Presenter: **Adelaide Maria Figueiredo**, Faculdade de Economia da Universidade do Porto and LIAAD INESC Porto, Portugal
Co-authors: Fernanda Otilia Figueiredo
In real situations, the evaluation of the global quality of either a product or a service depends on more than one quality characteristic. In order to monitor multivariate processes and identify the variables responsible for changes in the process, we shall use methods of multivariate data analysis, like the PCA (Principal Components Analysis) and the STATIS (Structuration des Tableaux a Trois Indices de la Statistique) methodology - a three-way data analysis method.

E862: **The total median and the total range statistics in phase I control charts**
Presenter: **Fernanda Otilia Figueiredo**, Faculdade de Economia da Universidade do Porto and CEAUL, Portugal
Co-authors: Maria Ivette Gomes, Subha Chakraborti
The classic procedures of inference developed under the assumption that the observations come from a normal population may be inappropriate if we have some disturbances in the data. It is thus important to find efficient and robust estimators of the process parameters. For instance, the sample mean, the sample standard deviation and the sample range, the most common location and scale estimators, are non-robust to deviations from the normality assumption. Although there are several robust location and scale estimators, in this paper we shall revisit the total median, the total range and some modified versions of the sample standard deviation, due to their simplicity of computation, and we shall consider their use in phase I control charts as well as in the development of non-parametric control charts. These estimators are related with the bootstrap sample associated to the observed sample, and can be considered in practice, as an advisable robust alternative to the usual ones. These estimators seem to be less influenced by the extreme observations than the usual estimators, and for this reason are more efficient than the usual ones for contaminated distributions or in the presence of outliers.

ES26 Room Gordon MATRIX ALGORITHMS AND HPC FOR LARGE SCALE DATA ANALYSIS II Chair: Costas Bekas

E809: **Identifying users of social networks from their data footprint: An application of large-scale matrix factorizations**
Presenter: **Michael Berry**, University of Tennessee, United States
Co-authors: Denise Koessler
Users of social media interact with the network and its users. Each interaction creates network specific data between the engaged users and the chosen social avenue. Over time, these engagements accumulate to describe the user's social fingerprint, an identity which encapsulates the user's persona on the network. The agglomeration of this information showcases the user's activity on the social network and establishes a traceable social fingerprint. These fingerprints can be tracked and stored as large matrices representing the quantity and occurrence of observed user behavior. We seek to apply large-scale matrix factorization techniques to establish the signature component vector of a social network user's identity. The preliminary results presented will demonstrate that a user's social fingerprint is both quantifiable and identifiable on a social network throughout time.

E519: **Data fusion based on coupled matrix and tensor factorizations**
Presenter: **Evrim Acar**, University of Copenhagen, Denmark
Co-authors: Anders Lawaetz, Morten Rasmussen, Rasmus Bro
In many disciplines, data from multiple sources are acquired and jointly analyzed for enhanced knowledge discovery. For instance, in metabolomics, biological fluids such as blood or urine are measured using different analytical techniques that are complementary in terms of the chemicals they can detect. Data sets obtained using different analytical methods can be jointly analyzed to identify the chemicals/biomarkers or patterns related to certain diseases. However, while we can collect huge amounts of data from multiple sources, we are still lacking the data mining tools for data fusion. Data fusion is a challenging task since data are often incomplete, heterogeneous, i.e., in the form of higher-order tensors and matrices, and large- scale. With a goal of addressing these challenges, data fusion is formulated as a coupled matrix and tensor factorization problem tailored to automatically reveal common (shared) and individual (unshared) components. In order to solve the coupled factorization problem, a gradient-based all-at-once optimization algorithm is used, which easily extends to coupled analysis of incomplete data sets. It is demonstrated that the proposed approach provides promising results in joint analysis of metabolomics measurements of plasma samples of a group of colorectal cancer patients and controls.

E775: **On incremental deterministic methods for dominant space estimation for large data sets**
Presenter: **Kyle Gallivan**, Florida State University, United States
Co-authors: Christopher Baker, Paul van Dooren
Low-rank matrix factorizations are important in several contexts to characterize and analyze, for example, large data sets and dynamical systems. For very large problems, operation counts are not as significant as data accesses and the effect of memory system parameters such as bandwidth and parallelism. For some applications with extreme time and/or size constraints, the matrix cannot even be effectively stored, giving the opportunity to access the matrix only by one or a few read-only passes with relatively limited read/write storage in which to produce the factorization. For several years low-rank incremental methods have been investigated in various contexts from very small signal processing devices to very large-scale data-driven problems. We revisit these low-rank incremental methods in light of recent progress on characterizing their capabilities, the situations in which they operate, and related algorithms.

E1260: **Fast projection methods for robust separable nonnegative matrix factorization**
Presenter: **Stephen Vavasis**, University of Waterloo, Canada
Co-authors: Nicolas Gillis
Nonnegative matrix factorization (NMF) has become a widely used tool for analysis of high-dimensional data. Although the NMF problem is NP-hard, certain special cases, such as the case of separable data with noise, are known to be solvable in polynomial time. We propose a very fast successive projection algorithm for this case. Variants of this algorithm have appeared previously in the literature; our principal contribution is to prove that the algorithm is robust against noise and to formulate bounds on the noise tolerance. A second contribution is an analysis of a

preconditioning strategy based on semidefinite programming that significantly improves the robustness against noise. We present computational results on artificial data and on a simulated hyperspectral imaging problem.

E900: **Experiments with randomized algorithms in the text to matrix generator toolbox**
Presenter: **Eugenia-Maria Kontopoulou**, University of Patras, Greece
Co-authors: Dimitris Zeimpekis, Efstratios Gallopoulos
Randomized matrix techniques developed in recent years to handle Big Data are gaining much attention. It is a fact that, in many applications, algorithms of this type are attracting the interest of theorists and practitioners as deterministic algorithms become impractical for massive data sets. Because of these developments, it becomes desirable to provide users with an environment that allows rapid familiarization with existing algorithms of this type and prototyping and evaluation of new ones. We will describe the incorporation of some of these randomized techniques in the Text to Matrix Generator toolbox (http://scgroup20.ceid.upatras.gr:8000/tmg/) and we will discuss their performance in Information Retrieval tasks through appropriate experiments.

ES31 Room Jessel STOCHASTIC SYSTEMS BIOLOGY Chair: Colin Gillespie

E247: **Stochastic simulation of chemically reacting systems using multi-core CPUs**
Presenter: **Colin Gillespie**, Newcastle University, United Kingdom
In recent years, computer simulations have become increasingly useful when trying to understand the complex dynamics of biochemical networks, particularly in stochastic systems. In such situations stochastic simulation is vital in gaining an understanding of the inherent stochasticity present, as these models are rarely analytically tractable. However, a stochastic approach can be computationally prohibitive for many models. A number of approximations have been proposed that aim to speed up stochastic simulations. However, the majority of these approaches are fundamentally serial in terms of central processing unit (CPU) usage. A novel simulation algorithm is discussed which utilises the potential of multi-core machines. The algorithm partitions the model into smaller sub-models. These sub-models are then simulated, in parallel, on separate CPUs. We demonstrate that this method is accurate and can speed-up the simulation by a factor proportional to the number of processors available.

E721: **A method for exploring the structural robustness of stochastic biological networks**
Presenter: **Chris Barnes**, University College London, United Kingdom
Synthetic biology can be defined as applying engineering approaches, such as part standardisation, abstraction and mathematical modelling, to the design and engineering of biologically based parts, novel devices and systems. One obstacle is that constructed devices have to function despite changes to external environmental conditions and internal fluctuations in binding, production and decay rates. Another challenge is that genetic circuits are fundamentally stochastic due to the low number of species involved. The ability of a system to perform some specified input-output behaviour under changes in parameters, network topology and model miss-specification is known in control theory as robust performance. The addition of positive and negative feedback loops can confer structural robustness, where the system becomes insensitive to the parameter values. We present a measure of robustness for stochastic biochemical models that coincides with the Bayesian model evidence. This allows us to exploit Bayesian model selection for the investigation of structural robustness and we employ an approximate Bayesian computation scheme to calculate the relative evidence of stochastic models governed by the chemical master equation. By jointly inferring the topology and parameters for a specified input-output behaviour we can rank designs based on their structural robustness.

E904: **Statistical inference of cellular behavior from heterogeneous single-cell data**
Presenter: **Heinz Koeppl**, ETH Zurich, Switzerland
Mathematical methods together with measurements of single-cell dynamics provide unprecedented means to reconstruct intracellular processes that are only partly or indirectly accessible experimentally. To obtain reliable reconstructions the pooling of recordings from several cells of a clonal population is mandatory. The population's considerable cell-to-cell variability originating from diverse sources poses novel computational challenges for such process reconstruction. A formulation in terms of a stochastic mixed-effect model invites the elimination of random extrinsic parameters from the inference problem. We discuss such inference techniques and explain how they vary according to the available data formats, such as live-cell data or population snapshot data. We apply the methods to data of two gene expression systems in budding yeast.

E036: **Fast approximate Bayesian computation for discretely observed Markov models using a factorised posterior distribution**
Presenter: **Theodore Kypraios**, University of Nottingham, United Kingdom
Many modern statistical applications involve inference for complicated stochastic models for which the likelihood function is difficult or even impossible to calculate, and hence conventional likelihood-based inferential techniques cannot be used. In such settings, Bayesian inference can be performed using Approximate Bayesian Computation (ABC). However, in spite of many recent developments to ABC methodology, in many applications the computational cost of ABC necessitates the choice of summary statistics and tolerances that can potentially severely bias estimation. We propose a new 'piecewise' ABC approach suitable for discretely observed Markov models that involves writing the posterior density of the parameters as a product of factors, each a function of only a subset of the data, and then using ABC within each factor. The approach has the advantage of side-stepping the need to choose a summary statistic and it enables a stringent tolerance to be set, making the posterior 'less approximate'. We investigate two methods for estimating the posterior density based on ABC samples for each of the factors and discuss their advantages and disadvantages. We illustrate the piecewise ABC approach for three examples; in each case, the approach enables 'exact matching' between simulations and data and offers fast and accurate inference.

E316: **The delayed-acceptance particle marginal random walk Metropolis**
Presenter: **Chris Sherlock**, Lancaster University, United Kingdom
Co-authors: Andrew Golightly, Alex Thiery
Delayed-acceptance particle marginal Metropolis-Hastings (daPMMH) algorithms can be applied when it is computationally expensive to calculate an unbiased stochastic approximation to the true posterior, but a computationally cheap deterministic approximation is available. An initial accept/reject stage uses the cheap approximation in the MH acceptance probability; proposals which are accepted at this stage are then subjected to a further accept-reject step which corrects for the error in the approximation. Since the expensive approximation is only evaluated for proposals which are accepted at the first stage, the cost of the algorithm is reduced. The full advantage of DA algorithms has, however, not been tapped. We focus on the random walk Metropolis (RWM) and consider the daPMRWM, developing theoretical expressions for limiting efficiency and acceptance rates in high dimension which inform a practical strategy for algorithm tuning. Both in theory and in a simulation study based upon a Markov jump process we also find that the optimal scaling of proposals is substantially larger for the DA algorithm than for the equivalent algorithm without the DA step. The simulation study verifies other heuristic properties suggested by our theory and for the example considered our optimised daPMRWM algorithm is nearly an order of magnitude more efficient than the equivalent optimised PMRWM algorithm.

ES40 Room Court DEPENDENCE MODELLING: THEORY AND PRACTICE Chair: Ivan Kojadinovic

E023: **On idempotent D-norms**
Presenter: **Michael Falk**, University of Wuerzburg, Germany
Replacing the spectral measure by a random vector Z allows the representation of a multivariate max-stable distribution with standard negative margins via a norm, called D-norm, whose generator is Z. We investigate the set of all generators in detail. This approach towards multivariate extreme value distributions entails the definition of a multiplication type operation on the set of D-norms leading to idempotent D-norms. We characterize the set of idempotent D-norms. Iterating the multiplication provides a track of D-norms, whose limit exists and is again a D-norm. If this iteration is repeatedly done on the same D-norm, then the limit of the track is idempotent.

E246: **When uniform weak convergence fails: Empirical processes for dependence functions via epi- and hypographs**
Presenter: **Johan Segers**, University of Louvain-la Neuve, Belgium
Co-authors: Axel Buecher, Stanislav Volgushev
For copulas whose partial derivatives are not continuous everywhere on the interior of the unit cube, the empirical copula process does not converge weakly with respect to the supremum distance. This makes it hard to verify asymptotic properties of inference procedures for such copulas. To resolve the issue, a new metric for locally bounded functions is introduced and the corresponding weak convergence theory is developed. Convergence with respect to the new metric is related to epi- and hypoconvergence and is weaker than uniform convergence. Still, for continuous limits, it is equivalent to locally uniform convergence, whereas under mild side conditions, it implies L^p convergence. Even in cases where uniform convergence fails, weak convergence with respect to the new metric is established for empirical copulas and tail dependence processes. No additional assumptions are needed for tail dependence functions, and for copulas, the assumptions reduce to existence and continuity of the partial derivatives almost everywhere on the unit cube. The results are applied to obtain asymptotic properties of minimum distance estimators, goodness-of-fit tests and resampling procedures.

E095: **Generalized additive modelling for conditional copulas**
Presenter: **Valerie Chavez-Demoulin**, University of Lausanne, Switzerland
Co-authors: Thibault Vatter
A generalized additive modelling framework to t conditional copulas is proposed. This framework allows copula-based models to depend on covariates in a parametric, semi-parametric or non-parametric way. The method uses penalized log-likelihoods maximized through specific Newton-Raphson type algorithms. Simulations designed to study the numerical properties of the method indicate that it performs well. A real dataset is considered as an application.

E741: **Nonparametric tests for constancy of a copula**
Presenter: **Axel Buecher**, Ruhr-Universitaet Bochum, Germany
Co-authors: Ivan Kojadinovic
The modeling and estimation of stochastic dependencies by copulas has attracted an increasing attention over the last years in various fields of application. Most of the available statistical inference procedures are based on the implicit assumption that the copula of a multivariate time series remains constant over time. We present methods that allow us to test for this assumption, both for marginals that are identically distributed and for those whose distribution may change over time. The asymptotics of the test statistics rely on a general weak convergence result for the sequential empirical copula process based on time series data. In the case of observations that are strongly mixing, it is shown how a generalized multiplier bootstrap allows us to get approximate access to the quantiles of the limiting distribution.

E029: **Computational statistics in copula modeling: Why and how**
Presenter: **Marius Hofert**, ETH Zurich, Switzerland
Co-authors: Martin Maechler
In the highly non-linear world of copulas, computational challenges (high dimensions, numerical precision, parallel computing) arise quickly. More importantly, numerical issues may not be obvious from computed results and therefore such results may actually be misleading. This has even led to wrong conclusions drawn from erroneous simulation studies in the recent literature and poses a particular challenge for practitioners who want to apply copula modeling in business practice. After giving a short motivation on the subject, we briefly present a new R package which aims at simplifying statistical simulation studies and which carefully deals with important tasks such as parallel computing, seeding, catching of warnings and errors, and measuring run time.

ES52 Room Bedford ADVANCES IN COMPOSITIONAL DATA ANALYSIS AND RELATED METHODS I Chair: Karel Hron

E859: **Zeros in compositional count data sets: The geometric Bayesian-multiplicative imputation**
Presenter: **Josep Antoni Martin-Fernandez**, University of Girona, Spain
Co-authors: Javier Palarea-Albaladejo, Karel Hron
A vector of counts is collected when the outcomes in each a number of identical and independent trials can fall in any of a fixed number of mutually exclusive categories. The analysis of this type of data is typically based on multinomial and Dirichlet models. These methods show important difficulties in those cases where the total sum of the vector is not of interest, suggesting that log-ratio methodology is a more general approach to the analysis of compositional count data. However, log-ratio methodology requires a preprocessing step where a proper treatment of the zeros is applied. In our approach we assume that zero values refer to unobserved positive values that may have been observed with a larger number of trials. According to this assumption, a treatment involving a Bayesian inference of the zero values and a multiplicative modification of the non-zero values is considered. This treatment offers the possibility to use valuable information from the personal knowledge of the analyst or from the data previously collected. In this way, an adequate strategy could be to consider a prior equal to the sample estimation of the expected value in the variable with a leave-one-out scheme. In other words, when we deal with a vector we are assuming that the other samples are our prior information. Because this approach is fully coherent with the properties of the geometric distribution of probability, we call it a geometric bayesian-multiplicative (GBM) replacement. An additional advantage of GBM replacement is that the posterior estimates do not depend on the number of categories, that is, conforms to the representation invariance principle. The performance of this technique is illustrated using real and simulated data sets.

E080: **N-way partial least squares for compositional data**
Presenter: **Michele Gallo**, University of Naples Orientale, Italy
Partial least squares (PLS) is a method for building regression models between independent and dependent variables. When a set of independent variables is measured on several occasions, the samples can subsequently be arranged in three-way arrays. In this case N-way partial least squares (N-PLS) can be used. N-PLS decomposes three-way array of independent variables and establishing a relation between the three-way array of independent variables and the array of dependent variables. Sometimes, the set of independent variables are parts of the same whole, thus each observation consists of vectors of positive values summing to a unit, or in general, to some fixed constant. When these data, known as compositional

data (CoDa), are analyzed by N-PLS, it is necessary to take into account the specific relationships between the parts that compositions are made of. The problems that potentially occur when one performs a N-way partial least squares analysis on compositional data are examined. A strategy based on the log-ratio transformations is suggested.

E389: **Left-censoring problems in compositional data sets**
Presenter: **Javier Palarea-Albaladejo**, Biomathematics and Statistics Scotland, United Kingdom
Co-authors: Josep Antoni Martin-Fernandez
Multivariate data representing part of a whole, usually called compositional data in statistics, are common in many fields of science-say chemical concentrations, food compositions, activity patterns, abundance of species, and so on. Their distinctive feature is that there is an inherent relationship between the parts, as they only convey relative information. This implies that ordinary statistical techniques are not generally adequate, and specialized statistical methods are required. Progress in compositional data analysis has been mostly driven by the log-ratio methodology. In this context, the focus is on imputation methods to deal with parts that include unobserved values falling below a certain threshold. Left-censoring problems often arise in experimental compositional data, either in the form of rounded zeros or values below certain limits of detection. We discuss the issue of introducing a number of methods that take into consideration the particular principles and nature of compositional data, from univariate non-parametric procedures to multivariate model-based approaches. These methods are supported with software that facilitates their practical implementation.

E863: **Compositional entropies in model based clustering**
Presenter: **Marc Comas-Cufi**, Universitat de Girona, Spain
Co-authors: Gloria Mateu-Figueras, Josep Antoni Martin-Fernandez
To cluster samples from a finite mixture density a model-based technique is recommended. Initially the cluster method selects the total number of mixture components. Assuming that the number of groups is less than or equal to the mixture components, the method hierarchically combines the components using an entropy criterion applied to posterior probabilities. Typically the criterion is based on the well-known Shannon entropy. In this work we show that any model-based cluster analysis applied to any type of data, not necessarily compositional, is enriched when the vector of posterior or individual's conditional probabilities (group memberships) are considered as elements of the simplex. In this way, entropy criterion based on the Aitchison distance and the compositional Kullback-Leibler divergence are introduced. Here the Aitchison distance between two compositions is defined as the Euclidean distance between the corresponding log-ratio coordinates. The compositional Kullback-Leibler divergence consists in a modification of the Jeffreys divergence. Both measures fulfil the two main compositional principles: scale invariance and subcompositional coherence. The performance of these compositional entropy criteria are evaluated and illustrated using real and simulated data sets.

E940: **Classical and robust principal component analysis for density functions using Bayes spaces**
Presenter: **Klara Hruzova**, Palacky University, Czech Republic
Co-authors: Karel Hron, Matthias Templ, Peter Filzmoser, Gianna Serafina Monti
Probability densities are used to describe relative contributions of Borel sets of real line to the overall probability of support of a random variable. This is one of the reasons for considering density functions as functional compositional data with a constant-sum-constraint equal to one. For this type of data it was necessary to construct a new Hilbert space, called Bayes space, that accounts for specific properties of densities as functional data carrying relative information. Recently, there have been several approaches to analyze statistically density functions in an ordinary way, e.g. using functional principal component analysis. The aim of this contribution is to show a different approach for the mentioned case of functional principal component analysis where densities are transformed using a functional version of centred logratio transformation (clr) for compositional data. Finally, the methodology will be applied to real-world data set from official statistics.

ES60 Room Torrington DESIGN OF EXPERIMENT IN INDUSTRY Chair: Kalliopi Mylona

E082: **Designed experiments for semi-parametric models and functional data with a case-study in tribology**
Presenter: **David Woods**, University of Southampton, United Kingdom
Co-authors: Christopher Marley, Susan Lewis
Experiments with functional data are becoming ubiquitous in science and engineering with the increasing use of online monitoring and measurement. Each run of the experiment results in the observation of data points that are realised from a smooth curve. Although large quantities of data may be collected from each run, it may still only be possible to perform small experiments with a limited number of runs. We describe a statistical methodology for an example from Tribology, concerning the wear-testing of automotive lubricants. Here, we investigated how lubricant properties and process variables affected the shape of a functional response measuring wear. Novel techniques were developed for the initial design of a screening study where the levels of some of the factors could not be set directly. A two-stage semi-parametric modelling approach was applied, using a varying coefficient model and principal components. New methods for the design of follow-up experiments for such models were also developed and applied. In addition to the new methodology, we present conclusions from the case study about which factors had substantial effects, and how they influenced the shape of the wear curves.

E508: **Optimal design of choice experiments with partial profiles**
Presenter: **Peter Goos**, Universiteit Antwerpen, Belgium
Co-authors: Daniel Palhazi Cuervo, Roselinde Kessels, Kenneth Sorensen
To limit the choice task complexity in choice experiments involving many attributes, some authors recommend the use of partial profiles. In the special case where respondents are indifferent between different alternatives, D-optimal designs are known for many scenarios. For more realistic cases in which respondents are not indifferent, no D-optimal designs are known, and experimenters have to resort to heuristic design construction algorithms. A new algorithm is presented to construct locally optimal partial-profile designs as well as Bayesian optimal partial-profile designs in these more realistic cases. The excellent performance of the new algorithm is demonstrated by showing that its results match the known D- optimal designs in case respondents are indifferent.

E659: **Minimum setup criterion to select designs with hard-to-change factors**
Presenter: **Andre Pinho**, Federal University of Rio Grande do Norte, Brazil
Co-authors: Linda Ho, Carla Vivacqua
Considering fierce competition, companies always seek strategies to improve their time to market. Although new simulation technologies are available, physical prototype testing still remains an important step in the product development cycle. It is common that prototypes are composed of several parts, with some more difficult to assemble than others. Moreover, typically, there is only one piece available of each part and a large number of different setups. Under these conditions, designs with randomization restrictions become attractive approaches. Considering this scenario, a new and additional criterion to construct split-plot type designs is presented. Designs with a small number of setups of the more difficult parts, which are especially useful for screening purposes, are discussed. The proposal of the minimum setup (MS) criterion is driven by real applications and represents an alternative to directly accommodate practical or operational requirements, disregarded by other criteria. The process

of sequentially minimizing the number of setups at each stratum with hard-to-change factors generates a design called MS split-plot type design. It is shown that although the new criterion can be used to differentiate designs with the same word length pattern, it is an independent criterion. Catalogs of selected 32-run split-split-plot and split-split-split-plot designs are presented.

E680: **Optimal experimental design in the mine industry**
Presenter: **Elvira Delgado-Marquez**, University of Castilla-La Mancha, Spain
Co-authors: Mariano Amo-Salas, Jesus Lopez-Fidalgo
The process of jam formations during the discharge by gravity of granular material stored in a 2D silo is analyzed. The aim is twofold. On the one hand, we propose optimal designs using the D-optimality criterion and a compound criterion with a cost function. Our results reveal that the efficiency of the design used by the experimenters may be improved dramatically. On the other hand, the estimation of the unknown parameters has been done using Least Squares and Maximum Likelihood. The experimenters use wrongly Least Square Estimates while the data come from an exponential distribution. Simulations show that the variance and the bias of the Maximum Likelihood Estimate is lower than the variance of the Least Square Estimate. Finally, a sensitivity analysis with respect to the most important parameter is performed.

E722: **Efficient blocking of orthogonal arrays with two and three levels based on design classification criteria**
Presenter: **Haralambos Evangelaras**, University of Piraeus, Greece
Co-authors: Christos Peveretos
When orthogonal arrays are used in practical applications, it is often difficult to perform all the designed runs of the experiment under homogeneous conditions. The arrangement of factorial runs into blocks is usually an action taken to overcome such obstacles. However, an arbitrary configuration might lead to spurious analysis results. The nice properties of orthogonal arrays are taken into consideration and an effort to establish an effective criterion for arranging experimental runs into blocks of the same size is made. This criterion is based on the so called J - characteristics of the corresponding design. Such "optimal" arrangements are given for orthogonal arrays with two and three levels and various run sizes.

ES62 Room Chancellor's APPLICATIONS OF FUNCTIONAL DATA ANALYSIS Chair: Alicia Nieto-Reyes

E037: **Assessing the effect of stroke on motor control using Bayesian penalized function-on-scalar regression**
Presenter: **Jeff Goldsmith**, Columbia University, United States
Co-authors: Tomoko Kitago
Common population-level effects of stroke on motor control are considered while accounting for possible subject-level idiosyncratic effects. Motor performance for each subject is assessed through repeated planar reaching motions from a central point to eight pre-specified targets arranged on a circle. We observe the kinematic data for hand position as a bivariate function of time for each reach. Our goal is to estimate the bivariate function-on-scalar regression with subject-level random functional effects and a functional principal components decomposition of residual curves; covariates of interest are stroke severity and target number. We express fixed effects, random effects and FPCA basis functions using penalized splines. Parameters are jointly estimated in a Bayesian context, and we implement a computationally efficient approximation algorithm using variational Bayes.

E272: **Kriging complex data for environmental applications: A critical inquiry into the functional approach**
Presenter: **Piercesare Secchi**, Politecnico di Milano, Italy
Co-authors: Alessandra Menafoglio
In many geophysical studies, data are available only in a few locations, due to the environmental constraints which prevent the collection of data at a higher spatial density. Hence, spatial prediction techniques, such as kriging, assume a key role. When dealing with the interpolation of a space-time field, classical geostatistics approaches may raise practical issues since the computational costs of performing spatial prediction becomes soon unaffordable, due to the very high number of data. To address this problem, we recently developed a novel Universal Kriging predictor for elements of a Hilbert Space: we proposed a Functional Data Analysis approach which is often justified by the typical high resolution of the data along the time coordinate. Moving from there, we investigate the potentiality of such an approach when dealing with complex data, through the analysis of its application to environmental case studies. Temperature curves embedded in L^2 and functional compositional data handled through Aitchison geometry are only two examples. The critical points to be addressed in real analyses are illustrated, with particular regard to the choice of the Hilbert space data are assumed to belong to, possibly exploiting their differential properties (e.g., by working in Sobolev spaces).

E366: **Improving the quality of biscuit production: Application of different functional data analysis approaches**
Presenter: **M. Carmen Aguilera-Morillo**, University of Granada, Spain
Co-authors: Ana M. Aguilera
The aim is to predict a binary response variable Y (quality of cookies) from a functional predictor X (resistance of dough), that is equivalent to the problem of classification of the sample curves in the two groups defined by the response categories (good or bad). Three different functional methodologies are applied: functional principal component logit regression, functional linear discriminant analysis based on functional partial least squares, and componentwise classification on the logit regression model. A P-spline approximation of the sample curves is proposed to improve the classification ability of these models and to suitably estimate the relationship between the quality of cookies and the resistance of dough. Inference results on the functional parameters and related odds ratios are obtained using the asymptotic normality of the maximum likelihood estimators under the classical regularity conditions. Specific interpretations of the principal components and the functional parameter are exposed.

E668: **Depth-based classification of multivariate functional data**
Presenter: **Mia Hubert**, KU Leuven, Belgium
Co-authors: Pieter Segaert
Recently, multivariate functional halfspace depth (MFHD) has been proposed as a new tool to study multivariate functional data. This depth function allows us to estimate the central tendency and the variability of multiple sets of curves. The application of this depth function to supervised classification is presented. The functional k-nearest neighbour classifier is compared to classifiers such as the within maximum depth, the minimum distance to the mean, and a classifier based on the DD-plot. Also new classifiers are proposed. By means of a simulation study, all these classifiers are compared. In particular, their behavior is studied in the presence of outlying curves.

E419: **Nonparametric functional methods for online signature recognition**
Presenter: **Gery Geenens**, University of New South Wales, Australia
The problem of automatic signature recognition has attracted attention for a long time, since signatures are well established in our everyday lives as the most common means of personal identification. There is therefore a clear need for accurate and reliable signature recognition systems, and it is no surprise that many digital procedures aiming at discriminating forgeries from genuine signatures have been proposed in biometrics, pattern recognition and engineering literature. However, it turns out that even those methods which claim to be functional or dynamic are actually based on a finite number of parameters describing the temporal evolution of some considered characteristics, like pen pressure or azimuth for example. Never has the problem been addressed from a purely functional point-of-view, that is, keeping the whole "signature- function" as the random object

of central interest. In contrast, the aim is mainly to use modern functional data analysis tools, like nonparametric functional regression ideas, to develop and implement an efficient signature recognition system, and to check whether this exclusively statistical method is able to match the currently used pattern recognition and biometrical methods in terms of simplicity, ease of implementation and, of course, efficiency at exposing fakes.

ES68 Room Woburn INFERENCE FOR COUNTING PROCESSES Chair: Paula R. Bouzas

E121: **Hypotheses testing: Poisson versus self-exciting**
Presenter: **Serguei Dachian**, Universite Blaise Pascal, France
Co-authors: Yuri Kutoyants
The problem of hypotheses testing is considered with a basic simple hypothesis: the observed sequence of points corresponds to a stationary Poisson process with known intensity. The alternatives are stationary self-exciting point processes. We consider one-sided parametric and one-sided nonparametric composite alternatives and construct locally asymptotically uniformly most powerful tests in both cases. The results of numerical simulations of the tests are also presented.

E249: **A dynamic contagion process with applications to finance & insurance**
Presenter: **Angelos Dassios**, London Scool of Economics, United Kingdom
The motivation is default contagion in financial mathematics. A point process is developed to model defaults. This model is inspired by stochastic intensity models, but also by branching theory. The model goes beyond Cox processes (ie, processes with a stochastic intensity). We will give a branching process definition and survey important results and applications to credit risk. Moreover, we will use the process as a claim arrivals process in general insurance and develop results concerning the probability of ruin. Applications of the mathematical tools developed for the case of delayed claims will also be discussed.

E598: **Semiparametric partially linear varying coefficient models with recurrent events**
Presenter: **Xin He**, University of Maryland, United States
Co-authors: Xingqiu Zhao, Xingwei Tong
The focus is on semiparametric regression analysis of panel count data, which arise naturally when recurrent events are considered. Such data frequently occur in medical follow-up studies and reliability experiments, for example. To explore the nonlinear interactions between covariates, a class of partially linear models is proposed with possibly varying coefficients for the mean function of the counting processes from recurrent events. The functional coefficients are estimated by B-spline function approximations. The estimation procedures are based on maximum pseudo-likelihood and likelihood approaches and they are easy to implement. The asymptotic properties of the resulting estimators are established, and their finite-sample performance is assessed by Monte Carlo simulation studies. The value of the proposed method is also demonstrated by the analysis of a cancer data set, where the new modeling approach provides more comprehensive information than the usual proportional mean model.

E429: **Characterizing self-exciting processes: Simulations driven by phase-type distributions**
Presenter: **Nuria Ruiz-Fuentes**, University of Jaen, Spain
Co-authors: Paula R. Bouzas
A self-exciting process is a very general type of counting process. It evolves with aftereffects, i.e. its intensity is also a stochastic process which depends on the past counts as well as any other influences. The counting processes are usually studied assuming a given stochastic model for the intensity. A conditional orderly self-exciting process is considered, without any given structure or prior information about the dependence on its past. In this framework, little research has been conducted to determine the intensity or count-conditional intensity process in order to characterize the self-exciting one. That is the aim based on a former proposed point estimator for the count-conditional intensity given any number of occurrences merely from several observed sample paths of the counting process. Due to it, the self-exciting process is not only characterized but the estimations of its probability mass function, its mean and the mean of its intensity are also derived. Phase-type distributions gather a wide range of positive-valued distributions so they can naturally play the role of an intensity. Simulations of a self-exciting process with intensity driven by a phase-type distribution illustrate the calculation of the point estimation proposed and the statistics mentioned.

E595: **Inference of a Cox process by FDA**
Presenter: **Paula Bouzas**, University of Granada, Spain
Co-authors: Nuria Ruiz-Fuentes
Different inference methods for counting processes and in particular for a Cox process (CP) can be found in the literature. However, concerning the main aim, most of them are focused on estimation and few on prediction and, regarding the counting process, it is usually assumed that its intensity has a given structure which is often unrealistic. FDA is a technique which allows modelling a stochastic process from raw data. Therefore, as the intensity of a CP is also stochastic, FDA can be used to model it and accordingly, the CP itself. This way of addressing the estimation of the intensity makes it possible to achieve the inference of a CP only from observed sample paths. Moreover, it is possible to estimate and forecast the mean and intensity processes and some counting and time statistics of a CP or a compound one. Furthermore, goodness- of-fit tests can be built to assess the intensity of observed data from any of those counting processes.

ES82 Room 349 STATISTICS IN ACTUARIAL SCIENCES Chair: Tim Verdonck

E208: **Tail dependence versus measures of association**
Presenter: **Vali Asimit**, City University London, United Kingdom
The analysis of concomitant extreme events has been simply connected to the concept of tail dependence. The problem becomes more acute in the presence of asymptotic independence, the issue that is addressed. It is intended to relate the degree of tail independence to some well-known measures of association such as Spearman's Rho and Kendall's Tau. Some numerical examples are provided in order to illustrate the author's view.

E228: **Life insurance policy termination and survivorship**
Presenter: **Emiliano Valdez**, Michigan State University, United States
Co-authors: Jeyaraj Vadiveloo, Ushani Dias
There has been some work leading to the development of statistical models in understanding the mortality pattern of terminated policies. However, there is scant literature on the empirical evidence of the true nature of the relationship between survivorship and persistency in life insurance. In a life insurance contract, when policies terminate due to voluntary non-payment of premiums, there is a possible hidden cost resulting from mortality antiselection. This refers to the tendency of policyholders who are generally healthy to select against the insurance company by voluntarily terminating their policies. We explore the empirical results of the survival pattern of terminated policies, using a follow-up study of the mortality of those policies that terminated from a portfolio of life insurance contracts. The data has been obtained from a major insurer who tracked the mortality of their policies withdrawn, for purposes of understanding the mortality antiselection, by obtaining their dates of death from the Social

Security Administration office. We modeled the time until a policy lapses and its subsequent mortality pattern. We find some evidence of mortality selection and we subsequently examined the financial cost of policy termination.

E471: **Sensitivity analysis of internal risk models**
Presenter: **Andreas Tsanakas**, City University London, United Kingdom
Co-authors: Pietro Millossovich
An internal risk model can be seen as consisting of (i) a random vector of *risk factors* and (ii) a *business structure*, understood as a real-valued function mapping the risk factors to an output of interest, such as aggregate loss. In typical insurance applications, the risk factors can be high dimensional and the business structure non-linear. Such models are often *black boxes*, in the sense that the only information available to a model reviewer is a Monte-Carlo sample from the risk factors and output. Sensitivity analysis is thus necessary to provide the reviewer with a better understanding of the model. A sensitivity measure is proposed that assesses the impact of each factor on the aggregate risk, as reflected by a distortion risk measure. It is then shown how the sensitivity measure can be calculated on a smooth approximation to the business structure, obtained by standard local linear regression methods. The proposed methodology is demonstrated through examples from life and non-life insurance and its applicability to the measurement of parameter risk is discussed.

E555: **Continuous chain ladder: Reformulating a classical insurance problem**
Presenter: **Maria Dolores Martinez-Miranda**, University of Granada, Spain
Co-authors: Jens Perch Nielsen, Stefan Sperlich, Richard Verrall
The single most important number in the accounts of a non-life insurance company is likely to be the estimate of the outlying liabilities. Since non-life insurance is a major part of our financial industry (amounting to up to 5% of BNP in western countries), it is perhaps surprising that mathematical statisticians and experts of operational research (the natural experts of the underlying problem) have left the intellectual work on estimating this number to actuaries. This important problem is established in a vocabulary accessible to experts of operations research and mathematical statistics and it can be seen as an open invitation to these two important groups of scholars to join this research. A number of new methodologies and approaches is introduced to estimating outstanding liabilities in non-life insurance. In particular, the classical actuarial technique is reformulated as a histogram type of approach and this classical technique is improved by replacing this histogram by a kernel smoother.

E747: **Micro level stochastic loss reserving for general insurance**
Presenter: **Katrien Antonio**, University of Amsterdam and KU Leuven, Netherlands
To meet future liabilities general insurance companies will set–up reserves. Predicting future cash–flows is essential in this process. The vast literature on (stochastic) loss reserving methods concentrates on data aggregated in run–off triangles. However, a triangle is a summary of an underlying data set with the development of individual claims. We refer to this data set as 'micro–level' data. Using the framework of Position Dependent Marked Poisson Processes and statistical tools for the analysis of recurrent events, a data set is analyzed with liability claims from a European insurance company. We use detailed information of the time of occurrence of the claim, the delay between occurrence and reporting to the insurance company, the occurrences of payments and their sizes and the final settlement. Our model specifications are (semi)parametric. We calibrate the model to historical data and use it to project the future development of open claims. An out–of–sample prediction exercise shows that we obtain detailed and valuable reserve calculations. For the case study, the micro–level model outperforms the results obtained with traditional loss reserving methods for aggregate data. A discussion of ongoing research in this area, and a sketch of opportunities in the domain of disability insurance is given.

ES103 Room Holden BEYOND CONDITIONAL INDEPENDENCE Chair: Tamas Rudas

E250: **Nonparametric testing of conditional independence**
Presenter: **Wicher Bergsma**, London School of Economics, United Kingdom
Random variables Y and Z are conditionally independent given X if, knowing the value of X, information about Y does not provide information about Z. Conditional independence relations are the building blocks of graphical models, applications of which include information extraction, speech recognition, computer vision, decoding of low-density parity-check codes, modeling of gene regulatory networks, gene finding and diagnosis of diseases, and graphical models for protein structure. A new method to test conditional independence is discussed. Existing literature on the topic is usually restricted to the normal and categorical cases, but recently nonparametric testing has also received a fair amount of attention. Our method is also nonparametric, but differs from previous ones in that it is based on the following decomposition, which gives some advantages. Denote by $\Psi_{g,h}(x)$ the conditional covariance between $g(Y)$ and $h(Z)$ given $X = x$. Conditional independence of Y and Z given X holds if and only if the following two conditions hold: 1. For arbitrary g and h, $E\Psi_{g,h}(X) = 0$; 2. For arbitrary g and h, $E\Psi_{g,h}(x)$ does not depend on x. Each condition can be tested separately. However, there are some technical difficulties which we explain and for which we provide a solution.

E786: **Learning structure in discrete nested Markov models**
Presenter: **Ilya Shpitser**, University of Southampton, United Kingdom
Co-authors: Thomas Richardson, James Robins, Robin Evans
Causal and probabilistic models with latent variables can be naturally represented by means of acyclic directed mixed graphs (ADMGs), containing directed arrows representing "direct causation", and bidirected arrows representing "unobserved confounding." Models represented by ADMGs can contain independence constraints which hold after a generalized conditioning operation called fixing, where a joint distribution is divided by a conditional rather than a marginal distribution as with ordinary conditioning. Fixing, which can be viewed as a non-causal version of an intervention, is represented graphically by removing arrows pointing to a fixed node in the graph. Recently, a factorization, and a number of Markov properties were given that capture post-fixing constraints giving a statistical model called the nested Markov model. A parameterization of discrete nested Markov models is described, and it is shown how this parameterization in an algorithm that learns representations of hidden variable DAG models from data. This algorithm is used to give evidence for a characterization of equivalence classes of mixed graphs of 4 nodes, where all graphs in an equivalence class agree on all independences after certain sequences of fixings (a generalization of Markov equivalence in DAGs). Finally, it is shown that an alternative log-linear parameterization gives a concise representation of hidden variable DAGs by allowing higher order interactions to be set to zero.

E467: **Extended relational models**
Presenter: **Anna Klimova**, IST Austria, Austria
Co-authors: Tamas Rudas
An extension of the relational model class to the set of non-negative distributions on a contingency table is described. Relational models generalize log-linear models and thus models for conditional independence. They are generated by a class of subsets of cells in a table, with model matrix equal to the indicator matrix of those subsets. Any distribution in a relational model is a common root of a set of binomials, derived from a kernel basis of the model matrix. For positive distributions, the binomial representations obtained from different kernel bases are equivalent, and a distribution in the model can be factored according to the model matrix. However, as demonstrated, if zero cell parameters are allowed, the equivalence fails and the factorization may or may not exist. An extended relational model is defined as the set of all non-negative distributions on the table whose cell

parameters satisfy the binomial representations for all possible choices of the kernel basis, and is thus a variety of the polynomial ideal associated with the model matrix. The support of any distribution in the extended model is isomorphic to a facial set of a polyhedron determined by the model matrix. For every distribution in the model whose support is a proper subset of the sample space there exists a sequence of strictly positive distributions in the model that converges to it. Therefore, the extended model is the closure of the relational model in the topology of the pointwise convergence. The extended models for various models of independence are shown.

E440: **On Bayesian network structure learning methods based on an integer linear programming approach**
Presenter: **Milan Studeny**, Institute of Information Theory and Automation of ASCR, Czech Republic

A special type of a conditional independence structure is a Bayesian network (BN) structure, ascribed to an acyclic directed graph. The idea of the geometric approach to statistical learning a BN structure is to represent it by a certain vector so that common scoring criteria, e.g. BIC or BDE, become affine functions of the vector representative. One such a vector representative is the characteristic imset, introduced in 2010, but there are other options. The aim is to make an overview of recent attempts at application of integer linear programming methods (and software packages) to learning a BN structure by finding a global maximum of a scoring criterion.

E528: **Independence on non-product spaces**
Presenter: **Tamas Rudas**, Eotvos Lorand University, Hungary
Co-authors: Anna Klimova

Independence, conditional independence and their generalizations are the basis of graphical models, and in many applications, similar structures on non-product spaces are of interest. For example, in feature selection problems discussed in machine learning, not all combinations of the presence and absence of features may be observable or meaningful. Relational models generalize log-linear models for sample spaces that are not Cartesian products. Ways of defining independence and conditional independence in these cases are investigated. A classical example which applies to sample spaces where in each cell, at least one feature is present, is Aitschison-Silvey independence, which assumes that there is no overall effect, that is, there is no common effect that would apply to all cells. A similar generalization of conditional independence is discussed. In this context, multiplicative models for features and for variables, defined as lack or presence of a feature, are fundamentally different. The former ones are curved exponential families and cannot be written to contain the overall effect. In these models, although cell probabiliies are products of effects associated with features present in the cell, these effects are not equal to the marginal probabilities of the relevant features, as is the case when the sample space is a Cartesian product. The models can be defined using generalized odds ratios, out of which exactly one is non-homogenous. When the model contains no overall effect, the observed marginal totals in the maximum likelihood estimates under multinomial sampling are not preserved and multinomial and Poisson likelihoods lead to different estimates.

ES125 Room Russell BAYESIAN SEMI- AND NONPARAMETRIC MODELLING IV Chair: Yongdai Kim

E578: **Convergence of posterior distribution in inverse problems with non-Gaussian noise**
Presenter: **Natalia Bochkina**, University of Edinburgh, United Kingdom

The aim is to present the rate of contraction of the posterior distribution for inverse problems where the noise is non-Gaussian, namely it belongs to the exponential family of distributions. A key feature of these models is that the information about the unknown function comes not only from the mean, as under Gaussian errors, but also from the variability or variance. This property may lead to improved rate of contraction compared to the rate under Gaussian errors, and may even lead to self-regularization of an inverse problem. The result is obtained as a limit of a finite-dimensional model with growing parameter dimension which is illustrated in two frameworks: discretization on a grid and an orthogonal basis decomposition.

E435: **Bayesian tensor decompositions for categorical data with many levels**
Presenter: **James Johndrow**, Duke University, United States
Co-authors: Anirban Bhattacharya, David Dunson

Bayesian tensor decompositions provide a scalable approach to inference for high-dimensional contingency tables. These models induce shrinkage on the coefficients in a saturated log linear model, and hence provide an alternative to explicitly parameterizing these interactions. Motivated by contingency tables having very many categories, the posterior concentration behavior is carefully studied, allowing the number of categories to grow with sample size n. We characterize the class of decomposition priors that achieve a near parametric convergence rate under mild conditions on sparsity and rank. Motivated by this new theory, new classes of tensor decomposition priors are developed, which lead to efficient posterior computation and induce excellent concentration including when the true model is close to a sparse log-linear interaction model. To assess finite sample properties, a comprehensive simulation study is conducted, showing substantially improved results over competitors in both computational time and ability to infer the dependence structure. Applications to marketing data are provided as an illustration. Applications beyond contingency tables to multivariate density estimation, and recommender system problems, are discussed.

E760: **A model for reversible Markov chains**
Presenter: **Konstantina Palla**, University of Cambridge, United Kingdom
Co-authors: David Arthur Knowles, Zoubin Ghahramani

A model for reversible Markov chains is presented. We use a completely random measure to construct graphs with weighted edges. More specifically, we use a gamma process to sample the nodes in the graph and the links between them. By enforcing symmetry to make the edges undirected we define a prior over random walks on graphs that results in a reversible Markov chain. The resulting prior over infinite transition matrices is closely related to the hierarchical Dirichlet process but enforces reversibility. A reinforcement scheme has recently been proposed with similar properties before, but the de Finetti measure is not characterised. We take the alternative approach of explicitly constructing the mixing measure, which allows more straightforward inference at the cost of no longer having a closed form predictive distribution. We use our process to construct a reversible HMM which we apply to real world datasets.

E239: **Sampling sigma-stable Poisson-Kingman mixture models**
Presenter: **Stefano Favaro**, University of Torino and Collegio Carlo Alberto, Italy
Co-authors: Maria Lomeli, Yee Whye Teh

Sigma-stable Poisson-Kingman models form a rich class of nonparametric priors including as special cases most of the priors currently known in the Bayesian nonparametrics such as the two parameter Poisson-Dirichlet process and the normalized generalized Gamma process. We study Markov chain Monte Carlo methods for sampling Bayesian nonparametric mixture models with a sigma-stable Poisson-Kingman mixing measure. In particular, we introduce a conditional slice sampling and a marginal Polya-urn sampler. The former is based on a stick-breaking representation for sigma-stable Poisson-Kingman models that we develop expressly. To obtain the marginal sampler we study two novel representations for the exchangeable random partition induced by a sigma-stable Poisson-Kingman mixture model.

Monday 16.12.2013 08:45 - 10:25 Parallel Session O – CFE

CSI03 Room Senate MEASURING SYSTEMIC RISK Chair: Monica Billio

C259: **Banking, insurance and the real economy: Interconnectedness and systemic risk**
Presenter: **Mardi Dungey**, University of Tasmania, Australia
Co-authors: Matteo Luciani, David Veredas
Systemic risk in financial markets is usually measured with reference to banks alone. However, the financial sector performs both maturity and risk transformation functions for the real economy via the banking and insurance sectors. And the interlinkages between all sectors mean that systemic episodes in financial markets can have significant impact on output and employment. Measurement of systemic risk where banking, insurance and real economy sectors are included shows the role of intervention policies in decreasing the systemic risk in the financial sector by reducing its interconnectedness with the real economy. Applied to the S&P500 firms in the US, we show that systemic risk in the banking sector peaked with the top of the US housing cycle, but in the insurance sector continued to rise until September 2008. The announcement and final ratification of TARP dramatically decreased the systemic risk associated with interconnectedness, compared with the sustained measures provided by methods applying a metric based on capital shortfall.

C255: **Sovereign, bank and insurance credit spreads: Connectedness and system network**
Presenter: **Monica Billio**, University of Venice, Italy
Co-authors: Mila Getmansky, Dale Gray, Andrew Lo, Robert Merton, Loriana Pelizzon
Macrofinancial risk has become increasingly important over time as global markets have become increasingly more connected. We apply several econometric measures of connectedness based on Granger-causality networks to the changes of sovereign risk of European countries and credit risk of major European, U.S., and Japanese banks and insurers to investigate the evolution of these connections. Credit risk for banks and insurers is measured using a version of the Merton Model (Contingent Claims Analysis) applied to risk-adjusted balance sheets. We highlight connections among banks, insurers, and sovereigns by quantifying the effects of risk transmission within and across countries and financial institutions.

C267: **Measuring credit risk in a large banking system: Econometric modeling and empirics**
Presenter: **Andre Lucas**, VU Amsterdam, The Netherlands
Co-authors: Xin Zhang, Bernd Schwaab
Two new measures for financial systemic risk are computed based on the time-varying conditional and unconditional probability of simultaneous failures of several financial institutions. These risk measures are derived from a multivariate model that allows for skewed and heavy-tailed changes in the market value of financial firms' equity. Our model can be interpreted as a Merton model with correlated Levy drivers. This model incorporates dynamic volatilities and dependence measures and uses the overall information on the shape of the multivariate distribution. Our correlation estimates are robust against possible outliers and influential observations. For very large cross sectional dimensions, we propose an approximation based on a conditional Law of Large Numbers to compute extreme joint default probabilities. We apply the model to assess the risk of joint financial firm failure in the European Union during the financial crisis.

CS108 Room G15 APPLIED ECONOMETRICS III Chair: Jean-Pierre Urbain

C625: **Evaluating real time VAR fasts based on survey-data assisted estimators**
Presenter: **Frieder Mokinski**, ZEW Mannheim, Germany
The aim is to compare and extend several methods aimed at improving real-time forecasts obtained from VAR models by incorporating information from survey forecasts. The first method uses the information in survey nowcasts: It augments the original regression model by adding to the vector of dependent variables their survey nowcasts and it imposes that survey nowcasts have the same explanatory variables and similar coefficient values as the corresponding actuals. The idea is that survey forecasts provide information about the conditional mean of the data-generating process that is not incorporated in realizations. The second method uses long-term survey forecasts to construct an informative prior for the unconditional mean of the variables involved in the model. The idea is that experts often realize well before it is obvious in realizations data that the long-term path of an economic variable has changed. Examples include potential growth and inflation. The forecast success of the two methods is evaluated on U.S. macroeconomic data, using survey forecasts from the Philadelphia Fed's Survey of Professional Forecasters as additional information and find that they can improve on forecasts obtained from VARs estimated without the survey information.

C1063: **On combination forecasts and historical average: Economic and statistical evidence**
Presenter: **Apostolos Thomadakis**, University of Surrey, United Kingdom
The out-of-sample predictability of monthly German stock returns is examined, and the issue of whether combinations of individual model forecasts are able to provide significant out-of-sample gains relative to the historical average is addressed. Empirical analysis over the period from 1973 to 2012 implies that firstly term spread has the in-sample ability to predict stock returns, secondly, and most importantly, this variable successfully delivers consistent out-of-sample forecast gains relative to the historical average, and thirdly, combination forecasts do not appear to offer a significant evidence of consistently beating the historical average forecasts of the stock returns. Results are robust using both statistical and economic criteria, and hold across different out-of-sample forecast evaluation periods.

C1093: **Inflation expectations spillovers between the European monetary union and the United States**
Presenter: **Lars Winkelmann**, Freie Universitaet Berlin, Germany
Co-authors: Aleksei Netsunajev
Time varying spillovers of inflation expectations between the European Monetary Union (EMU) and the United States (US) are quantified at a medium and long term expectation horizon. We employ a recently proposed Markov switching VAR model to identify inflation expectations-shocks from forward break-even inflation rates. Zero restrictions reflecting instantaneous spillovers as well as the anchoring of inflation expectations are formally tested. We establish that impulse responses to the inflation-expectations-shocks do not vary significantly between the detected regimes. Changes in spillovers are triggered by switching inflation-expectations-shock volatilities. Results document increasing inter linkages in expectation formation processes and a substantial impact of the European sovereign debt crisis on US expectations.

C1022: **Precautionary savings over the life cycle: A simple two step weighted locally constant least squares estimator**
Presenter: **Alexandra Soberon**, Universidad de Cantabria, Spain
Co-authors: Juan Manuel Rodriguez Poo
A new method is proposed for estimating a structural model of optimal life-cycle savings, controlling both by uncertainty on health care expenditures and households risk aversion. The main interest of the estimator with respect to others in the literature is that it allows for heterogeneity of unknown form, varying parameters of unknown form in the Euler equation and endogeneity. The estimator of the function of interest turns out to have the simple form of a two step weighted locally constant least squares estimator. Additionally, some marginal integration techniques are required to compute a subset of the functionals of interest. We establish the asymptotic properties of these estimators. To illustrate the feasibility and

possible gain of using this method, we conclude with an application of households precautionary savings over the life cycle. From this empirical application we obtain the following conclusions: Households accumulate wealth at least in two periods in life, when they are younger, preventing against uncertainty in future income and when they are older, savings for retirement and bequests. Furthermore, national health programs have a negative impact over precautionary savings. This effect is stronger for younger households. Finally, comparing by educational levels, we obtain that households with low education levels are more risk averse than those ones with a higher level.

C644: **Monetary credibility, endogenous pass through and imported inflation uncertainty**
Presenter: **Jan Roestel**, Christian Albrechts Universitaet zu Kiel, Germany
Co-authors: Helmut Herwartz

The aim is to investigate if the uncertainty of import prices is mainly determined in the local currency or rather according to foreign conditions, and measure the potential implications for the import of inflation uncertainty. First and second order moment interactions among import prices, local currency prices and foreign producer currency prices are estimated for 20 mostly developed economies and a time period of 3 decades. From trivariate VAR MGARCH models with in-mean interaction effects we found evidence for first and second stage pass through of foreign price pressures to local prices both in terms of inflation levels and inflation uncertainties. The exposure to foreign influences appears especially relevant under lack of local monetary credibility: Evidence suggests that local inflation uncertainty relaxes the rigidity of import prices in the local currency, which, in turn, facilitates the import of foreign currency prices.

CS17 Room B34 TREND FILTERING AND STATISTICAL SIGNAL PROCESSING Chair: Serge Darolles

C290: **A quadratic Kalman filter**
Presenter: **Guillaume Roussellet**, Banque de France - CREST, France
Co-authors: Alain Monfort, Jean-Paul Renne

A new filtering and smoothing technique for non-linear state-space models is proposed. In our framework, observed variables are quadratic functions of latent factors following a Gaussian VAR. We augment the state vector, stacking it with its vectorized outer-product. The Compound auto-regressive (Car) dynamics property of this augmented state vector is exploited to construct the new filter. Since no adaptation of the linear Kalman algorithm is needed, our method does not involve computational-intensive calculations. A simulation study shows that in terms of filtering, our method performs at least as well as the extended and unscented Kalman approaches. Furthermore, the blindness of those competitors in a fully quadratic measurement equation case is resolved.

C320: **Practical applications of digital signal processing in trading**
Presenter: **Rafael Molinero**, Molinero Capital Management, United Kingdom

Strengths and weaknesses of digital signal processing algorithms, as well as the complexity inherent to market movements, will be discussed from a practical approach. We will try to illustrate how hedge funds may use digital signal processing to identify trading opportunities. We will also discuss the high level of noise which characterizes markets rendering much more difficult the utilization of these algorithms.

C391: **Model for price and trades high-frequency dynamics**
Presenter: **Emmanuel Bacry**, Ecole Polytechnique CNRS, France
Co-authors: Jean-Francois Muzy

The aim is to introduce a multivariate Hawkes process that accounts for the dynamics of market prices through the impact of market order arrivals at microstructural level. The model is a point process mainly characterized by 4 kernels. It allows one to account for both stylized facts of market prices microstructure (including random time arrival of price moves, discrete price grid, high frequency mean reversion, correlation functions behavior at various time scales) and the stylized facts of market impact (mainly the concave-square-root-like/relaxation characteristic shape of the market impact of a meta-order). Moreover, it allows one to estimate the entire market impact profile from anonymous market data. It is shown that these kernels can be empirically estimated from the empirical conditional mean intensities.

C723: **Multitask statistical learning for financial return prediction**
Presenter: **Ruocong Zhang**, Telecom ParisTech, France
Co-authors: Stephan Clemencon

Financial asset prediction is one example of an application in which the tasks are dependent with an unknown underlying structure. This structure may change over time, so that clustering large historical data would generally be unsuitable for prediction. For example, some stock prices would behave alike under one set of market conditions but very differently under another. Learning in a multitask framework would enable more consistent predictions, but also a better risk assessment on the predicted assets. We propose multitask Laplacian learning, a new method for jointly learning clusters of closely related tasks. Unlike standard multitask methodologies, the graph of relations among the tasks is not assumed to be known a priori, but is learned by the multitask Laplacian algorithm. The algorithm builds on kernel based methods and exploits an optimization approach for learning a continuously parameterized kernel. Multitask Laplacian learning can find application in many cases in which tasks are related with each other at varying degrees, some strongly, others weakly. Our experiments highlight such cases in which multitask Laplacian learning outperforms independent learning of tasks and a state of the art multitask learning method. In addition, they demonstrate that our algorithm partitions the tasks into clusters each of which contains well correlated tasks. Beyond theoretical results and computational advantages, the performance of the approach promoted in the context of financial return prediction is supported by strong empirical evidence.

CS23 Room B36 HIGH-FREQUENCY VOLATILITY FORECASTING Chair: Ana-Maria Fuertes

C1249: **Forecasting mean reversion in exchange rates: An empirical evaluation of support vector regression and Bollinger bands**
Presenter: **Farhad Bahramy**, Lancaster University, United Kingdom
Co-authors: Sven F Crone

Support Vector Regression (SVR) algorithms have received increasing interest in forecasting, promising nonlinear, non-parametric and data driven regression capabilities for time series prediction. But despite evidence on the nonlinear properties of foreign exchange markets, applications of SVR in price or return forecasting have demonstrated only mixed results. However, prior studies were limited to using only autoregressive time series inputs to SVR. The aim is toevaluate the efficacy of SVR to predict the Euro-US Dollar exchange rate using input vectors enhanced with explanatory variables on mean-reversion movements derived from Bollinger Bands technical indicators. Using a rigorous empirical out-of-sample evaluation of multiple rolling forecast origins, we assess the accuracy of different SVR input vectors, including upper and lower BB, binary trading signals of BB, and combinations of the above. As a result, a local SVR model using autoregressive lags in conjunction with BB bands and BB indicators, and recalibrated yearly, outperforms the random walk on directional and all other error metrics, showing some promise for an SVR application.

C115: **Overnight news and daily equity trading risk limits**
Presenter: **Katja Ahoniemi**, Imperial College Business School, United Kingdom
Co-authors: Ana-Maria Fuertes, Jose Olmo
Value-at-Risk (VaR) is widely used by banks to set day-ahead trading limits of different trading desks. A novel bivariate modeling approach for VaR is proposed that deals separately with the daytime and overnight segments of the trading day, and introduces a covariance term. This modeling strategy is motivated by inefficiencies in the opening price discovery process due to price staleness and news spillover effects, which induces a non-zero ex ante covariance between overnight and daytime returns. The model proposed is compared with various approaches. These include a simpler model without covariance, models with overnight returns redefined by moving the open price k minutes further into the trading day, and univariate models fitted to overnight-adjusted realized volatility measures. Using high-frequency data on the S&P 500 and Russell 2000 equity portfolios, we provide conclusive evidence on the superiority of the covariance modeling approach.

C675: **On daily stock volatility predictability: The role of market conditions, sector type and horizon**
Presenter: **Marwan Izzeldin**, Lancaster University, United Kingdom
Co-authors: Vasileios Pappas, Ana-Maria Fuertes
A hundred stocks from 10 sectors over the period 2000 to 2010 are considered to examine the impact of market conditions, sector type and horizon on the forecasting performance of various volatility models. Out-of-sample forecasts are generated with standard daily GARCH and with ARFIMA/HAR models fitted to intraday realised-based realised measures. We find that forecast performance is sensitive to market conditions/ horizon and sector type. ARFIMA generates the best daily volatility forecasts across all sectors for very short horizons up to 4 days ahead whereas HAR leads for longer horizons from 5 to 22 days. Engineering and Information Technology stocks are the easiest to forecast whereas Healthcare are the worst.

C028: **On a model-free approach to forecasting realized volatility of the S&P 100 stock index**
Presenter: **Ana-Maria Fuertes**, City University London, Spain
Co-authors: Julian Andrada-Felix, Fernando Fernandez-Rodriguez
The benefits of the model-free Nearest Neighbour (NN) method as a flexible approach to forecasting daily variability of the S&P 100 stock index are investigated. The benchmark forecasts are those obtained from time-series models able to capture parametrically long memory and specific nonlinear dependencies. For the out-of-sample forecast evaluation, we consider both the relatively calm market episode from 2002 to 2007 and the high-volatility period from 2008 to 2012. The Sharpe ratio of an options straddle-based trading strategy is employed as our main objective economic measure of predictability. For a range of transaction costs, signal-filtering rules and forecast horizons the NN predictions entail superior profits than model-based predictions. The best trading performance arises from the directional combination of both approaches through a unanimity rule. The results also reveal that the ranking obtained from common statistical forecast accuracy criteria does not have a direct mapping onto profitability.

CS98 Room B29 FINANCIAL ECONOMETRICS III Chair: Matthias Fengler

C1149: **Volatility prediction using a high-frequency-based component model**
Presenter: **Diaa Noureldin**, The American University in Cairo, Egypt
A high-frequency-based component model suitable for short- and medium-term volatility prediction is proposed. In a joint model for the daily return and a realized measure (e.g. realized variance), the conditional variance of the daily return has a multiplicative component structure: a long-run (secular) component, and a short-run (transitory) component driven by a high-frequency-based signal. Despite being a fixed-parameter model, its component structure allows for a time-varying intercept as well as time-varying dynamic parameters, that evolve with the long-run components of both the return and the realized measure. We discuss the model properties and estimation by quasi-maximum likelihood. The empirical analysis reveals strong out-of-sample performance at both short- and medium-term forecast horizons compared to some benchmark models.

C884: **Integer-valued trawl processes with applications to financial econometrics**
Presenter: **Almut Veraart**, Imperial College London, United Kingdom
Co-authors: Ole Barndorff-Nielsen, Asger Lunde, Neil Shephard
A new continuous-time framework for modelling serially correlated count and integer-valued data is introduced. The key component in our new model is the class of integer-valued trawl processes, which are serially correlated, stationary, infinitely divisible processes. We analyse the probabilistic properties of such processes in detail and, in addition, study volatility modulation and multivariate extensions within the new modelling framework. Moreover, we illustrate in a simulation study how our new models can be estimated. In an empirical study, we apply our new modelling framework to high frequency financial data and show that the class of integer-valued trawl processes can describe key features of high frequency financial data well.

C1019: **Asymmetric volatility spillovers**
Presenter: **Lukas Vacha**, UTIA AV CR, Czech Republic
Co-authors: Jozef Barunik, Evzen Kocenda
Based on the negative and positive realized semivariances, we modify a well-known volatility spillover index. The resulting asymmetric volatility spillover indices are easy to compute and account well for negative and positive parts of volatility. We apply the modified indices on the 30 U.S. stocks with the highest market capitalization over the period 2004-2011 to study intra-market spillovers. We provide evidence of sizable volatility spillover asymmetries and a markedly different pattern of spillovers during periods of economic ups and downs.

C1172: **Dispersed information in FX trading: A martingale representation**
Presenter: **Victoria Halstensen**, European University Institute, Italy
Informational heterogeneity is an important feature of real world foreign exchange markets. The abstraction from this feature is therefore likely a major reason for the poor empirical performance of many exchange rate models. Explicitly modeling this heterogeneity reveals a close correspondence between standard macroeconomic exchange rate models and the jump-diffusion type martingale representation of asset prices. This explains the better empirical performance of statistical jump diffusion models. Introducing dispersed information and news releases on macroeconomic conditions into a monetary model of exchange rates results in price dynamics that appear similar to the classical jump diffusion commonly assumed for many asset prices. The presence of dispersed information in conjunction with news releases on macroeconomic conditions leads to time-coordinated expectation revisions that produce jumps in the exchange rate path. The aim is to bridge the gap between the two modeling approaches. Employing newly developed statistical methods for disentangling jumps from integrated variance of jump diffusions using bi-power variation at an hourly frequency, I illustrate that jump activity is closely linked to news releases. This supports the notion that exchange rate jump activity arises as a consequence of expectation revisions from the arrival of new price relevant information.

C942: **Additive modeling of realized variance with tests for parametric specifications**
Presenter: **Matthias Fengler**, University of Sankt Gallen, Switzerland
Co-authors: Enno Mammen, Michael Vogt
The heterogeneous autoregressive (HAR) model is widely applied for modeling and forecasting realized variance (RV) data. As a core assumption, the daily, weekly and monthly variance component functions modeled linearly in the regressors. We extent the (HAR) model to an additive model and develop the diagnostic tools to investigate such a linearity assumption. In carrying out the tests in 17 series of RV data from 2003-2010, we find the linear specification often rejected, mostly in the daily component function. In visual terms, however, actual deviations from linearity are moderate; typically, we find mildly convex shapes for the daily component functions and concave ones for weekly and monthly functions. These findings may explain the limited superiority of extensions of the baseline HAR model.

CS36 Room B33 MODELING AND FORECASTING RISK AND RETURNS Chair: Maral Kichian

C399: **Downside and upside variance risk-premium**
Presenter: **Mohammad Jahan-Parvar**, Federal Reserve Baord of Governors, United States
Co-authors: Bruno Feunou, Cedric Okou
A new decomposition of the variance risk-premium is proposed in term of up and downside variance risk-premium. As is shown in the literature, the difference between upside and downside semi-variance is a measure of skewness. Building on this observation, an equilibrium model is constructed that admits closed form solutions for the premia that agents demand for bearing (semi)variance and skewness risks. The decomposition of equity premium establishes that 1) downside variance premium is the main component of the variance premium and 2) skewness is a priced factor and has significant prediction power for equity premium. This model is an easier to implement alternative to extreme tail jump/disaster premium. The empirical investigation highlights the positive and significant link between downside variance risk-premium and the equity premium, and a significant and negative relation between upside variance risk-premium and the equity premium.

C1117: **On the long-term dynamics of oil prices: Learning from combining forecasts**
Presenter: **Maral Kichian**, CREA Universite Laval, Canada
Co-authors: Jean-Thomas Bernard, Lynda Khalaf, Clement Yelou
Oil price dynamics have long been at the center of various active debates. A contentious -although popular- view is that the long run evolution of the real price of oil should not be approached via structural models since statistical models produce more reliable forecasts. We revisit the evidence supporting this view by using the U.S. Department of Energy's Energy Information Administration (EIA) forecasts. The agency adopted a formal model since 1965, the National Energy Modeling System (NEMS), that provides a natural reference benchmark. While learning which model performs best is interesting, the ultimate objective is to produce accurate forecasts. We express our research question along these lines, which to the best of our knowledge has not been considered with regards to EIA. Three important results emerge from our analysis. First, the EIA forecast differs markedly whether real-time or completely out-of-sample exercises are conducted, or whether the agency's own published assessment criterion is applied relative to the usual scholarly benchmark, i.e., the mean-square error (MSE) criterion. Second, we fail to find uniformly conclusive evidence in favor of statistical models. This result is noteworthy given the above cited popular view. For example, when the EIA's criterion is applied on continuously updated forecasts, the EIA outperforms all considered statistical models in the short -up to 3 years ahead- and long run -11 years and more ahead-, yet is dominated in the medium run -from 4 to 10 years-. These results are not contradicted with our purely out-of sample analysis, even with the MSE criterion. Third, the quadratic model dominates - although not uniformly - in interestingly enough scenarios. Forecast combinations, at least with the considered weights, dominate in rare although noteworthy cases. Interestingly, these rare occurrences concur with the above-cited medium run horizon in which the EIA forecasts under-perform.

C1269: **Endogeneity in parametric duration models with applications to clinical risk indices**
Presenter: **Anand Acharya**, Carleton University, Canada
Co-authors: Lynda Khalaf, Marcel Voia, David Wensley
An exact, identification-robust method is provided to correct for endogeneity in the location-scale family of parametric duration models using instrumental variables and the generalized Anderson Rubin statistic. An empirical application specifying the log-normal and Weibull duration model is carried out with a large ($n = 10044$) clinical patient data set. Using the trauma status of a critically ill patient as an instrument, we correct for the endogenous risk of mortality index in a commonly specified length of stay regression. Our analysis suggest that increased risk of mortality results in an increased length of stay in intensive care. In particular, the methods developed here correct for the upward bias of conventional estimates. Using the instrument and methods suggested in this paper, routinely collected patient risk indices are meaningful for informing policy in the health care setting.

CS112 Room G16 TIME SERIES ECONOMETRICS II Chair: Liudas Giraitis

C122: **On the great recession and the great moderation**
Presenter: **Lola Gadea**, University of Zaragoza, Spain
Co-authors: Ana Gomez-Loscos, Gabriel Perez-Quiros
The collapse of the global economy in 2008 following the outbreak of the financial crisis, and the ensuing economic developments of the so-called Great Recession (GR), has led many economists to question whether the relative oasis in which we had lived for almost a quarter of a century, known as the Great Moderation (GM), had indeed come to an end. The aim is to offer evidence that the cyclical features associated with the GM (decrease in output volatility and disappearance of the third phase of the cycle) still remain in force in spite of the GR and of the fact that it were even to last as long as the GM. These findings have important implications for the exit pattern of the current and future recessions.

C254: **Spatial price transmission on agricultural commodity markets under different volatility regimes**
Presenter: **Sebastien Ganneval**, Paris One Pantheon Sorbonne, France
Since 2007, agricultural commodity markets have been affected by a rise in volatility. In this context, we propose to analyze the impact of volatility on spatial price transmission between physical markets for three grain commodities (barley, corn and wheat). Our methodology is based on cointegration theory. Threshold vector error-correction models (TVECM) with two regimes are estimated. Unlike other studies on spatial price transmission we consider volatility as the exogenous transition variable. Thus, with ARCH processes, we modelize the volatility on the reference market for one of the relevant commodities. This specification allows us measuring the speed of adjustment of prices to their long term equilibrium value under low and high volatility regimes. Our empirical study is realized with weekly prices for several French physical markets over 2005-2011.

C1032: **Dynamic estimation of trade costs from real exchange rates**
Presenter: **Efthymios Pavlidis**, Lancaster University Management School, United Kingdom
Co-authors: Nicos Pavlidis
The difficulty of measuring trade costs is well documented. We propose a nonlinear state-space model that enables us to extract information about

changes in trade costs directly from real exchange rates. The model is well grounded in theory and nests numerous widely used empirical real exchange rate models. By employing two centuries of data on the Dollar-Sterling rate, we provide evidence of substantial variation in trade costs over time. Our results indicate that the process of economic integration that started after World War II has stopped during the last decades. We also contrast our results with a measure from the gravity literature and explore implications for the Purchasing Power Parity puzzle.

C1075: **A seemingly unrelated cointegrating polynomial regression analysis of the EKC**
Presenter: **Peter Grabarczyk**, Technical University Dortmund, Germany
Co-authors: Martin Wagner
Unit root and cointegration techniques are widely used for the analysis of Environmental Kuznets curves, where it is by now understood that the EKC, if present, is a nonlinear cointegrating relationship that has to be estimated with methods designed for this type of nonlinear relationships. We investigate the properties of a recently developed fully modified OLS estimator for systems of equations including deterministic variables, integrated processes and integer powers of integrated processes as explanatory variables and stationary errors. The regressors are allowed to be endogenous and the errors serially correlated as well as cross-sectionally correlated. Such systems are referred to as seemingly unrelated cointegrating polynomial regressions. We study the properties of various SUR and pooled estimators, as well as of coefficient and poolability tests based thereon. The empirical application studies the EKC hypothesis for CO2 and SO2 emissions data for 19 early industrialized countries.

C1256: **Modeling the impact of forecast-based regime switches on time series**
Presenter: **Koen Bel**, Erasmus University Rotterdam, Netherlands
Co-authors: Richard Paap
Forecasts of key macroeconomic variables may lead to policy changes of governments, central banks and other economic agents. Policy changes in turn lead to structural changes in macroeconomic time series models. To describe this phenomenon we introduce a logistic smooth transition autoregressive model where the regime switches depend on the forecast of the time series of interest. This forecast can either be an exogenous expert forecast or an endogenous forecast generated by the model. Results of an application of the model to US inflation shows that (i) forecasts lead to regime changes and have an impact on the level of inflation; (ii) a relatively large forecast results in actions which in the end lower the inflation rate; (iii) a counterfactual scenario where forecasts during the oil crises in the 1970s are assumed to be correct leads to lower inflation than observed.

CS72 Room B30 COMPUTATIONAL METHODS FOR OPTION PRICING Chair: Lars Stentoft

C959: **A kernel approach to energy contracts valuations in high dimensions**
Presenter: **Denis Mazieres**, Birkbeck University of London, United Kingdom
Co-authors: Helyette Geman, Simon Hubbert
A kernel-based approach to value energy contracts in high dimensions is introduced. This method utilises both Tensor of Radial Basis Functions (TBF) and Radial Basis Functions (RBF). Applications of these methods are provided for multi-exercise options in energy markets. These methods are then studied from a performance perspective.

C628: **Convexity adjustments for affine term structure models**
Presenter: **Agatha Murgoci**, CBS, Denmark
Co-authors: Raquel Medeiros Gaspar
Practitioners are used to value a broad class of exotic interest rate derivatives simply by performing what is known as convexity adjustments (or convexity corrections). Convexity adjustments, however, are highly dependent on both the particular interest rate model one chooses and the exact payoff under analysis. The focus is on affine term structure (ATS) models and, in this context, conjecture convexity adjustments should be related affine functionals. An unified framework is then developed to compute various convexity adjustments. These adjustments can be computed without the need of Taylor approximations for non-Gaussian ATS models. Unlike previous adjustments, they are exact solutions up to an ODE. We derive the system of Riccatti ODE-s one needs to compute to obtain the exact adjustment for several examples.

C850: **Business conditions, market volatility and option prices**
Presenter: **Christian Dorion**, HEC Montreal, Canada
Options are bets on volatility. Volatility is highly counter-cyclical, but the volatility models used to value options typically disregard macroeconomic risk when predicting future volatility. The asset pricing properties of these models suffer from this omission; their option pricing errors tend to increase as business conditions deteriorate. We propose the use of a model in which the conditional expected level of volatility evolves with business conditions and, as a result, that accounts for macroeconomic risk. This risk is quantified using a mixed data sampling (MIDAS) structure to account for changes in the recently introduced Aruoba-Diebold-Scotti (ADS) Business Conditions Index. The new model outperforms existing ones in explaining asset returns and in the pricing of options, especially when business conditions are at their worst.

C529: **On the pricing framework for option valuation under GARCH and non-normality**
Presenter: **Lars Stentoft**, Copenhagen Business School, Denmark
Two different pricing frameworks are typically used in the literature characterizing option prices under GARCH with non-normal innovations: the equilibrium pricing approach and the no-arbitrage approach. Each framework can accommodate various forms of GARCH and innovation distributions, but empirical implementation and tests are typically done in one framework or the other because of the difficulty to find a model suitable for both. The aim is to contribute to this literature by comparing and documenting the empirical performance of a GARCH specification with skewed and leptokurtic Johnson su innovations which can be implemented in both pricing frameworks. The two frameworks are tested empirically with constant as well as time varying risk price of risk.

CS116 Room B20 FINANCIAL APPLICATIONS IV Chair: Spyridon Vrontos

C559: **The real-life performance of market timing with moving average and time-series momentum rules**
Presenter: **Valeriy Zakamulin**, University of Agder, Norway
The aim is to revisit the myths about the superior performance of the market timing strategies with moving average and time-series momentum rules. These active timing strategies are very appealing to investors because of their extraordinary simplicity and because they promise enormous advantages as compared to their passive counterparts. It is maintained that the reported performance of these market timing strategies contains a substantial data-mining bias and is further manipulated by cherry picking and neglecting important market frictions. In order to provide estimates for the real-life historical performance of the market timing strategies, as well as their expected future performance, using a bootstrap procedure out-of-sample tests of these two timing models are performed where it is accounted for realistic transaction costs.

C905: **Structural credit risk model under stochastic volatility: A particle-filter approach**
Presenter: **Di Bu**, The University of Queensland, Australia

Merton's structural credit risk model is extended to account for the fact that the underlying asset volatility follows a stochastic process. With the presence of stochastic volatility, the transformed-data maximum likelihood estimation (MLE) method can no longer be applied. We devise a particle filtering algorithm to estimate this model, and the technique developed here is based on the general non-linear and non-Gaussian filtering approach with sequential parameter learning. A simulation study is conducted to ascertain the performance of the estimation method. Meanwhile, we implement this model on the real data of Dow Jones Industrial average and find that ignoring stochastic volatility in the structural credit risk model can lead to significantly underestimating the firm's credit risk.

C1038: **Optimal cut-off points for multiple causes of business failure models**
Presenter: **Marialuisa Restaino**, University of Salerno, Italy
Co-authors: Alessandra Amendola

In studies involving bankruptcy prediction models, the attention is focused on the classification of firms into groups according to their financial status and the prediction of the status for new firms. The researchers have concentrated long time on the binary events (healthy vs bankrupt), without taking into account that the firms may exit the market for several different reasons (such as merger, acquisition, voluntary liquidation and so on). In order to predict the probability of being bankrupt at certain time points, different statistical models and methods can be used (discriminant analysis, logit and probit, survival analysis). In each model an optimal cutoff point is needed in order to classify/predict the firms' status. Some methods have been developed for two-group classification, and the researchers have pointed out their advantages and disadvantages. Until now, there are no references on how to determine optimal thresholds when the groups are more than two. Here, a method based on the optimization of misclassification error is proposed for determining optimal cutoff points when there are multiple causes of business failure. The proposed procedure has been tested by a simulation study and on a real data set.

C1059: **Modelling the credit card ownership in Europe: A count data approach**
Presenter: **Alfonso Pellecchia**, University of Salerno, Italy
Co-authors: Alessandra Amendola, Luca Sensini

Credit cards, both as mean of payment and borrowing, rise many economic issues. Firstly, the credit card industry can be viewed as a two-sided network industry characterized by externalities that could harm competition. Particularly, the fact that consumers hold or use credit cards from multiple networks is known as 'multi-homing' and in some theoretical models is of great importance in determining the outcome of the industry. Moreover, some studies show that the multiple credit cards can be seen as a device to access to more financing, making family bankruptcy more likely. Then, we model the number of credit cards held by European household using demographic, socio-economic and geographical variables as potential predictors. Unlike most of the existing literature, we employ standard and advanced count data models, which are more suitable to model integer nonnegative values. Our estimates show that such variables are statistically significant in explaining the credit card ownership, although substantial differences among European countries exist. Our results could be a first step toward a more in deep understanding of multi-homing and related phenomena (e.g. 'co-holding') characterizing the credit card industry.

C1148: **Optimal structure of central counterparties clearing houses for minimizing default risk**
Presenter: **Marius Frunza**, Schwarzthal Kapital, France

After the recent crisis Central Counterparties Clearing Houses (CCP) gained momentum with the new regulations(Dodd-Franck, EMIR) concerning the clearing of Over the Counter(OTC) derivatives. By interposing themselves in transactions, CCPs help to manage counterparty risk for market members and facilitate the netting positions. Nevertheless CCPs are themselves exposed to various risk, the most important being the risk due to the propagation of losses from the default of one of its Clearing members. A classic CCP has a multi-'tranche' structure that allows to absorb the losses due to members defaults. The waterfall is structure in the following tranches: Initial Margin, Default Fund of the defaulter, CCP Equity, Default Funds of the remaining Clearing members and other haircuts. In the case of a Clearing Member's default the loss will be amortized by each tranche, depending on the magnitude of the exposure at that time. If the loss cannot be amortized by the waterfall, the CCP enters in default. Nevertheless the Clearing Members start to observe losses before the technical default of the CCP. The aim is to study through a simulating approach the loss propagation across the waterfall under various assumptions. Based on these observations we propose an optimal structure of the waterfall in order to minimize the expected economic losses for the surviving Clearing Members in the cases of a CCP member's. The resilience of the proposed waterfall structure is studied through stress tests including scenarios of distressed markets.

CS118 Room B18 CONTRIBUTIONS TO VOLATILITY MODELS AND THEIR APPLICATIONS Chair: Francesco Audrino

C1094: **Efficient stochastic modelling of financial volatility within non-Gaussian paradigm**
Presenter: **Farrukh Javed**, Lund University, Sweden
Co-authors: Krzysztof Podgorski

Two features of volatility of the stock returns are addressed: the asymmetric response of volatility to positive and negative 'shocks', and positive serial correlation over long time lags. In the past, the asymmetric power ARCH model (A-PARCH) belonged to the most popular models attempting to capture these asymmetry and long memory effects. We discuss a new model based on the generalized Laplace distributions that equally efficiently models these features however additionally has the ability to retrieve non-Gaussian feature of the residuals, which A-PARCH has difficulty to account for. It is more flexible in its formulation and has one additional parameter, the shape parameter. The heavy tails/long memory are modelled by a shape parameter rather than power (making a possible additional parameter allowing for more flexibility if needed). We compare the two models using S&P 500 data as well as present a Monte Carlo study based on the data simulated from the models that demonstrate that the new model is an attractive alternative to A-PARCH and related models.

C1155: **Modelling conditional volatility with a continuum of hidden regimes and fundamental factors**
Presenter: **Rachid Belhachemi**, American University in Cairo, Egypt
Co-authors: Mohammed Bouaddi

A key issue to model conditional densities of returns in all financial markets is the time-variation in conditional volatility. The focus is usually on the classic econometric approach to volatility modelling by the mean of the generalized autoregressive conditional heteroskedasticity (GARCH) models where the conditional mean and the conditional volatility depend only on historical prices. We propose a new family of distributions in which the conditional distribution depends on a latent continuous factor with a continuum of states. The most important contribution is the inclusion of a latent continuous factor with a continuum of states unlike the Markov chains where the state variable is discrete with a few number of states. The distribution contains eight parameters. Five of these parameters capture the effect of the physical distribution of the sate variable on the conditional distribution of returns. The parameters have economic interpretations in terms of volatility and correlations of the returns with the hidden continuous state. The distribution has an interpretation in terms of a mixture distribution with time varying mixing probabilities. We give the whole characterization of this distribution, e.g., stochastic representation, moment generating function, higher order moment. The empirical results show that this distribution outperforms all its competitors in terms of capturing the stylized facts known for stock returns, namely, volatility clustering, skewness, excess kurtosis and regime dependence.

C1177: **Volatility forecasting using latent information**
Presenter: **Chainarong Kesamoon**, Autonomous University of Barcelona, Spain
Co-authors: Joan del Castillo

Volatility has become a standard risk measure in financial market. Accurate forecasting of volatility is relevant both to risk calculation and derivatives valuation, but forecasting is difficult because financial time series often exhibit time varying volatility and volatility clusters. GARCH models and stochastic volatility (SV) models are two main classes of models used to capture the dynamics of volatility. The parameters estimation is generally more complicated in SV models since the volatility is a latent random variable. The normal inverse Gaussian (NIG) model, that is a parametric Levy process closed under convolution, is estimated by considering it as a generalised linear model (GLM) with random effects. This allows us to estimate unobserved volatility and use this information to improve the forecasting. The GLM with random effects can be extended to incorporate exogenous information such as trading volume, open, close, high and low prices into the models and this estimation method is still applicable. This method is applied to three currency exchange rates: GBP/USD, JPY/USD and EUR/USD from 2006 to 2010 including the financial crisis of 2008. We find that the estimated volatilities are highly correlated to standard models and the forecasts are improved by time series method.

C983: **Stationarity and ergodicity regions for score driven dynamic correlation models**
Presenter: **Erkki Silde**, VU Amsterdam, Netherlands
Co-authors: Francisco Blasques, Andre Lucas

Stationarity and ergodicity (SE) regions are described for a recently proposed class of score driven dynamic correlation models. These models have important applications in empirical work. The regions are derived from average contraction sufficiency conditions and take a non-standard form. We show that the non-standard shape of the sufficiency regions cannot be avoided by reparameterizing the model or by rescaling the score steps in the transition equation for the correlation parameter. This makes the result markedly different from the volatility case. Observationally equivalent decompositions of the stochastic recurrence equation yield regions with different sizes and shapes. We illustrate our results with an analysis of time-varying correlations between UK and Greek equity indices. We find that also in empirical applications different decompositions can give rise to different conclusions regarding the stability of the estimated model.

C1026: **A semi-parametric locally stationary ARCH Model**
Presenter: **Lionel Truquet**, ENSAI, France
Co-authors: Valentin Patilea

An ARCH(p) model with a time-varying constant and fixed, but unknown, coefficients of the lag variables is considered. This model is a compromise between the stationary ARCH model of Engle and the time-varying ARCH models. The time-varying constant allows us to capture volatility non stationarity. Moreover, assuming constant coefficients for the lag variables could be appealing for ARCH models users. The time-varying constant is estimated using a kernel estimator, while the lag variables coefficients are estimated using weighted marginal (unconditional) moment conditions. Our procedure could be interpreted as an extension to nonlinear time-series framework of the classical least-squares aproach for semi-parametric partially linear models. We derive the parametric rate of convergence and the asymptotic normality for the lag variables coefficients estimates when the weights belong to a class of functions depending on the past. Moreover, for Gaussian inputs, we construct an asymptotically efficient estimator in the semi-parametric framework. We also prove the uniform convergence of the kernel estimator of the time-varying constant. The results are derived under minimal moment conditions on the non stationary process, in particular without additional restrictions on the lag variable coefficients with respect to the classical ARCH(p) modeling. The problem of testing whether the lag variables coefficients are constant or not is also investigated. Simulation and real data applications illustrate the new modeling approach.

CS95 Room B35 RECENT DEVELOPMENTS IN SEASONAL ADJUSTMENT II Chair: Rosa Ruggeri-Cannata

C523: **On trend-cycle-seasonal interactions**
Presenter: **Denise Osborn**, University of Manchester, United Kingdom
Co-authors: Irma Hindrayanto, Jan P.A.M. Jacobs

Traditional unobserved component models assume that the trend, cycle and seasonal components of an individual time series evolve separately over time. Although this assumption has been relaxed in recent papers that focus on trend/cycle interactions, it remains at the core of all seasonal adjustment methods applied by official statistical agencies around the world. An unobserved components model is developed that permits nonzero correlations between seasonal and nonseasonal shocks, hence allowing testing of the uncorrelated assumption that is traditionally imposed. Identification conditions for estimation of the parameters are discussed from the perspectives of both analytical and simulation results, while applications to observed time series illustrate the model and its implications for seasonal adjustment.

C570: **Theoretical and empirical analysis of the reg-components approach for seasonal adjustment**
Presenter: **Dominique Ladiray**, INSEE, France

The aim is to review the Reg-Component models and the REGCMPNT program made available in 2011. We focus on the application of these models to Seasonal Adjustment. Special attention is paid to the compliance of the program to the ESS guidelines on seasonal adjustment and to the detection and correction of outliers in this specific framework.

C583: **Nonparametric price transmission analysis**
Presenter: **Tatyana Krivobokova**, Georg-August-Universitaet Goettingen, Germany
Co-authors: Rosales Francisco

The (threshold) vector correction model is a simple parametric model that relates two cointegrated prices time series. Depending on the goal, certain parametric models for the price differences are estimated. In particular, it is of interest how well two markets are integrated, which is described by a certain parameter of this model. Clearly, if the underlying time series fail to fulfil the stated assumptions, parametric models could completely fail. Another drawback of such parametric models is that they are hardly extendable to the multivariate case, where the simultaneous integration of several time series is of interest. Therefore, the aim is to develop a different, nonparametric and much more flexible model for the price transmission analysis, which can include complex seasonal patterns and be extended to the multivariate case. The first step is to obtain a nonparametric model for a collection of time series. As an example, pork prices in 15 countries of the European Union are considered, observed weekly over a period of the last 17 years. The idea is to model all series simultaneously, so that each individual time series can be represented as a sum of an overall smooth trend, which is common for all series, a smooth deviation from this overall trend, that includes a subject-specific seasonal pattern and a stationary process in the residuals. Finally, a new concept of the market integration description should be developed.

C619: **Seasonal adjustment in times of crisis using a dedicated algorithm for automatic model identification procedure**

Presenter: **Sylwia Grudkowska**, National Bank of Poland, Poland

In times of dramatic change in a mean of a time series the quality of seasonal adjustment results is often decreasing. Such disturbances are frequently observed, e.g. during economic crises. The empirical studies imply that under those circumstances the application of the Ramp effect in the pre-processing part of seasonal adjustment improves the degree of the linearization of a time series. However, due to lack of an automatic identification procedure of this effect, it is rarely used in practice. For this reason an algorithm for detecting Ramps, which enhances the results from the Automatic Model Identification procedure of TRAMO/SEATS in this specific situation, has been developed.

Monday 16.12.2013 08:45 - 10:25 Parallel Session O – ERCIM

ES04 Room 104 STOCHASTIC APPROXIMATION FOR BIG DATA Chair: Christoforos Anagnostopoulos

E1250: **Abstract stochastic approximation for online statistical algorithms**
Presenter: **David Leslie**, University of Bristol, United Kingdom
Co-authors: Stephen Perkins
The purpose is to develop the ODE approach to stochastic approximation for use when the parameter of interest is an element of a function space instead of a Euclidean space. This allows us to prove the consistency of Newton-type online algorithms for estimating mixing densities, with fewer assumptions than previously needed.

E1259: **Parameter estimation for hidden Markov models with intractable likelihoods**
Presenter: **Sumeetpal Singh**, Cambridge University, United Kingdom
Approximate Bayesian Computation (ABC) is a popular technique for approximating likelihoods and is often used in parameter estimation when the likelihood functions are analytically intractable. Although the use of ABC is widespread in many fields, there has been little investigation of the theoretical properties of the resulting estimators. We give a theoretical analysis of the asymptotic properties of ABC based maximum likelihood parameter estimation for hidden Markov models. We also discuss how Sequential Monte Carlo methods provide a natural method for implementing likelihood based ABC procedures.

E1263: **Gradient free parameter estimation for hidden Markov models with intractable likelihoods**
Presenter: **Nikolas Kantas**, Imperial College, United Kingdom
Maximum Likelihood estimation (MLE) for the static model parameters of hidden Markov models (HMMs) is addressed. We will consider the case where one cannot or does not want to compute the conditional likelihood density of the observation given the hidden state because of increased computational complexity or analytical intractability. Instead we will assume that one may obtain samples from this conditional likelihood and hence use approximate Bayesian computation (ABC) approximations of the original HMM. We will then use gradient methods to perform MLE in an on-line and offline setting by implementing simultaneous perturbation stochastic approximation (SPSA) and Sequential Monte Carlo (SMC) for the ABC approximation of the HMM. ABC approximations will induce a bias and we will investigate its effect using numerical examples.

E1264: **Truncated stochastic approximation with moving bounds**
Presenter: **Teo Sharia**, Royal Holloway University of London, United Kingdom
A wide class of truncated stochastic approximation procedures with moving random bounds will be discussed. This class of procedures has three main characteristics: truncations with random moving bounds, a matrix-valued random step-size sequence, and a dynamically changing random regression function. While we believe that the proposed class of procedures will find its way to a wider range of applications, the main motivation is to accommodate applications to parametric statistical estimation theory. The proposed method allows for incorporation of auxiliary information into the estimation process, and is consistent and asymptotically efficient under certain regularity conditions.

ES28 Room Bloomsbury ADVANCES IN ROBUST DATA ANALYSIS Chair: Agustin Mayo

E586: **Detection of outliers in megavariate data**
Presenter: **Alicia Nieto-Reyes**, Universidad de Cantabria, Spain
Co-authors: Javier Cabrera
Real data sets commonly have a megavariate structure: small sample size and high dimensional space. The detection of outliers can be studied from the point of view of which directions contain outliers or from the perspective of the sample, i.e. seeking for the elements of the sample that are outliers as a whole. Due to the sample size being actually small, this last type of outliers are difficult to detect. For instance, microarray data have a megavariate structure and are known for containing outliers even after normalization. While seeking for outlier genes in a given microarray has been broadly studied, there only exist graphical techniques to find microarrays that are outliers with respect to the given microarray sample; the problem being, they are usually misleading. An analytical technique to detect outliers specifically designed for megavariate data is introduced. It is quite simple and so, pleasant for either statisticians and biologists, who have to deal with microarray data.

E731: **A robust approach in cluster weighted modeling**
Presenter: **Francesca Greselin**, The University of Milano Bicocca, Italy
Co-authors: Luis Angel Garcia Escudero, Alfonso Gordaliza, Salvatore Ingrassia, Agustin Mayo-Iscar
The aim is to introduce robust estimation in model based clustering, via the Cluster-Weighted approach (CWM). The latter has been shown to be an effective way to provide data modeling and classification for multivariate data. It was originally proposed in the context of media technology and recently extended to a quite general setting. Our aim, here, is to reconsider the EM algorithm which provides parameter estimation, by trimming a fraction α of the data and using restrictions on covariance eigenvalues. This approach enables the model to avoid fitting a small localized random pattern in the data rather than a proper underlying cluster structure. To select appropriate bounds for constrained estimation, a data-driven procedure is illustrated, and applied on real data.

E097: **Advances in the small sample distribution of the trimmed mean**
Presenter: **Alfonso Garcia-Perez**, UNED, Spain
The trimmed mean has an intractable exact distribution. Its large-sample approximation is asymptotically normal under some conditions. When the sample size is small and the data are normally distributed, a Student's $t-$distribution can be used as an approximation for the standardized trimmed mean. Nevertheless, if data are suppose to come from a normal distribution (i.e., no contamination is assumed in the data), the trimmed mean is not really needed. Although there are some Edgeworth expansions used as approximations, it is well known that these approximations are accurate only in the center of the distribution, but not in the tails where they can be even negative. The only accurate approximations for the distribution, when the sample size is small and the distribution not normal, are the saddlepoint approximations, although these approximations are almost impossible to apply and the elements involved in them, difficult to interpret. We propose a new approximation in this context based on the von Mises expansion. Finally, the trimming fraction is choice using the new concept of the $p-$value line. Impartial trimming is explored and some applications are given.

E769: **Covariance matrix estimation: Robustness and dimensionality**
Presenter: **Pietro Coretto**, Universita degli Studi di Salerno, Italy
The sample covariance matrix performs poorly in situations with outliers and high dimensionality. Most high-breakdown point estimates of multivariate location and scatter cannot cope with high-dimensional settings. We propose a ML-type estimator of location and scatter based on an improper density function that is positive everywhere. The estimator smoothly trims data points far away from the center. The robustness of the procedure is based on tunings that have a straightforward interpretation. The proposed estimator is easy to compute even in high-dimensional

settings. We compare its behavior with that of the Stahel-Donoho estimator, FMCD, S-estimates, and the Orthogonalized Gnanadesikan-Kettenring estimator. We explore datasets where the following features are combined: (i) moderate to small values of the ratio n/p, with n and p being the sample size and the dimension of the data space respectively; (ii) moderate to large proportions of contamination; (iii) deviations from elliptical shapes. Empirical evidence confirms that contaminated high-dimensional datasets present extraordinary challenges to data analysis.

ES36 Room Woburn BAYESIAN APPROACHES TO INTEGRATIVE GENOMICS Chair: Michele Guindani

E062: **An integrative Bayesian modeling approach to imaging**
Presenter: **Marina Vannucci**, Rice University, United States
Co-authors: Francesco Stingo, Michele Guindani, Vince Calhoun
A Bayesian hierarchical modeling approach for imaging genetics is presented, where the interest lies in linking brain connectivity across multiple individuals to their genetic information. We have available data from a functional magnetic resonance (fMRI) study on schizophrenia. Our goals are to identify brain regions of interest (ROIs) with discriminating activation patterns between schizophrenic patients and healthy controls, and to relate the ROIs' activations with available genetic information from single nucleotide polymorphisms (SNPs) on the subjects. For this task we develop a hierarchical mixture model that includes several innovative characteristics: it incorporates the selection of ROIs that discriminate the subjects into separate groups; it allows the mixture components to depend on selected covariates; it includes prior models that capture structural dependencies among the ROIs. Applied to the schizophrenia data set, the model leads to the simultaneous selection of a set of discriminatory ROIs and the relevant SNPs, together with the reconstruction of the correlation structure of the selected regions.

E283: **Bayesian approaches for high-dimensional biological networks**
Presenter: **Francesco Stingo**, UT MD Anderson, United States
Co-authors: Christine Peterson, Marina Vannucci, Yang Ni, Veera Baladandayuthapani
Recent advancements in Gaussian graphical models for the analysis of biological networks will be illustrated. First, in order to take into account the known biological nonlinear interactions, it will be shown how to construct gene regulatory networks that allow for both linear (Gaussian) and non-linear protein interactions. Then, a modeling approach for joint inference of multiple networks will be introduced. This approach allows us not only to share information between sample groups when appropriate, but also to obtain a measure of relative network similarity across groups. The application of proposed methodologies to the inference of protein networks will be illustrated, where the protein expressions were obtained using reverse phase protein arrays (RPPA).

E474: **Dissecting high-dimensional phenotypes with Bayesian sparse factor analysis of genetic covariance matrices**
Presenter: **Sayan Mukherjee**, Duke University, United States
Quantitative genetic studies that model complex, multivariate phenotypes are important for both evolutionary prediction and artificial selection. For example, changes in gene expression can provide insight into developmental and physiological mechanisms that link genotype and phenotype. However, classical analytical techniques are poorly suited to quantitative genetic studies of gene expression where the number of traits assayed per individual can reach many thousands. The aim is to derive a Bayesian genetic sparse factor model for estimating the genetic covariance matrix (G-matrix) of high-dimensional traits, such as gene expression, in a mixed effects model. The key idea of the model is that it needs only to consider G-matrices that are biologically plausible. An organism's entire phenotype is the result of processes that are modular and have limited complexity. This implies that the G-matrix will be highly structured. In particular, it is assumed that a limited number of intermediate traits (or factors, e.g., variations in development or physiology) control the variation in the high-dimensional phenotype, and that each of these intermediate traits is sparse - affecting only a few observed traits. The advantages of this approach are two-fold. First, sparse factors are interpretable and provide biological insight into mechanisms underlying the genetic architecture. Second, enforcing sparsity helps prevent sampling errors from swamping out the true signal in high-dimensional data. The advantages of our model are demonstrated on simulated data and in an analysis of a published Drosophila melanogaster gene expression data set.

E716: **Computational decision theory to explore posterior models in cancer genomics**
Presenter: **Chris Holmes**, University of Oxford, United Kingdom
Bayesian models have proved highly useful in the analysis of genetic variation arising in cancers. We have previously developed Bayesian Hidden Markov Models for this task where the hidden states relate to structural changes in DNA known to be key drivers of cancer initialisation and progression. We will discuss the use of decision theory to help elicit knowledge contained in the posterior models. That is, having conditioned a model on data how can we explore the posterior model for interesting, highly probable, state sequences that associate with adverse clinical outcomes.

ES51 Room Montague SEMIPARAMETRIC ESTIMATION OF SIMULTANEOUS EQUATION MODELS Chair: Giampierro Marra

E141: **Bayesian latent Gaussian simultaneous equation models**
Presenter: **Thomas Kneib**, University of Goettingen, Germany
Bayesian seemingly unrelated regression models combine two or more regression equations based on correlated multivariate normal errors. In this framework, flexible predictor specifications can easily be transferred from univariate regression models and can be conveniently estimated based on efficient Markov chain Monte Carlo simulation techniques. We will discuss extensions where seemingly unrelated regression is used for the latent representation of simultaneous equation models. As examples, we will consider bivariate binary regression models, binary ordinal probit models and bivariate Bayesian quantile regression. For each of these model classes, we will utilize penalized splines to introduce flexibility with respect to the predictor specification and will also discuss potential extensions for spatial effects, random effects or interactions. The developed methodology will be illustrating two examples related to the decoding of neuronal coding hypotheses and species interactions in an analysis of biodiversity.

E372: **Bivariate response vector generalized additive models**
Presenter: **Thomas Yee**, University of Auckland, New Zealand
The VGAM framework has developed to handle a wide range of data-type responses. One such example are bivariate binary responses. The aim is to give an overview of at least 3 models for handling these: the bivariate odds-ratio model, bivariate probit model and a loglinear model for two binary responses. The framework naturally allows smoothing for all its parameters, and it is implemented in the VGAM R package which has a vglm() function that works in a similar manner to the standard glm() function. There is also a vgam() function that shares similarities with gam()-like functions from other R packages. Well over 100 models/distributions are currently implemented. Time allowing, other types of bivariate responses will be considered, and some new details about P-spline VGAMs will be described.

E533: **Generalized sample selection model**
Presenter: **Malgorzata Wojtys**, Warsaw University of Technology, Poland
Co-authors: Giampiero Marra
Non-random sample selection is a problem frequently encountered among observational studies, arising when a response variable of interest is

observed only for a restricted, non random sample. A generalized sample selection model is considered, defined by a system of two regression equations, the outcome equation and selection equation, expressed as generalized additive models. Each of the two equations contains a parametric linear part related to qualitative covariates and a non-parametric component in the form of a sum of smooth functions of continuous covariates. Selectivity arises if outcome variables of these two equations are mutually dependent. A fitting method is proposed, based on penalized likelihood simultaneous equation estimation where the smooth functions are approximated using the regression spline approach and the bivariate distribution of the outcomes of the two equations is represented in the form of an Archimedean copula. Theoretical results related to asymptotic consistency of the procedure are presented and small sample performance of the method is illustrated using simulated data as well as the real world example. Moreover, the problem of copula model selection in the context of sample selection models is addressed.

E299: **Copula regression spline models for binary outcomes**
Presenter: **Rosalba Radice**, Birkbeck University of London, United Kingdom
Co-authors: Giampiero Marra, Malgorzata Wojtys

A model for binary responses which can deal simultaneously with the issues of (i) unobserved confounding, (ii) residual confounding and (iii) non-linear dependence between the treatment and outcome variable is presented. Unobserved confounding arises when variables which are associated with both the treatment and outcome (hence confounders) are not available. Residual confounding occurs when observed confounders are insufficiently accounted for in the analysis. Treatment and outcome might exhibit a non-linear dependence that cannot be modelled using a linear measure of association (e.g. correlation coefficient). Problem (i) is addressed using a two-equation structural latent variable framework, where one equation describes a binary outcome as a function of a binary treatment whereas the other determines whether the treatment is received. Issue (ii) is tackled by modelling flexibly covariate-response relationships using smooth functions for the continuous confounders, whereas problem (iii) is dealt with by assuming normality for the marginal distributions of the two equations while introducing non-linear dependencies using copulas. Related model fitting and inference procedures are also developed, and asymptotic considerations presented. The method is illustrated using data from the 2008 Medical Expenditure Panel Survey where the aim is to estimate the effect of private health insurance on health care utilization.

ES101 Room Bedford NEW PERSPECTIVES ON FUNCTIONAL DATA ANALYSIS Chair: Elvira Romano

E521: **Uncertainty evaluation in functional kriging with external drift**
Presenter: **Rosaria Ignaccolo**, University of Turin, Italy
Co-authors: Maria Franco Villoria

Functional data featured by a spatial dependence structure occur in many environmental sciences when curves are observed, for example, along time or along depth. To predict a curve at an unmonitored site taking into account exogenous variables a functional kriging model with external drift can be applied. However, uncertainty evaluation of such a predictor is still an open issue. To evaluate uncertainty of the predicted curves, a semi-parametric bootstrap approach for spatially correlated data is adapted to functional data. Confidence bands are obtained by ordering the bootstrapped predicted curves in two different ways, one based on functional depth and the other by means of distance between curves. The proposed approach is illustrated on pollutant functional data gathered by the monitoring network of Piemonte region (Italy), with the drift defined in terms of meteorological functional covariates and orographical scalar variables. Confidence bands are added to predicted pollutant curves on 10 validation sites and the two ordering strategies compared.

E878: **Some extensions of FDA when dealing with particular curves**
Presenter: **David Nerini**, Mediterranean Institute of Oceanology, France
Co-authors: Claude Mante, Pascal Monestiez

Most of statistical multivariate methods have been discussed and extended to cope with functional data. Classification methods, PCA, linear model are some classical examples where intense efforts, both from a theoretical or applied point of view, have been made to solve the technical problems that arise when dealing with curves. However, few theoretical developments or applied statistical cases have been tackled when observations arrive as several curves or as parametrized curves. Starting from data in oceanography, we will consider some statistical problems on the use of PCA, linear model or kriging that may appear when observations arrive as several curves or when observations can be described with parametrized curves.

E1029: **Sparse principal component analysis for multilevel functional data**
Presenter: **Todd Ogden**, Columbia University, United States
Co-authors: Xiaochen Cai, Ernest Greene

The situation in which noisy functional observations are available for each of several sessions for each of several individuals is considered. We describe a method to decompose the variance of the data into three components: individual level variation, session level variation, and noise. This is accomplished using a wavelet basis in order to allow for localized features in the data and to ensure sparsity in the transformed domain. The methodology is illustrated on data from a human vision study.

E1242: **Modelling binomial, Poisson and gamma observations spatially distributed over irregularly shaped domains**
Presenter: **Laura Sangalli**, Politecnico di Milano, Italy
Co-authors: Matthieu Wilhelm

A novel method is introduced for the analysis of spatially distributed data from an exponential family distribution, able to efficiently deal with data occurring over irregularly shaped domains. The proposed generalized additive framework can handle all distributions within the exponential family, including binomial, Poisson and gamma outcomes, hence leading to a very broad applicability of the model. Specifically, we maximize a penalized log-likelihood function where the roughness penalty term involves a suitable differential operator of the spatial field over the domain of interest. Space-varying covariate information can also be included in the model in a semi-parametric setting. The proposed model exploits advanced numerical analysis techniques, and specifically makes use of the Finite Element Method, that provide a basis for piecewise polynomial surfaces.

ES72 Room Gordon ADVANCES AND ISSUES IN CREDIT SCORING Chair: Gilbert Saporta

E451: **Satisfaction, human values and other intangible dimensions as drivers of new credit scoring models**
Presenter: **Furio Camillo**, University of Bologna, Italy
Co-authors: Caterina Liberati

Intangible values have not been exploited in credit risk modeling. More specifically the usage of the information coming from personal values reputation and other intangible assets can be used as loan collateral to distressed borrowers that lack sufficient predictive risk indicators. This is particularly true for specific customer segments: high revenue individuals (affluent segments) not involved in corporations. Lending against intangibles is a recent credit market innovation, and in recent literature we found evidence consistent with the fact that this credit practice has not been a negative mutation in credit markets. A predictive kernel discriminant has been performed, matching marketing data (opinions survey and individual semiometrics) with usual individual credit scoring drivers. Results highlight standard risk indicators express only the high level of affluent segment homogeneity. Marketing data have been crucial in prediction and interpretation of risk factors especially for adjusting bank

CRM actions according to individual credit scoring. Data complexity, related to characteristics of affluent segment which is very spread in Italian financial system, has been treated ad hoc: the selection of relevant variables and the model identification have been combined simultaneously via a linear reconstruction of a kernel rule.

E684: **Penalized and ensemble methods in scoring**
Presenter: **Stephane Tuffery**, Rennes University - ENSAI, France
Penalized regression methods (ridge, lasso, group lasso, elastic net), and particularly penalized logistic regression, have proved their efficiency in high dimensional statistics. We will show that they are also useful in the more standard situations of scoring, where the number of predictors is much smaller than the sample size. From the perspective of optimizing the predictive power, greater attention will be paid to ridge regression than lasso regression, inasmuch as the lasso selection of relevant variables may impose greater penalty than that which optimizes the predictive power. These methods allow for the consideration of a larger set of predictors to obtain as much useful information as possible to predict the response, without seeing the appearance of the overfitting phenomenon or inconsistent coefficients, as would result in a non-penalized model. Moreover, the penalty parameter allows for the adjusting of the bias-variance tradeoff of the model, as the number of predictors selected at each step of a random forest aggregated model. In this latter method, the bias-variance tradeoff can be smoothly adjusted by selecting the size of the subset of predictors randomly drawn at each iteration. We also show that the random forest mechanism, thanks to de-correlation between the successive models, can be applied to strong base classifier, such as the logit, as opposed to the bagging methods. Furthermore, not only the predictor but also its cut-points can be randomly chosen at each iteration.

E686: **A two stage mixture model for predicting EAD**
Presenter: **Jonathan Crook**, University of Edinburgh, United Kingdom
Co-authors: Mindy Leow
Using a large portfolio of defaulted loans and their historical observations, we estimate EAD on the level of obligors by estimating the outstanding balance of an account, not only for the account at the time of default, but at any time over the entire loan period. We theorize that the outstanding balance on a credit card account at any time during the loan is not only a function of the spending and repayment amounts by the borrower, but is also subject to the credit limit imposed by the card issuer. Therefore, we predict for outstanding balance using a two-step mixture model. We first develop a discrete-time repeated events survival model to predict for the probability of an account having a balance greater than its limit. Next, we develop two panel models with random effects to estimate for balance and limit, where the final prediction for balance would be a product of the estimated balance and limit, depending on how likely the borrower is to have a balance greater than the limit. Using this mixture model, we find that we are able to get good predictions for outstanding balance, not only at the time of default, but at any time over the entire loan period, which would allow us to make predictions for outstanding balance and hence EAD before default occurs, for delinquent accounts. Overall r-square values achieved are 0.46 when looking over the entire loan period for all delinquent accounts, and 0.55 when only looking at default-time observations.

E755: **Reject inference techniques and semi supervised methods implemented in the process for granting credit**
Presenter: **Asma Guizani**, Computational Mathematics Laboratory, Tunisia
Co-authors: Besma Souissi, Salwa Benammou, Gilbert Saporta
Credit scoring techniques are used to estimate the probability of default for loan applicants in the process of granting credit. Like any statistical model, scoring is built based on historical data to help predict the creditworthiness of applicants. The method has the drawback of not estimating the probability of default for refused applicants which means that the results are biased when the model is built on only the accepted data set. Reject inference techniques attempt to solve the problem of selection bias by estimating the risk of default of the rejected applicants. First, we present some reject inference techniques like re-weighting, iterative reclassification and parceling, then, techniques that belong to semi-supervised learning mixing labelled and unlabelled data are considered. These techniques are related to mixed classification, adaboost algorithm and gentle adaboost algorithm. ROC analysis is used to compare the performances of different techniques.

ES74 Room 349 TIME SERIES EXTREMES Chair: Johan Segers

E282: **Extremal properties of stochastic volatility models**
Presenter: **Anja Janssen**, University of Hamburg, Germany
Discrete-time stochastic volatility (SV) models have become a standard tool for the modeling of economic time series as they are able to reflect many of the well-known stylized facts of financial markets. With regard to their extremal behavior, the standard SV model specifications have in common that consecutive observations are asymptotically independent and thus their extremal index is equal to 1. However, on a pre-asymptotic level SV models may still show a clustering of large values and we are therefore interested in the second order behavior of extremal dependence. Different concepts may be applied which allow for a finer analysis of asymptotic independence: See, in particular, the coefficient of tail dependence and the notion of hidden regular variation. However, the standard model specifications for SV models do not reflect the broad spectrum of possible second order behavior since their asymptotic properties are mainly determined by the heavy-tailed i.i.d. innovations. With a view towards empirical results for real life processes we suggest an alternative model which allows for more freedom in the asymptotic dependence structure. We analyze this model in the framework of hidden regular variation under the use of suitable general results.

E416: **On copulas, quantiles, ranks and spectra: An L1-approach to spectral analysis**
Presenter: **Stanislav Volgushev**, Ruhr University Bochum, Germany
Co-authors: Holger Dette, Marc Hallin, Tobias Kley
An alternative method for the spectral analysis of a strictly stationary time series is discussed. A "new" spectrum is defined as the Fourier transform of the differences between copulas of the pairs with lag k and the independence copula. This object is called copula spectral density kernel and allows us to separate marginal and serial aspects of a time series. It is shown that it is intrinsically related to the concept of quantile regression. Like in quantile regression, which provides more information about the conditional distribution than the classical location-scale model, the copula spectral density kernel is more informative than the spectral density obtained from the auto-covariances. In particular the approach allows us to analyse the tails of a time series by considering high or low quantiles. In order to estimate the copula spectral density kernel rank-based Laplace periodograms are introduced, which are calculated as bilinear forms of weighted L1-projections of the ranks of the observed time series onto a harmonic regression model. The asymptotic properties of the proposed estimators are commented upon and several possible extensions are discussed.

E564: **Statistical analysis of extremal serial dependence of time series**
Presenter: **Holger Drees**, University of Hamburg, Germany
Modeling the dependence between consecutive observations in a time series plays a crucial role in risk management. For example, the risk of large losses from a financial investment is increased if extreme negative returns tend to occur in clusters, and heavy rainfall on several consecutive days could trigger a catastrophic flooding. The aim is to recall the so-called coefficient of tail dependence as an important measure of the strength of serial dependence between extremes which allows for a refined characterization of dependence structures, especially in the case of asymptotic

independence. A general class of empirical processes introduced previously enables us to analyze the asymptotic behavior of estimators of the coefficient of tail dependence in a unified framework. Bootstrap versions of these empirical processes yield asymptotic confidence intervals. In an application it is shown how to use these results to discriminate between time series of GARCH-type and time series from common stochastic volatility models. An analysis of a time series of returns of the German blue stocks index however reveals that probably none of these time series models describe the extremal dependence structure accurately.

Monday 16.12.2013 10:55 - 13:00 Parallel Session P – CFE

CS26 Room Bloomsbury A TIME SERIES APPROACH TO RISK MANNAGEMENT Chair: Christophe Chorro

C398: **Construction, management, and performance of sparse Markowitz portfolios**
Presenter: **Juan-Pablo Ortega**, CNRS Universite de Franche-Comte, France
Co-authors: Julie Henriques
The aim is to study different implementations of the sparse portfolio construction and rebalancing method introduced in a previous work. This technique is based on the use of a l_1-norm (sum of the absolute values) type penalization on the portfolio weights vector that regularizes the Markowitz portfolio selection problem by automatically eliminating the dynamical redundancies present in the time evolution of asset prices. Specific recommendations are made as to the different estimation techniques for the parameters needed in the use of the method and its good performance is proven in realistic situations involving different rebalancing frequencies and transaction costs. The empirical findings show that the beneficial effects of the use of sparsity constraints are robust with respect to the choice of trend and covariance estimation methods used in its implementation.

C522: **Testing for leverage effect in non linear financial time series**
Presenter: **Hanjarivo Lalaharison**, Pantheon-Sorbonne-Paris, France
Co-authors: Christophe Chorro, Dominique Guegan, Florian Ielpo
The finite distance properties of three likelihood-based estimation strategies for GARCH processes with non-Gaussian conditional distributions are discussed: (1) the maximum likelihood approach; (2) the quasi maximum likelihood approach; (3) a multi-steps recursive estimation approach (REC). A Monte Carlo test is first run which shows that the recursive method may be the most relevant approach for estimation purposes. It then turns to a sample of SP500 returns. It is confirmed that the REC estimates are statistically dominating the parameters estimated by the two other competing methods. Regardless of the selected model, REC estimates deliver the more stable results. Finally, this new empirical estimation method is used to question the importance of introducing a leverage effect parameter in GARCH-type specifications. It is found that the in and out of sample forecast performances are not improved by this extra parameter once an appropriate asymmetric distribution is used for the residuals.

C826: **Wavelet shrinkage of noisy chaotic signals and risk measures**
Presenter: **Matthieu Garcin**, Universite Paris I - Pantheon Sorbonne, France
Co-authors: Dominique Guegan
By filtering wavelet coefficients, one is able to construct a good estimator of a pure signal from noisy data. For a simple linear noise influence, we already know well an optimal filter design in the sense of a good reconstruction of the pure signal. We propose a more realistic framework where the influence of the noise is non-linear. In particular, we propose an optimal method to filter the wavelet coefficients of a noisy discrete dynamical system in order to construct a good estimator of the pure signal. Some examples with simple chaotic dynamical systems are presented. Moreover, this technique is applied to de-noise financial signals and to build specific risk measures.

C851: **Derivative pricing with non-Gaussian GARCH models and their continuous time limits**
Presenter: **Alex Badescu**, Universite du Quebec a Montreal, Canada
Co-authors: Robert Elliott, Juan-Pablo Ortega
The weak convergence of risk-neutralized non-Gaussian GARCH models to continuous-time stochastic volatility models is investigated. A discretized version of a general Girsanov principle is used for deriving the pricing models within an asymmetric GARCH setting. When the driving noise is Gaussian distributed, we recover a well-known limiting result which was obtained by applying the minimal martingale measure. However, the skewness of the GARCH innovations induces a non-zero market price of volatility risk for its corresponding risk-neutralized continuous-time limit. We provide extensive simulation exercises to analyze the numerical convergence of GARCH option and variance swap prices to their bivariate diffusion counterparts.

CS99 Room Woburn ADVANCES IN FINANCIAL ECONOMETRICS Chair: Richard Luger

C1084: **VaR-implied tail-correlation matrices**
Presenter: **Stefan Mittnik**, University of Munich, Germany
Empirical evidence suggests that asset returns correlate more strongly in bear markets than conventional correlation estimates imply. To estimate such phenomena, a method for the pairwise derivation of tail-correlation coefficients, on the basis of Value-at-Risk (VaR) estimates, has been previously proposed. We generalize and extend this approach by going beyond pairwise estimation. This enables us to directly determine complete tail-correlation matrices and has several advantages. First, restrictions on the correlation matrix, such as positive semidefiniteness, a property required for valid risk aggregation and Markowitz-type portfolio optimization, can be imposed straightforwardly. Second, more efficient estimates can be obtained by using overidentification strategies. Moreover, closed-form expressions for the complete tail-correlation matrix facilitate analytical tractability. Properties of alternative strategies for estimating VaR-implied correlation matrices are investigated. An empirical application to a 30-asset universe illustrates the practical applicability and relevance of the approach in portfolio applications.

C1037: **The cross-quantilogram: Measuring quantile dependence and testing directional predictability between time series**
Presenter: **Heejoon Han**, Kyung Hee University, Korea, South
Co-authors: Oliver Linton, Tatsushi Oka, Yoon-Jae Whang
The cross-quantilogram, which measures the quantile dependence between stationary time series based on the correlogram of quantile hits, is considered. It can be applied to test the hypothesis that one time series has no directional predictability to another time series. We derive its asymptotic distribution, which is not free of nuisance parameters. To construct credible confidence intervals, we first propose the stationary bootstrap method and prove that it provides asymptotically valid confidence intervals. We investigate the finite sample performance of this procedure and show that it works well. We also define a self-normalized version of the cross-quantilogram. The self-normalized version has an asymptotically pivotal distribution whose critical values have been tabulated so that there is no need for long run variance estimation or even bootstrap. We provide two empirical applications. First, we use the cross-quantilogram to detect predictability from stock variance to excess stock return. Compared to existing tools, our method provides more complete relationship between a predictor and stock return. Second, we apply the cross-quantilogram to the question of systemic risk. We investigate the systemic risk and market exposure of individual financial institutions, such as JP Morgan Chase, Goldman Sachs and AIG.

C1043: **Simultaneous confidence intervals for dynamic default rates**
Presenter: **Jinyu Yu**, Georgia State University, United States
Co-authors: Richard Luger
Statistical inference is considered in the context of a generalized version of the widely used Vasicek credit default model, whereby the purely

static model is extended to allow for autocorrelated default rates and macroeconomic risk factors. The proposed inference method proceeds by numerically inverting a likelihood ratio test and then exploits projection techniques to produce simultaneous confidence intervals for general non-linear functions of the model parameters, including multi-step ahead expected losses. An extensive simulation study reveals that the new method outperforms the delta method and even the usual parametric and non-parametric bootstrap procedures. The results of an empirical application to U.S. bank loan losses show that moving from the static to the dynamic default rate distribution significantly lowers the implied economic capital requirements.

C204: **Measure of downturn connectedness and systemic risk in financial institutions**
Presenter: **Hui-Ching Chuang**, Yuan Ze University, Taiwan
A series of systemic risk measures is proposed which are designed to capture the cluster of the tail events among financial institutions. These measures are based on the network topology constructed by the Granger-causality in risk (GCR) test which examines whether the occurrence of a large downside risk in one institution can help to predict the occurrence of a large downside risk in another institution. Empirical results show that the individual outward spillover measure is positively related to the size of financial institutions in U.S. from 1994 to 2012. In addition, the total degree of the GCR measure is highly correlated with the downturn market conditions, such as credit spread, recession probability of NBER and the estimated financial sector default rates.

C1145: **A simple NIG-type approach to calculate value at risk based on realized moments**
Presenter: **Christian Lau**, Martin Luther University, Germany
An alternative model for VaR calculation based on realized variance (RV) is presented. It is straightforward to expand the RV concept by realized skewness and realized kurtosis. Using a simple EWMA approach, we calculate one-day forecasts for the moments. Given the forecasts, the parameters of a sophisticated distribution can be estimated with methods of moments. The normal inverse Gaussian distribution is a feasible candidate, because it exhibits higher moments and an easy analytical solution for parameter calculation with methods of moments exits. Using this technique, we calculate VaR for the DAX. Although this model is comparatively simple, the empirical study shows good results in terms of backtesting.

CS42 Room Bedford UNCERTAINTY AND REAL-TIME TURNING POINTS DETECTION I Chair: Gianluigi Mazzi

C412: **Forecasting recessions in real time**
Presenter: **Anne Sofie Jore**, Norges Bank, Norway
Co-authors: Knut Are Aastveit, Francesco Ravazzolo
The aim is to review several methods to define and forecast classical business cycle turning points in Norway, a country which does not have an official business cycle indicator. It compares the Bry and Boschan rule (BB), an autoregressive Markov Switching model (MS), and Markov Switching models (FAMS) augmented with a factor or alternative surveys, respectively, using several vintages of Norwegian Gross Domestic Product as the business cycle indicator. Timing of the business cycles depends on the vintage and the method used. BB provides the most reliable definition of business cycles. A forecasting exercise is also presented: the BB applied to density forecasts from survey models provides more timely predictions for Norwegian turning points than the Markov Switching models both using final vintage and, above all, real-time data.

C482: **Real time trend-cycle prediction via linear filters that minimize revisions and turning point detection**
Presenter: **Silvia Bianconcini**, Bologna, Italy
Co-authors: Estela Bee Dagum
Chronologies of business cycle turning points are currently maintained in the United States by the National Bureau of Economic Research (NBER), and in Europe by the Centre for Economic Policy Research (CEPR). The identification of a new turning point in the economy requires additional data, which is only available after the turning point is reached. As a consequence, these turning points can only be identified with a time lag. For real time analysis, official statistical agencies analyze trend-cycle estimates, generally derived using asymmetric moving average techniques. But the use of these nonparametric asymmetric filters introduces revisions as new observations are added to the series, and delays in detecting true turning points. The aim is to consider a reproducing kernel representation of commonly applied nonparametric trend-cycle predictors to derive asymmetric filters that monotonically converge to the corresponding symmetric one. Three specific criteria of bandwidth selection are considered, namely: (1) minimization of the transfer function which implies an optimal compromise between reducing revisions and phase shift; (2) minimization of the gain function which implies revisions reduction, and (3) minimization of the phase shift function which implies reduction of the time lag to detect a true turning point. Hence, a family of real time trend-cycle predictors is obtained with the important properties of either minimization of revision, or fast detection of turning points or an optimal compromise between these two. The behavior of the proposed procedure is illustrated with real series mainly used to identify and predict true macroeconomic turning points.

C1216: **Detecting and predicting economic accelerations, recessions, and normal growth periods in real-time**
Presenter: **Christian Proano**, The New School for Social Research, United States
The dichotomous characterization of the business cycle in recessions and expansions has been the key over the last fifty years. However, there are various reasons, both theoretical and empirical, to question the adequacy of this dichotomous, recession/expansion approach for our understanding of the business cycle dynamics, as well as for prediction of future business cycle developments. In this context, the contribution is twofold: On the one hand, a positive rate of growth in the level of economic activity can be considered as the normal state of modern economies due to both population and technological growth, it proposes a new non-parametric algorithm for the detection and dating of economic acceleration periods which differentiate themselves from periods of "normal growth", and economic recessions. On the other hand, it uses a composite ordered probit model for the estimation and forecasting of these three business cycle phases using an automatized model selection and forecast combination approach using monthly macroeconomic and financial data on the German economy. The empirical results do not only show the empirical relevance of this new algorithm, but also significant asymmetries in the determinants of the different business cycle phases.

C503: **Balance sheet recessions and time-varying coefficients in a Phillips curve relationship: An application to Finnish data**
Presenter: **Katarina Juselius**, University of Copenhagen, Denmark
A modified Phillips curve has been previously introduced where the natural rate of unemployment is a function of the real interest rate instead of a constant. It was argued that the effect of the interest rate on the macro economy is likely to be diluted during a balance sheet recession such as the ones recently seen in many countries. In the late eighties, after having deregulated credit and capital movements, Finland experienced a housing boom which subsequently developed into a serious economic crisis similar to the recent ones. To learn from the Finnish experience the aim is to estimate the modified Phillips curve and use a Smooth Transition model to distinguish between ordinary periods and balance sheet recessions.

C486: **A system for euro area and member states turning point detection**
Presenter: **Gian Luigi Mazzi**, Eurostat, Luxembourg
Co-authors: Monica Billio, Laurent Ferrara
The availability of a timely and reliable system for detecting turning points is essential for the correct follow up of the cyclical situation and it is crucial for policy and decision makers in their daily work. Usually, turning points detection systems identify them either on the classical or on the

growth cycle. A multivariate model is presented detecting simultaneously turning points of both cycles within the so called ABCD framework. The chosen model is a Markov-switching VAR with multiple regimes to capture different phases of both reference cycles. The model is specified by minimising error two type and maximizing the timeliness. The model is simulated in real-time by using the chronologies for both classical and growth cycle as a benchmark. The concordance index and the QPS are used to discriminate among different model specifications. It is also shown how the model, originally developed for the euro area, has been extended to its largest countries. Country economic specificities and data availability have been taken into account building up national models. Finally, a real-time comparison of direct and indirect turning points detection for the euro area is presented and the resulting discrepancies are discussed.

CS49 Room Gordon RECENT ADVANCES IN FORECASTING FINANCIAL TIME SERIES Chair: Katerina Panopoulou

C174: **Forecasting exchange rate volatility: Multivariate Realized GARCH Framework**
Presenter: **Janine Balter**, Saarland University, Germany
Co-authors: Elena-Ivona Dumitrescu, Peter Reinhard Hansen

A new simple and parsimonious way to jointly model and forecast the returns, realized and conditional measures of volatility and correlation of different foreign exchange markets, is proposed. Our multivariate realized GARCH model combines the simplicity and flexibility of the GARCH approach and the advantages of high-frequency (intraday) data in terms of statistical precision in the estimation of volatility. It is not prone to efficiency issues as the existing approaches considered in this literature (two-step approaches or Dynamic Conditional Correlation models) while it accounts for the spillover effects between the markets under analysis. It hence makes it possible to investigate the dynamics of exchange rates volatility and the propagation of shocks in foreign exchange markets during both crisis and calm periods. In this context, it is possible to identify which market is at the origin of a shock and it gives us an understanding about the markets' asymmetry in response to good/bad news. Most importantly, the multivariate realized GARCH model allows us not only to analyze the volatility spillovers between the exchange markets, but also to forecast the volatility and correlations one-step and multi-step ahead. An empirical application of this methodology in the case of the euro / dollar / yen exchange rates is provided. The superior out-of-sample performance of our method with respect to that of existing models indicates that it is a useful tool for addressing risk in cross-border economic transactions.

C232: **Measuring the market risk of freight rates: A forecast combination approach**
Presenter: **Christos Argyropoulos**, University of Piraeus, Greece
Co-authors: Ekaterini Panopoulou

The aim is to contribute in three ways: First, we re-evaluate the performance of popular Value-at-Risk (VaR) estimation methods on freight rates amid the adverse economic consequences of the recent financial and sovereign debt crisis. Second, we provide a detailed and extensive backtesting methodology in order to identify possible weaknesses associated with the standard backtesting criteria. A newly proposed method is employed in order to evaluate the performance of each method. Last, we propose a combination forecast approach for estimating VaR. Our findings suggest that both the parametric and simulation methods produce accurate estimates of daily VaR. More importantly, our combination methods produce more accurate estimates at both the 1% and the 5% significance level and for all the sectors under scrutiny, while in some cases they may be viewed as conservative since they tend to overestimate nominal VaR.

C572: **Out-of-sample equity premium prediction: A complete subset quantile regression approach**
Presenter: **Spyridon Vrontos**, University of Piraeus, Greece
Co-authors: Loukia Meligkotsidou, Ekaterini Panopoulou, Ioannis Vrontos

A quantile regression approach is proposed to equity premium forecasting based on complete subset combinations of forecasts. The proposed approach extends the complete subset mean regression framework to a quantile regression setting, which enables us to construct robust and accurate equity premium forecasts. It also puts forward a recursive algorithm that selects, in real time, the best complete subset for each predictive regression quantile. Then, robust point forecasts of the equity premium are based on the synthesis of these 'optimal' complete subset quantile forecasts. The approach delivers statistically and economically significant out-of-sample forecasts relative to both the historical average benchmark and the complete subset mean regression approach.

C577: **Volatility forecasts and the MSE criterion**
Presenter: **Theologos Pantelidis**, University of Macedonia, Greece
Co-authors: Nikitas Pittis

The aim is to explain the poor forecasting performance of GARCH models reported in many empirical studies. A number of previous applied studies compare the conditional volatility forecasts of GARCH models to those of a model that assumes constant conditional variance (homoscedasticity) and conclude that the GARCH models do not outperform the homoscedastic one. In these studies, the comparison of the relative forecasting accuracy of competing models is usually based on the Mean Squared Error (MSE) criterion. Some recent theoretical studies argue that the poor forecasting performance of GARCH models found in empirical studies is deceptive, since the utilization of the squared shocks as a proxy for the unobserved true conditional volatility increases the MSE of the volatility forecasts (MSE-inflation), rendering the MSE criterion invalid for proper evaluation of the forecasting accuracy of GARCH models. However, the presented analytical results show that the utilization of the squared shocks as a proxy for the unobserved true conditional volatility "inflates" equally the MSEs of the forecasts of the GARCH and homoscedastic models and thus the MSE criterion remains capable to evaluate properly the relative forecasting performance of the two models.

C575: **High frequency risk measures**
Presenter: **Denisa Georgiana Banulescu**, University of Orleans, France
Co-authors: Christophe Hurlin, Gilbert Colletaz, Sessi Tokpavi

A high frequency risk (HFR) measure is proposed whose aim is to quantify both the risk on durations and the risk on price returns, while taking into account the irregular timing of trades or any other type of microstructure event. It consists of estimating a couple of two measures: an intraday irregularly spaced Value at Risk (ISIVaR) defined in event time, and a liquidity risk indicator (i.e. the Time-at-Risk (TaR) measure) used to forecast the expected duration between two consecutive trades. A backtesting procedure is subsequently implemented to assess the accuracy of the forecasts series. The empirical application is performed for three financial stocks, namely SP500, Bank of America and Microsoft. First, the results confirm the theoretical inverse relationship between the clustering of price durations and volatility returns: the shorter the price durations, the more the volatility returns is important. It hence implies that a more intense trading activity could imply more frequent revisions on prices and vice versa. Second, the procedure performs well, which means that overall the ISIVaR and TaR forecasts are quite accurate. Finally, the results are robust to the change of the asset, the coverage rate, or the backtesting approach.

CS115 Room 104 FINANCIAL APPLICATIONS V **Chair: Matthias Scherer**

C1023: **Dividend predictability in large markets: A mixed data sampling approach**
Presenter: **Nikolaos Kourogenis**, University of Piraeus - Research Center, Greece
Co-authors: Panagiotis Asimakopoulos, Stylianos Asimakopoulos, Emmanuel Tsiritakis
One of the main theoretical implications of the present value approach on firm valuation is the hypothesis that dividend yield has a predictive power on future dividend growth. The relevant literature, however, was not able to provide evidence that clearly supports this hypothesis. We cope with the two main reasons that raise the econometric complexity on testing the dividend growth predictability hypothesis, namely, the seasonality effects that appear when higher frequency data are used, and the effect of price volatility on the computation of dividend yield. Specifically, an application of a the Mixed Data Sampling (MiDaS) technique allows us to use monthly dividend data in order to test the hypothesis that dividend yield explains the future annual dividend growth. In order to cancel out the effects of price variance on dividend yield we use a smoothing technique, and we identify the component of the smoothed dividend yield that offers predictive power. Empirical evidence from US, UK, Japan, Germany, France, Italy and Canada, strongly supports the dividend growth predictability hypothesis.

C1067: **Macroprudential oversight, risk communication and visualizations**
Presenter: **Peter Sarlin**, Abo Akademi University, Finland
The role of visualizations in macroprudential oversight at large, and especially for the purpose of risk communication, is dicussed. Risk communication is divided into two tasks: internal and external. Internal communication concerns spreading information about systemic risks within, but at various levels of, the organization, such as among divisions, groups or analysts, whereas external communication refers to the task of dissemination information about systemic risks to the general public. The term visualization has a wide meaning and relates to a number of interdisciplinary disciplines, in particular information visualization and visual analytics. We mainly focus on the background and theory of information visualization and visual analytics, as well as techniques provided within these disciplines. Then, we define the task of visualization in internal and external risk communication, as well as turns to a discussion about the type of available data for identifying and assessing systemic risk in macroprudential oversight. This provides a broad overview of which tasks visualizations should support and what are the underlying data to be visualized. Finally, the paper ends with two applications of mapping techniques on high-dimensional systemic risk indicators.

C1068: **CAPM, components of beta and the cross section of expected returns**
Presenter: **Tolga Cenesizoglu**, HEC Montreal and LSE, Canada
Co-authors: Jonathan Reeves
It is demonstrated that a conditional version of the Capital Asset Pricing Model (CAPM) explains the cross section of expected returns, just as well as the three factor model of Fama and French. This is achieved by measuring beta (systematic risk) with short-, medium- and long-run components. The short-run component of beta is computed from daily returns over the prior year. While the medium-run beta component is from daily returns over the prior 5 years and the long-run component from monthly returns over the prior 10 years. More immediate changes in risk such as changes in portfolio characteristics are captured in the short-run beta component, whereas, more slowly changing risk due to the business cycle is captured in the medium- and long-run beta components.

C1174: **Accounting for large losses in portfolio selection**
Presenter: **Alexandra Dias**, University of Leicester, United Kingdom
During financial crises equity portfolios have suffered large losses. Methodologies for portfolio selection taking into account the possibility of large losses have existed for decades but their economic value is not well established. The economic value in reducing the probability of large losses in portfolio selection is investigated. We combine mean-variance analysis with semi-parametric estimation of potential portfolio large losses. We find that strategies that reduce the probability of large losses outperform efficient minimum variance portfolios, especially when semi-parametric estimation is used. Our results are robust to transaction costs.

C1205: **Mathematical models for inverse problems arising from financial markets**
Presenter: **YASUSHI OTA**, Doshisha, Japan
When we apply the Black-Scholes model to financial derivatives, one of most interesting problems is reconciling the deviation between the expected and observed values. In particular, if we apply its model to default probability estimation, under the no arbitrage property of financial markets, since we cannot utilize only the equation which does not include a real drift coefficient to estimate the default probability, we must be careful of the deviation. Taking this deviation into account, we derived a new mathematical model based on the Black-Scholes model, formed an arbitrage model, and formulated a new mathematical approach for inverse problems arising from financial markets. In our new model and approach, we have derived the Black-Scholes type equation of the form containing a real drift coefficient and found the mathematical method that using market data enables us to reconstruct a real drift coefficient. For our model we applied the linearization technique and obtained the useful integral equation. Moreover we demonstrated the stability of the linearized problem, and confirmed the existence of arbitrage in the financial markets.

C635: **The information content of forward skewness inferred from option portfolios**
Presenter: **Anastasios Kagkadis**, Durham University, United Kingdom
Co-authors: Panayiotis Andreou, Dennis Philip
The aim is to create a term structure of risk-neutral third moments using portfolios of the S&P 500 index option contracts and extract risk-neutral forward skewness measures of monthly market returns. We explore the ability of forward skewness to predict a wide range of macroeconomic variables, asset prices, as well as crash risk, uncertainty and investor sentiment variables over and above forward variance. Predictive ability is gauged using both individual and joint parameter statistical significance as well as increase in the explanatory power of the models when forward skewness coefficients enter in the predictive model specifications. The results show that forward skewness is particularly successful in predicting growth in variables related to real activity, consumption, money supply, credit and inflation as well as changes in exchange rates, with forecasting power increasing with the horizon. Moreover, it improves the predictability of stock market returns for horizons between three and six months ahead and also helps to forecast equity market tail risk, systemic risk, uncertainty and investor sentiment based on mutual fund flows for horizons up to twelve months ahead. Overall, the results provide evidence that forward skewness encapsulates important information about future macroeconomic and asset market conditions over and above forward variance.

CS79 Room Montague TOPICS IN TIME SERIES AND PANEL DATA ECONOMETRICS Chair: Martin Wagner

C022: **Monitoring stationarity and cointegrating relationships**
Presenter: **Dominik Wied**, TU Dortmund, Germany
Co-authors: Martin Wagner
A monitoring procedure for stationarity and cointegration relationships is proposed. Under the null hypothesis, we have a stationary or cointegrating process, while, under the alternative, there is some point after a historical period at which the process jumps to non-stationarity or looses the contegration property. The detector of the procedure relies on comparing suitable partial sums of regression residuals. We provide analytic results of the asymptotic behavior of the procedure under both the null hypothesis and the alternative, investigate the finite sample performance by Monte Carlo simulations and apply it to several macroeconomic datasets.

C1201: **Cross-sectional averages versus principal components**
Presenter: **Jean-Pierre Urbain**, Maastricht University SBE, Netherlands
Co-authors: Joakim Westerlund
In spite of the increased use of factor-augmented regressions in recent years, little is known from a theoretical point of view regarding the relative merits of the two main approaches to estimation and inference, namely, the cross-sectional average and principal components estimators. Light is shed on this issue.

C127: **On size and power of heteroscedasticity and autocorrelation robust tests**
Presenter: **David Preinerstorfer**, University of Vienna, Austria
Co-authors: Benedikt Poetscher
Testing restrictions on regression coefficients in linear models often requires correcting the conventional F-test for potential heteroskedasticity or autocorrelation amongst the disturbances, leading to so-called heteroskedasticity and autocorrelation robust test procedures. These procedures have been developed with the purpose of attenuating size distortions and power deficiencies present for the uncorrected F-test. We develop a general theory to establish positive as well as negative finite-sample results concerning the size and power properties of a large class of heteroskedasticity and autocorrelation robust tests. Using these results we show that nonparametrically as well as parametrically corrected F-type tests in time series regression models with stationary disturbances have either size equal to one or nuisance-infimal power equal to zero under very weak assumptions on the covariance model and under generic conditions on the design matrix. In addition, we suggest an adjustment procedure based on artificial regressors. This adjustment resolves the problem in many cases in that the so-adjusted tests do not suffer from size distortions. At the same time their power function is bounded away from zero. As a second application we discuss the case of heteroskedastic disturbances.

C1002: **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 has been widely used in applied macroeconomics for over two decades. An analysis of standard tests like the Phillips-Perron (PP) or Augmented-Dickey-Fuller (ADF) tests in the presence of large initial values is provided. The asymptotic behavior is derived in the stationary, the unit-root, and the near integrated case. In a small 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 PP/ADF distributions, causing size and power distortions of standard PP and ADF tests. The economic implications of these findings are demonstrated in an application using annual growth data for several OECD countries.

C026: **Some extensions of regression based cointegration analysis**
Presenter: **Martin Wagner**, Technical University Dortmund, Germany
The analysis of cointegrating relationships in a regression framework is typically carried out using modified least squares estimators that employ corrections for the effects of endogeneity and error serial correlation in order to obtain limiting distributions that allow for asymptotic standard inference. Several such estimation procedures are available in the literature. We discuss extensions of such approaches along two dimensions. On the one hand, we discuss the applicability of modified least squares estimators in cointegrating regressions that are linear in parameters, but nonlinear in I(1) variables. Typical examples of such relationships are (environmental) Kuznets curves or translog cost or production functions. Especially the latter poses new challenges for estimation and inference using when trying to use, e.g. the so-called fully modified OLS estimation principle. On the other hand, we discuss estimation when it is not the functional form that is allowed to be more general, but the dynamic properties of the variables. In particular, we discuss cointegration analysis based not on summed up i.e. integrated stationary but integrated locally stationary processes. We consider localized fully modified OLS estimation for this case.

CS106 Room 349 TIME SERIES ECONOMETRICS III Chair: Gareth Peters

C670: **Identification in structural VAR models with different volatility regimes**
Presenter: **Emanuele Bacchiocchi**, University of Milan, Italy
The aim is to propose a new specification for structural VAR models that accounts for possible heteroskedasticity in the observable variables. The model allows for different propagation mechanisms in different volatility regimes. Identification rules for the structural parameters are provided when the information coming from the heteroskedasticity observed in the data is combined with that suggested by the economic theory in the form of traditional parameters constraints. A *structure condition* is developed, which only depends on the structure of the constraints and the form of heteroskedasticity characterizing the data. It is shown that this condition, if combined with an appropriate order condition, provides a sufficient condition for identification. The practical implementation, based on checking for the non-singularity of a certain matrix, is simple, and does not require the estimation of the unknown parameters. The empirical relevance of the methodology is discussed through a set of examples and an empirical application concerning the relationships between term structure of interest rates and output growth.

C744: **On the moving quantile effects in financial time series**
Presenter: **Virmantas Kvedaras**, Vilnius University, Lithuania
Co-authors: Isao Ishida
The purpose is to introduce and investigate analytically some properties of a class of nonlinear time series models with the moving order statistics present in the data generating process. This endogenizes the regime changes and allows for non-linear size effects, e.g. where the impact of extreme and ordinary events substantially differs. We show by simulations that such effects can produce realizations looking as if the structural breaks were present in the data and having substantially flatter sample autocorrelation functions. Since the usual tests for omitted non-linearity have insufficient power against such type of non-linearity, a suitable test is proposed. Some empirical illustrations are presented.

C1040: **Adaptive trend estimation in financial time series via multiscale change-point-induced basis recovery**
Presenter: **Anna Louise Schroder**, LSE, United States
Co-authors: Piotr Fryzlewicz

Low-frequency financial returns can be modelled as centered around piecewise-constant trend functions which change at certain points in time. We propose a new stochastic time series framework which captures this feature. The main ingredient of our model is a hierarchically-ordered oscillatory basis of simple piecewise-constant functions. It differs from the Fourier-like bases traditionally used in time series analysis in that it is determined by change-points, and hence needs to be estimated from the data before it can be used. The resulting model enables easy simulation and provides interpretable decomposition of non-stationarity into short- and long-term components. The model permits consistent estimation of the multiscale change-point-induced basis via binary segmentation, which results in a variable-span moving-average estimator of the current trend, and allows for short-term forecasting of the average return.

C1051: **Evaluation of volatility predictions in a VaR framework**
Presenter: **Vincenzo Candila**, University of Salerno, Italy
Co-authors: Alessandra Amendola

An approach to indirectly assess the volatility predictions generated by a set of competing models is looking at their Value at Risk (VaR) measures. Unfortunately, the traditional tests for their evaluation suffer from low statistical power. Recently, a loss function approach has been proposed, based on the comparison of profits and losses of a portfolio with the VaR. A new asymmetric loss function that overpenalizes the models with a larger number of VaR violations is proposed. To test its performance, a Monte Carlo experiment is carried out comparing a set of models with different characteristics, one of which is closer to the true data generating process (DGP). The inference is based on the distribution of loss function for the DGP. The accuracy measure matches the occurrences of rejections for each model with those of the DGP. The proposed asymmetric loss function yields a better accuracy. Moreover, a good performance is reached for what concerns the rankings: the frequency at which the model closer to the DGP has the smallest loss function value increases with the use of the asymmetric loss function.

C1062: **Identifying causal relationships in case of non-stationary time series**
Presenter: **Angeliki Papana**, University of Macedonia, Greece
Co-authors: Catherine Kyrtsou, Dimitris Kugiumtzis, Cees Diks

The standard linear Granger non-causality test is effective only when time series are stationary. In case of non-stationary data, a vector autoregressive model (VAR) in first differences should be used instead. However, if the examined time series are cointegrated, a VAR in first differences will also fail to capture the long-run relationships. The vector error-correction model (VECM) has been introduced to correct a disequilibrium that may shock the whole system. The VECM accounts for both short run and long run relationships, since it is fit to the first differences of the non-stationary variables, and a lagged error-correction term is also included. An alternative approach of estimating causality when time series are non-stationary, is to use a non-parametric information-based measure, such as the symbolic transfer entropy (STE) and its multivariate extension partial symbolic transfer entropy (PSTE). The two approaches, namely the VECM and the STE / PSTE, are evaluated on simulated and real data. The advantage of the PSTE is that it can be applied directly to the non-stationary data, whereas no integration / co-integration test is required in advance. On the other hand, the VECM can discriminate between short run and long run causality.

C1209: **Spatial GARCH(p,q) model**
Presenter: **Hans Arnfinn Karlsen**, University of Bergen, Norway

A space time GARCH(p,q) model is considered. Generalisation of the Garch structure from time series to a pure spatial series appears to be difficult since the Garch model has a conditional specification where the resulting process is non-Gaussian. However, this is possible for a spatial model that includes a time parameter. We presents such a space time Garch model together with conditions ensuring second order stationarity. This result could be seen as an extension of the corresponding well known result for ordinary Garch models.

Monday 16.12.2013 10:55 - 13:00 Parallel Session P – ERCIM

ESI02 Room Senate THEORETICAL PERSPECTIVES IN COMPUTATIONAL STATISTICS Chair: Alastair Young

E098: **Some recent advances in the theory of particle filters**
Presenter: **Ajay Jasra**, National University of Singapore, Singapore
The particle filter is a computational algorithm that is used in many disciplines, including economics, engineering and population genetics. The theory of particle filters is by now reasonably well-understood. Despite this substantial effort, there still is a significant amount of work to be done to fully understand particle filters. We will summarize the recent advances and future work to be done in this challenging area. We will detail some of our work on particles in high-dimensions and adaptive particle filters, as well as that of others on time-uniform behaviour of the particle filter.

E924: **Continuous sequential importance sampling and estimation for stochastic processes**
Presenter: **Gareth Roberts**, University of Warwick, United Kingdom
Co-authors: Paul Fearnhead, Krysztof Latuszynski, Giorgos Sermaidis
Recent work is presented on a sequential importance sampler which provides online unbiased estimation for stochastic processes, and especially irreducible diffusions (that is ones for which the reduction to the unit diffusion coefficient case by the Lamperti transform is not possible). For this family of processes, exact simulation (ie free from discretisation error) using recently developed retrospective simulation techniques is typically not possible. Thus the work significantly extends the class of discretisation error-free Monte Carlo methods available for diffusions. The methods are most useful in the multi-dimensional setting, where many interesting families of models (particularly in finance and population modelling) exhibit the irreducibility property.

E989: **Effects of high-dimensional nuisance parameters on parametric inference**
Presenter: **Alastair Young**, Imperial College London, United Kingdom
Co-authors: Thomas DiCiccio, Todd Kuffner
A major challenge of frequentist parametric inference lies in the appropriate handling of high-dimensional nuisance parameters. Existing asymptotic theory indicates that higher-order accuracy, as well as inferential correctness in relation to conditionality arguments, can be obtained, even in the presence of nuisance parameters, by several routes. These include: analytic distributional approximation based on small-sample asymptotic methods; parametric bootstrap; objective Bayes approaches. Quantification of the effects of nuisance parameter estimation on the inference sheds insight to the relative effectiveness of the different routes in key contexts and to the practical benefits of elimination of nuisance parameters by theoretical means.

ES120 Room B02 MODELS AND GOODNESS-OF-FIT TESTS Chair: Wenceslao Gonzalez-Manteiga

E927: **Goodness of fit tests under informative sampling design**
Presenter: **Abdulhakeem Eideh**, Al-Quds Unicersity, Occupied Territories of Palestine
Sampling designs for surveys are often complex and informative, in the sense that the selection probabilities are correlated with the variables of interest, even when conditioned on explanatory variables. In this case conventional analysis that disregards the informativeness can be seriously biased, since the sample distribution differs from that of the population. Most of the studies in social surveys based on data collected from complex sampling designs. Standard analysis of survey data often fails to account for the complex nature of the sampling design such as the use of unequal selection probabilities, clustering, and post-stratification. The effect of the sample design on the analysis is due to the fact that the models in use typically do not incorporate all the design variables determining the sample selection, either because there may be too many of them or because they are not of substantive interest or in practice it is not possible to anticipate all uses of the design variables. However, if the sampling design is informative in the sense that the outcome variable (variable of interest) is correlated with the design variables not included in the model, even after conditioning on the model covariates, standard estimates of the model parameters can be severely biased, leading possibly to false inference. The analysis of categorical responses survey data, under informative sampling, is addressed. Also we discuss goodness of fit tests based on likelihood ratio and Kullbak-Leibler Information measure.

E963: **An exact Kolmogorov-Smirnov test for the binomial distribution with unknown probability of success**
Presenter: **Arnab Hazra**, Indian Statistical Institute, India
An exact Kolmogorov-Smirnov goodness-of-fit test for the binomial distribution with an unknown probability is developed. This test is conditional, with the test statistic being the maximum absolute difference between the empirical distribution function and its conditional expectation given the sample total. Exact critical values are obtained using a modification of an algorithm previously used in case of Poisson distribution. We explore properties of the test, and we illustrate it with three examples. We also include some simulations in order to check the power of the procedures. The new test seems to be the first exact binomial goodness-of-fit test for which critical values are available without simulation or exhaustive enumeration.

E198: **Spatial matern fields generated by non-Gaussian noise**
Presenter: **Jonas Wallin**, Lund University, Sweden
Co-authors: David Bolin
The aim is to study non-Gaussian extensions of a recently discovered link between certain Gaussian random fields, expressed as solutions to stochastic partial differential equations, and Gaussian Markov random fields. We show how to construct efficient representations of non-Gaussian random fields generated by generalized asymmetric Laplace noise and normal inverse Gaussian noise, and discuss parameter estimation and spatial prediction for these models. Finally, we look at an application to precipitation data from the US.

E1039: **Slepian model for moving averages driven by a non-Gaussian noise**
Presenter: **Krzysztof Podgorski**, Lund University, Sweden
Co-authors: Igor Rychlik, Jonas Wallin
In analysis of extreme behavior of a time series or a stochastic process the Rice formula is often used to obtain the distribution of the process at instants of level crossings (biased sampling distribution). A Slepian model for a general stationary process is defined as any explicit and convenient process with the distribution coinciding with the biased sampling distribution. The original Slepian model has been developed for a stationary Gaussian process and due to its simplicity is very helpful in analyzing the behavior of the process at extreme levels. Here, a Slepian model is derived describing the distributional form of a moving average driven by a non-Gaussian noise as observed at level crossings. We take a generalized Laplace noise although the methodology can be applied to more general processes. A computational method of sampling from the corresponding biased sampling distribution of the underlying gamma process is obtained. This is used for efficient simulation of the behavior of a non-Gaussian process at the high level crossing. Additionally, we present the asymptotic behavior of the process at high level crossings that is fundamentally different than that in the Gaussian case.

E1197: **Robust correlation coefficient goodness-of-fit test for the extreme value distribution**
Presenter: **Aldo Corbellini**, Faculty of Economics - University of Parma, Italy
Co-authors: Abbas Mahdavi
A simple robust method is provided to test the goodness of fit for the extreme value distribution (Type I family) by using the new diagnostic tool called the Forward Search method. The Forward Search is a powerful general method that provides diagnostic plots for finding outliers and discovering their underlying effects on models fitted to the data and for assessing the adequacy of the model. The Forward Search algorithm has been previously developed for regression modeling and multivariate analysis frameworks. One of the powerful goodness-of-fit tests is represented by the correlation coefficient test, but this test suffers from the presence of outliers. We introduce the Forward Search version of this test that is not affected by the outliers. Also by using the transformation study, an application to the two-parameter Weibull distribution is investigated. The performance and the ability of this procedure to capture the structure of data are illustrated by some simulation studies

E1157: **On the effect of the correlation on the variable selection using random forest**
Presenter: **Baptiste Gregorutti**, Universite Pierre et Marie Curie, France
Co-authors: Bertrand Michel, Philippe Saint-Pierre
In many situations, the selection of the most relevant variables is necessary to explain an outcome of interest. This task is often tricky particularly when correlations between variables are involved. The main goal is to eliminate irrelevant variables and to select a small size model with a good prediction error. We focus on variable selection using the random forest algorithm in presence of correlated predictors. First of all, we provide a theoretical study of the permutation importance measure in the case of an additive regression model. As already observed in simulation studies, we show that permutation importance strongly depends on the correlation between the variables. The Recursive Feature Elimination (RFE) algorithm is then considered for variable selection. This algorithm recursively eliminates the variables using permutation importance measure as a ranking criterion. A large simulation study based on several designs highlights (i) the effect of correlation on permutation importance measure, (ii) the efficiency of the RFE algorithm for selecting a small number of variables together with a good prediction error. Finally, this selection algorithm is applied to the context of aviation safety where the flight data recorders are analysed for the prediction of a dangerous event.

ES01 Room B18 HIGH DIMENSIONAL DATA ANALYSIS: PAST, PRESENT AND FUTURE Chair: S. Ejaz Ahmed

E088: **The KLAN for complex grouped variable selection**
Presenter: **Xiaoli Gao**, Oakland University, United States
Existing grouped variable selection methods rely heavily on the prior group information, thus may not be reliable if an incorrect group assignment is used. We propose a family of novel shrinkage variable selection operators by controlling the k-th largest norm (KLAN). The proposed KLAN method exhibits some flexible group-wise variable selection naturally even though no correct prior group information is available. We also construct a group KLAN shrinkage operator using a composite of KLAN constraints. Neither ignoring nor relying completely on prior group information, the group KLAN method has flexibility of controlling the within group strength and therefore can reduce the damage caused by incorrect group information. Finally, we investigate an unbiased estimator of the degrees of freedom of (group) KLAN estimates in the framework of Steins unbiased risk estimation (SURE) . Some small sample simulation studies are performed to demonstrate the advantage of both KLAN and group KLAN as compared to the LASSO and group LASSO, respectively.

E175: **Invisible fence methods for the identification of differentially expressed gene sets**
Presenter: **J. Sunil Rao**, University of Miami, United States
Co-authors: Jiming Jiang, Thuan Nguyen
The fence method involves a procedure to isolate a subgroup of correct models (of which the optimal model is one) by constructing a barrier to eliminate incorrect models. Once constructed, the optimal model can be selected from within the fence according to various criteria that can be made flexible. We extend the fence to situations where a true model may not exist or be among the candidate models leading to a new procedure called the invisible fence (IF). A fast algorithm using a subtractive measure of fit is developed allowing scaling to particular high dimensional data problems. The main application considered here is microarray gene set analysis. A special feature of this data is that the number of genes dominates the sample size. We demonstrate consistency properties of the IF and compare it to other methods of gene set analysis (GSA) which can have inconsistencies under certain situations. Numerical and real data studies demonstrate the promising performance of the IF method.

E383: **Efficient estimation in high dimensional spatial regression models**
Presenter: **Abdulkadir Hussein**, University of Windsor, Canada
Co-authors: Marwan Al-Momani, S. Ejaz Ahmed
Some spatial regression models are considered and an array of shrinkage and absolute penalty estimators for the regression coefficients are developed. The estimators are compared analytically and by means of Monte Carlo simulations. The usefulness of the proposed estimation methods is illustrated by using data sets on crime distribution and housing prices.

E580: **Complex Bayesian space-time models: Some tools for their efficient development**
Presenter: **Pilar Munoz**, Universitat Politecnica de Catalunya, Spain
Co-authors: Alberto Lopez
Space time models are being increasingly used in order to analyze stochastic processes related to the environment, econometrics, epidemiology, health sciences, and a long etcetera. The estimation and prediction of these processes is non-trivial and the data set needed for estimating and predicting them correctly, in general, is very large, requiring a high computational cost. Among the statistical procedures that have contributed most to estimating these kinds of problems are the hierarchical Bayesian space time models (HBSTM). These procedures consist of starting with a complex structure and make it hierarchically flexible up to a basic one within a Bayesian perspective. The objective is to set up some norms or guides for helping and solving the most common difficulties of fitting that kind of data. Specifically, we address: the correct theoretical model definition; the explanation of different spatial structures and how they affect the estimation algorithm; and, most importantly, provide alternatives for selecting the correct initial values for the hyperpriors parameters. This model has been programmed in R code and applied to the estimation of temperature data sets in the Strait of Gibraltar.

E394: **Shrinkage estimation for high dimensional data analysis**
Presenter: **Ejaz Ahmed**, Brock, Canada
Co-authors: Xiaoli Gao
In high-dimensional data settings where $p > n$, many penalized regularization approaches were studied for simultaneous variable selection and estimation. However, with the existence of covariates with weak effect, many existing variable selection methods, including Lasso and its generations, cannot distinguish covariates with weak and no contribution. The prediction based on a submodel of selected covariates only may not be efficient. A high-dimensional shrinkage estimation strategy is proposed to improve the prediction performance of a submodel. Such a high-dimensional shrinkage estimator is constructed by shrinking a weighted ridge estimator in the direction of a predefined candidate submodel. Under an asymptotic distributional quadratic risk criterion, its prediction performance is explored analytically. It is shown that the proposed high- dimensional

shrinkage estimator performs better than the weighted ridge estimator. Further, it improves the prediction performance of any candidate submodel generated from most existing Lasso-type variable selection methods. The relative performance of the proposed high-dimensional shrinkage strategy is demonstrated by both simulation studies and real data analysis.

ES09 Room B29 STATISTICAL METHODS FOR NON-PRECISE DATA II **Chair: Angela Blanco-Fernandez**

E565: **Symmetry of the fuzzy-valued mean and median of symmetric random fuzzy numbers**
Presenter: **Beatriz Sinova**, University of Oviedo, Spain
Co-authors: Maria Rosa Casals, Maria Angeles Gil

Among the valuable properties of the mean and the median of a random variable as measures quantifying its central tendency, one should highlight that associated with their behavior in case of dealing with symmetric distributions. In such a case, the mean equals the symmetry point and the median either equals the symmetry point (if it is unique) or it is any of the values in an interval which is symmetry w.r.t. such a point. An analysis is to be performed on the behavior of the usual extensions of the mean and the median for random fuzzy numbers (i.e., fuzzy number-valued random elements) when symmetric random fuzzy numbers are considered. In this respect, one can conclude that both extended measures always correspond to fuzzy numbers which are symmetric with respect to the symmetry point. An introductory approach will be also given to the symmetry of a random fuzzy number w.r.t. a symmetric 'fuzzy point'. The conclusions will be illustrated by means of several examples.

E590: **Analyzing the effect of the scientific background in "Likertizing" fuzzy rating responses to a questionnaire**
Presenter: **Sara de la Rosa de Saa**, Universidad de Oviedo, Spain
Co-authors: Peter Filzmoser, Maria Asuncion Lubiano

The advantages of using continuous rating scales over the well-known Likert-type scales in opinion surveys have been discussed in recent papers. Continuous scales based on fuzzy numbers, like the fuzzy rating scale, have already been used. Two real-life questionnaires are analyzed, and their replies are simultaneously given in two scales, namely, Likert and fuzzy rating ones. For both questionnaires, each free fuzzy response is identified with the Likert category (more concretely, with its usual integer encoding) with the smallest distance, and this association is compared with the real one for the two surveys. The results show the highest percentage of coincidences (around 86%) for the questionnaire in which the respondents had a higher background and expertise (most of them were mathematicians, engineers and computer scientists). By contrast, lower coincidences (around 78%) are obtained for the other questionnaire whose respondents came from a more diverse background.

E630: **Generalized probabilistic inference, membership functions and measures of fuzzy sets**
Presenter: **Giulianella Coletti**, University of Perugia, Italy
Co-authors: Barbara Vantaggi

Consistency with a reference framework of a likelihood function and an uncertainty measure (probability, completely monotone measure , maxitive function) is studied. These measures arise, for example, when the likelihood function and the prior are defined on different spaces, so the prior needs to be extended on the space where the likelihood function is defined. When these not-additive measures have the role of "a prior" the aim is to update them through a generalized Bayesian inferential procedure. An interpretation of the membership of fuzzy sets inside these frameworks is provided, through a coherent conditional non-additive measure μ regarded as function of the conditioning event. The semantic seems to be similar to the probabilistic one: the membership function consists of the degree that You claim that a variable X has the property φ under the hypothesis that X assumes the value x. From a syntactical point of view many differences and many common features can occur. It is established which properties of fuzzy sets can be preserved in a not-additive ambit. Some results concerning the measure of uncertainty μ of a fuzzy event (which in this interpretation coincides with a Boolean event of the kind "You claim that X is φ") are given.

E632: **Coherence for statistical matching with misclassification**
Presenter: **Barbara Vantaggi**, University La Sapienza, Italy
Co-authors: Marco Di Zio

In several applications there is the need to combine different data sources. A specific case is the so-called statistical matching that aims at combining information available in distinct sample surveys (distinct observations) referred to the same target population, characterised by having a set of common variables. The focus is on the case when two data sets are integrated, the common variables are categorical and in one of the data sources the common variables are affected by misclassification. The classical approach consists in assuming an error model for the misclassified variables, and this allows us to make a point-wise inference. However, assumptions on the error mechanism are difficult to test with the data at hand because of the lack of common observations in the two data sources. The use of coherence allows us to make inference by avoiding specific assumptions concerning the misclassification mechanism, the related inference will take into account all the sets of compatible probabilities with respect to a broad class of misclassification models. In this setting, the inference will not provide anymore a point estimation but a class of probability distributions. An example showing advantages and drawbacks of the proposed method is given.

E1151: **Fuzzy indicators: An application for measuring sustainable development**
Presenter: **Shohreh Mirzaei Yeganeh**, TU wien, Austria
Co-authors: Reinhard Viertl

Over the last decade, decision-makers have had an increasing attitude to set quantitative targets for meeting policy goals. Sustainable development (SD) is one of the recently introduced concepts with focus on development, ensuring both the well-being of those currently living and the potential for the well-being of future generations. A variety of indicators have been introduced by national and international organizations to measure trends in economic, social and environmental dimensions of SD. Measurements of SD-related variables are often subject to considerable uncertainty. Besides statistical variability, such data are also fuzzy. For example, the amount of CO_2 emission in a given period can not be measured as a precise number. This fuzziness is not statistical in nature, since statistical models are suitable to describe variability, but for a single measurement there is no variability. The uncertainty of a single measurement can be described by a so-called fuzzy number. To determine a realistic SD indicator, the underlying fuzziness of the data must be quantitatively described. This is possible with the help of fuzzy numbers (respectively, fuzzy vectors). Based on this, it is possible to obtain more realistic values for sustainable development indicators.

ES20 Room B36 SEMIPARAMETRIC ESTIMATION METHODS FOR DURATION MODELS WITH CENSORED DATA Chair: M.Carmen Pardo

E102: **Semi-parametric estimation of imperfect preventive maintenance effects and intrinsic wear-out.**
Presenter: **Laurent Doyen**, University Grenoble Alpes, France

A major issue for industrial systems is the joint management of ageing and maintenance. An efficient maintenance and a controlled ageing allow the extension of the operating life on equipment. A system subject to corrective maintenance (CM) and preventive maintenance (PM) actions is considered. CM is carried out after failure. It is done at unpredictable random times. Its aim is to quickly restore the system in working order. Then, CM effects are assumed to be As Bad As Old (ABAO, i.e. leave the system in the state as it was before maintenance). PM are carried out when the system is operating and intend to slow down the wear process and reduce the frequency of occurrence of system failures. They are supposed to be done at predetermined times. Their effects are assumed to follow a Brown-Proschan (BP) model: each PM is As Good As New (AGAN, i.e. renews the system) with probability p and ABAO with probability $(1-p)$. In this context a semi-parametric estimation method is proposed: non-parametric estimation of the first time to failure distribution and parametric estimation of the PM effect, p. The originality is to consider that BP effects (ABAO or AGAN) are unknown.

E425: **Cox regression for data sets containing a part of current status data**
Presenter: **Laurent Bordes**, Universite de Pau et des Pays de l'Adour, Mexico
Co-authors: Martha Lorena Avendano, Ma. Carmen Pardo, Valentin Patilea

Let T be a lifetime or a failure time and Z a vector of covariates possibly depending on time. To account the effect of covariates on the duration T we assume a proportional hazards model. This model has been largely studied in presence of incomplete data. We consider the case where the covariates Z are observed and where the duration T is either observed or it is left or right censored (current status). A latent variable model is introduced, that allows us to derive a Breslow type estimator of the cumulative baseline hazard rate function for a fixed value of the regression parameter, and then to write a profile likelihood function, which depends on the regression parameter only. Finally, estimators of all the unknown parameters are derived. Based on general asymptotic results for profile likelihood estimators, and using empirical processes theory, several results of consistency and weak convergence for the estimators are obtained. The semiparametric efficiency of the regression parameter estimator is also derived. Finite distance properties of the estimators are investigated through a Monte Carlo study and an application to a real data set.

E437: **Semiparametric inference for ARA models**
Presenter: **Eric Beutner**, Maastricht University, Netherlands
Co-authors: Laurent Bordes

A class of semiparametric virtual age models is considered, called arithmetic reduction of age models of first type. It is shown that for this kind of semiparametric model the often employed profile likelihood approach does not work. An estimation method based on the minimum distance principle is then proposed. The estimators are studied through a Monte Carlo study.

E589: **Semiparametric inference for lifetime data with competing risks, additive risks and different mechanisms of missingness**
Presenter: **Jean-Yves Dauxois**, INSA-IMT Toulouse, France
Co-authors: Laurent Bordes, Pierre Joly

The aim is to consider a semiparametric model for lifetime data with competing risks and missing causes of death. It is assumed an additive hazards model on each cause-specific hazard rate function and that a random right censoring occurs. The goal is to estimate the regression parameters as well as functional parameters like the baseline cumulative hazard rate functions, the cause-specific cumulative hazard rate functions or the cumulative incidence functions. Preliminary estimators of the unknown (Euclidean and functional) parameters are first introduced when cause of death indicators are missing completely at random (MCAR). They are obtained using the observations with known cause of failure. The advantage of considering such a simple MCAR model appears when it is shown that, in this case, the information given by the observed lifetimes with unknown failure cause can be used to improve the previous estimates in order to reach an asymptotic optimality criterion. This is the main purpose of the work. However, since it is often more realistic to consider a missing at random (MAR) mechanism, estimators of the regression and functional parameters under this model are also derived. The large sample behaviors of the estimators are obtained thanks to martingale and empirical process techniques. A simulation study is also provided in order to compare the behavior of the three types of estimators under the different mechanisms of missingness. It is shown that the improved estimators under MCAR assumption are rather robust when only the MAR assumption is fulfilled. Finally, two illustrations on real datasets are also given, one in Survival Analysis and the other in Reliability.

E972: **Kaplan-Meier integrals for bivariate survival data**
Presenter: **Philippe Saint-Pierre**, University Pierre et Marie Curie, France
Co-authors: Olivier Lopez

The study of dependent durations arises in many fields. For example, in epidemiology, one can be interested in the lifetimes of twins or in the time before healing after using two kinds of drugs on sick eyes. Such bivariate survival data are often incomplete due to censoring. Some approaches have been proposed in the literature to deal with such data. Among the nonparametric estimators proposed in the literature, most of them estimate the survival function but are not adapted to estimate the failure time distribution as they often do not define proper distributions. In the presence of bivariate right-censoring, we introduce a class of nonparametric estimators based on a copula approach. The estimator allows us to study integrals of the distribution as it is designed to estimate the joint distribution of failure times. The estimator can be seen as a generalization of the empirical distribution to censored data. Asymptotic i.i.d. representation and asymptotic normality for these integrals are obtained. This estimator is used to estimate several quantities as the survival function, dependence measures as the Kendall's coefficient or the regression function when the response and the covariate are both censored. Finally, these methods are applied to a database on twin lifetimes.

ES94 Room B33 OUTLIERS AND STRUCTURAL CHANGES IN TIME SERIES Chair: Roland Fried

E418: **A multiple filter test for change point detection in renewal processes with varying variance**
Presenter: **Michael Messer**, Goethe University Frankfurt, Germany
Co-authors: Marietta Kirchner, Julia Schiemann, Jochen Roeper, Ralph Neininger, Gaby Schneider

Non-stationarity of the event rate is a persistent problem in modeling time series of events, such as neuronal spike trains. Motivated by a variety of patterns in neurophysiological spike train recordings, a general class of renewal processes is defined. This class is used to test the null hypothesis of stationary rate versus a wide alternative of renewal processes with finitely many rate changes (change points). The test extends ideas from the filtered derivative approach by using multiple moving windows simultaneously. To adjust the rejection threshold of the test a Gaussian process is used, which emerges as the limit of a cadlag filtered derivative process. The benefits of the multiple filter test are analyzed and the increase in test power against a single window test is studied. A multiple filter algorithm is also developed, which can be used when the null hypothesis is rejected in order to estimate the number and location of change points. In a sample data set of spike trains recorded from dopamine midbrain neurons in anesthetized mice in vivo, the detected change points agreed closely with visual inspection, and in over 70% of all spike trains which were classified as rate non-stationary, different change points were detected by different window sizes.

E436: **Robust online-detection of structural changes in uni- and multivariate time series**
Presenter: **Matthias Borowski**, University of Muenster, Germany
Co-authors: Roland Fried
The online- (also real-time or sequential) surveillance of data stream time series takes place in many fields. These time series are often not stationary but show enduring and suddenly changing trends and jumps. Due to the high frequency of measurement the data are often corrupted by a changing amount of noise and outliers. Furthermore, single observations as well as long stretches of data can be missing due to technical problems. Moreover, the coherences between the univariate components of multivariate streams may change over time. The online-detection of structural changes in such complex time series is a challenging task calling for fast and robust (multivariate) methods that are based on weak assumptions and which are feasible in practice. A filtering procedure is presented that can be used for the online-detection of trend changes and jumps. Based on this univariate method a procedure for the online-detection of incoherences between two univariate time series is developed. Both procedures assume an unknown locally linear signal carrying the relevant information and are therefore based on robust regression in moving windows.

E491: **Modelling asset prices within an HMM: Investment strategies including assets and bonds**
Presenter: **Stefanie Grimm**, Fraunhofer Institute of Industrial Mathematics ITWM, Germany
Co-authors: Christina Erlwein-Sayer, Peter Ruckdeschel, Joern Sass
An HMM framework for stock returns is considered to develop investment strategies for portfolios consisting of two stocks and a bond. In this framework, drift and volatility of the returns are driven by a discrete time hidden Markov chain and can therefore switch between different market regimes. Filtering techniques are used to obtain optimal parameter estimates. The forecasted asset distribution supports an investor to decide how to switch his investment between assets at given times. A mixed strategy is also analysed, here the investor is allowed to diversify his investment between all assets. To examine the results of these strategies data sets of DAX stock returns for the period of 2006-2012 are considered. Using historical as well as simulated data, the performances of the investment strategies are compared with those of purely investing in one of the assets. The findings show that the model can improve investment decisions in switching market situations. Finally, the algorithm of the parameter estimation is robustified to account for additive outliers in the observation process.

E496: **Robust filtering and extreme value statistics for hydrological data**
Presenter: **Peter Ruckdeschel**, Fraunhofer ITWM, Germany
Co-authors: Bernhard Spangl
River discharge data are extreme events with strong seasonal and regional variations. Modern research pools data collected at different times and locations to improve risk prediction at a certain point and time. Besides extreme value statistics, this calls for filtering techniques to remove seasonalities, and regional effects. Among the extreme events of such data, it is not clear whether they are recurrent, so non-robust approaches attributing overly high influence to single extreme events could impair future risk predictions. This motivates the use of robust variants of filtering and extreme value procedures. The aim is to analyze data comprising daily average discharge time series of rivers from various sites in Austria collected over the last 35 years. It is limited to considering seasonal effects. For filtering, a linear, time-invariant, time-discrete state space model is fitted to the data by means of a robust EM algorithm. By filtering a detrended and deseasonalized series is extracted, to which a Generalized Pareto distribution is robustly fitted.

E710: **A systematic comparison of statistical process monitoring methods for high-dimensional, time-dependent processes**
Presenter: **Eric Schmitt**, KU Leuven, Belgium
Co-authors: Mia Hubert, Bart de Ketelaere, Marco Reis, Tiago Rato
Most multivariate statistical process monitoring (SPM) methods rely on some form of Principal Component Analysis (PCA) methodology in order to deal with features of high-dimensional and time-dependent processes. These methodologies can be divided into two broad categories: non-adaptive and adaptive models. Non-adaptive models include the classic PCA approach and Dynamic PCA for data with auto-correlation, while Recursive PCA and Moving Window PCA, derived for non-stationary data, are adaptive. To our knowledge, there is no comprehensive investigation indicating when each method is more appropriate or examining their relative performance. Therefore, a comparison of the performance of these methods on different process scenarios is performed, and guidelines are outlined on the selection of the most appropriate monitoring strategy for each scenario. The process scenarios are drawn from a selection of simulated and real data cases with high-dimensionality and time dependence properties. These characteristics make many of the commonly used methods, which assume stationarity and nonautocorrelated variables, unsuitable for use. Guidelines for the selection of parameter values for PCA methods, such as the number of lags for DPCA or the forgetting factor for RPCA are also discussed. Additionally, approaches for simulating data with desirable properties for testing the methods are covered.

ES42 Room B34 ADVANCES IN COMPOSITIONAL DATA ANALYSIS AND RELATED METHODS II Chair: J. Antoni Martin-Fernandez

E218: **Statistical analysis of compositional tables in coordinates**
Presenter: **Karel Hron**, Palacky University, Czech Republic
Co-authors: Kamila Facevicova, Ondrej Vencalek
Compositional tables as a special case of compositional data represent a continuous counterpart to the well-known contingency tables, which carry relative information about a relationship between two (row and column) factors. The nature of compositional tables requires a specific geometrical treatment, represented by the Aitchison geometry on the simplex. Its properties allow a decomposition of the original table into its independent and interactive parts and consequently investigation of the relationship between factors. The corresponding statistical analysis should be performed in orthonormal coordinates (with respect to the Aitchison geometry). The resulting isometric logratio (ilr) coordinates provide representation of compositional data in the real space, where all standard statistical methods can be applied. The aim is to search for such ilr coordinates that are easy to handle and well-interpretable. We introduce a new formula that assigns such coordinates through decomposition of the interaction table into smaller tables; each coordinate then corresponds to odds-ratio in these tables. Theoretical results are applied to a real-world example.

E507: **Simplicial regression: a useful tool in ecological risk assessment**
Presenter: **Gianna Monti**, University of Milano-Bicocca, Italy
Co-authors: Sonia Migliorati, Karel Hron, Klara Hruzova, Eva Fiserova
The assessment of different toxic compounds on living organisms is usually based on fitting non-linear regression models. Such models are able to describe the variation in the effect on an exposed organism caused by changing the dose or concentration level of a particular compound. In order to control the ecological risk assessment, the quantitative relationships between the log-concentration and the magnitude of the response, i.e. probability of death or probability of inhibition in living organisms, is therefore of interest. The concentration-response curve is a result of the best-fit selected curve among different non-linear regression models, such as logistic, probit or Weibull models. Since the response contains percentage of inhibited organisms exposed to different levels of concentrations, a reformulation of the basic task is proposed by introducing simplicial regression of the observed proportion of inhibitions to the log-concentration values using the log-ratio methodology. This approach, leading to an easy interpretable regression model, enables us to preserve the relative scale of the response as well as to perform consequent statistical inference like confidence and prediction bounds. An analytical comparison of the standard and simplicial approaches is provided and also some real-world data examples are reported.

E925: **On compositional association**
Presenter: **Juan Jose Egozcue** , Technical University of Catalonia, Spain
Co-authors: Vera Pawlowsky-Glahn, David Lovell
Proportionality between parts of a composition across a sample is a meaningful measure of association: when two parts are perfectly proportional, the variance of their log-ratio is zero. The variance of the log-ratio of two parts has been adopted as a measure of association in compositional data analysis, substituting Pearson correlation, which is known to produce spurious results. However, when two parts are not perfectly proportional, it is not clear how to interpret the variance of their log-ratio. Normalization of log-ratios and testing for proportionality become thus crucial in an exploratory analysis. A normalization of the variation matrix, made of all variances of log-ratios between couples of parts, is proposed. Based on regression on the principal axis, an asymptotic F- test of proportionality is developed. A second test checks whether the lack of association between two parts should or should not involve other parts in the analysed composition. It relays on a multiple regression of log-ratios and takes the form of an F-test. The combination of the variation matrix normalization and the two tests are proposed as a standard presentation of sample compositional association.

E476: **Compositional data analysis of the resource efficiency in manufacturing**
Presenter: **Valentin Todorov**, UNIDO, Austria
Co-authors: Klara Hruzova, Karel Hron, Peter Filzmoser
The statistical analysis of compositional data cannot be performed for the original observations until they are expressed in orthonormal coordinates (with respect to the Aitchison geometry on the simplex, the sample space of compositions). The specific nature of compositional data become particularly visible, when regression analysis is performed. In case, when finding a regression relation within compositions is of interest, it is necessary to form special coordinate systems in order to capture all the relevant information. Moreover, in the presence of outlying observations in compositional data the orthogonal regression (that is able to handle the regression problem statistically) should be replaced by its robust counterpart. The theoretical findings are applied to a study of resource efficiency in manufacturing, carried out by UNIDO. The UNIDO INDSTAT database contains data on gross output and value added for the manufacturing sectors. Additional data on input costs (raw materials, energy and services) was collected for about 30 countries. The output for each particular sector equals to the sum of value added and all other input costs. Thus the data set to be considered has a compositional structure. The logratio approach applied to this compositional data set reveals the dependence of the resource efficiency on the technology characteristics of the sectors. Further, robust orthogonal regression is used to study the relation of value added to the other components of the input expenses as well as the structure of the input.

E1127: **Sparse principal balances for high-dimensional compositional data**
Presenter: **Mehmet Can Mert**, Vienna University of Technology, Austria
Co-authors: Peter Filzmoser, Karel Hron
The analysis of compositional data has to take care of the fact that the relevant information is contained only in the ratios between the measured variables, and not in the reported values. The focus is on high-dimensional compositional data, as they appear in chemometrics (e.g. mass spectral data), proteomics or genomics. The goal is to perform a dimension reduction of such data, where the new directions should allow for interpretability. An approach named principal balances turned out to be successful, but is limited to low dimension. The concept of sparse principal component analysis is used here to construct principal directions, so-called sparse principal balances, that are sparse (contain many zeros), build an orthonormal basis in the sample space of the compositional data, are efficient for dimension reduction, and are applicable to high-dimensional data.

ES58 Room B30 NEW METHODS FOR GROWTH PROCESSES Chair: Christine Mueller

E220: **Analysis of multivariate growth curves with cubic smoothing splines**
Presenter: **Tapio Nummi**, University of Tampere, Finland
Co-authors: Nicholas Mesue
A novel method for estimation and testing of growth curves when the analysis is based on spline functions is presented. The foundation of the new method is on the use of a spline approximation of natural cubic smoothing splines. For the approximated spline model an exact F-test is developed. This method also applies under a certain type of correlation structures that are especially important in the analysis of repeated measures and growth data. We tested the method on real data and also investigated it by simulation experiments. The method proved to be a very powerful modeling and testing tool especially in situations, where the growth curve may not be easy to approximate using simple parametric models. The method is also easily extended to more general multiple response situations.

E388: **Prediction of growth curves: A Bayesian estimation approach**
Presenter: **Simone Hermann**, TU Dortmund University, Germany
In constructional engineering, material fatigue is an important topic of research because it is important to determine the lifetime of a product. In many cases, experiments are very expensive and therefore seldom performed. That is a problem in survival analysis. A group of engineers at the TU Dortmund conducted an experiment, in which they set several prestressed concrete beams under cyclic load, starting with initial cracks. The observed crack widths exhibit irregular jumps with increasing frequency which influence the growth process substantially. The aim is to describe the process in a way that makes it possible to predict the development of the curve. An ordinary diffusion process will be expanded by a jump process with increasing intensity rate. The model will be estimated with Bayesian methods and a prediction procedure will be proposed.

E453: **Depth based estimators with application to crack growth**
Presenter: **Christoph Kustosz**, TU Dortmund University, Germany
Co-authors: Christine Muller
Based on data from experiments conducted on prestressed concrete, different growth models to describe crack growth are introduced. Due to specific properties of prestressed concrete, non-standard errors are considered by the choice of an appropriate robust estimation method. Based on simplicial depth, estimators and test statistics for which limiting distributions exist are proposed and the corresponding confidence intervals and parameter tests are derived. As a result general crack growth models which treat stochastic errors in a non-parametric way can be constructed. By a change point identification method resting upon these estimators the data can be analysed with respect to phase changes. Additionally, a method to construct prediction intervals for crack growth is proposed, which is illustrated by real data examples.

E857: **Wald tests under matrix normalisation**
Presenter: **Tassos Magdalinos**, University of Southampton, United Kingdom
It is well known that systems of regression equations exhibiting different persistence degrees along each equation do not necessarily conform to standard asymptotic theory of estimation and testing. The key difference with the standard asymptotic framework of inference is that sample moment matrices require matrix-valued normalisations, a complication that may result to a reduction in the asymptotic rank of sample moment estimators and the associated test statistics. In hypothesis testing, an additional complication arises from the interaction between the matrix-valued normalisation and the matrix of restrictions imposed by the null hypothesis, which may lead to further asymptotic degeneracy and non-standard limit distributions for Wald type test statistics. Sufficient conditions are provided that guarantee standard chi-squared inference for Wald tests in

this general multivariate modelling framework. Applications include regression models with deterministic components and cointegrated systems of mildly integrated time series with roots that induce potentially different persistence rates.

E501: **Prediction intervals of crack growth**
Presenter: **Christine Mueller**, TU University Dortmund, Germany
The prediction of the growth of cracks is an important topic in engineering sciences. In particular, prediction intervals are of interest to express the precision of the prediction. Several approaches for prediction intervals for growth models are discussed. Some of these approaches use only the observations of the growth process for which the prediction should be done. Here especially approaches of nonlinear models and SDE models based on the Paris-Erdogan equation are considered. Additionally, approaches are studied which use also the information of other growth processes. Here a special mixed linear model is considered.

ES61 Room B35 NONPARAMETRIC FUNCTIONAL DATA ANALYSIS Chair: Alicia Nieto-Reyes

E035: **Exploratory nonparametric functional data analysis using the spatial depth approach**
Presenter: **Robert Serfling**, University of Texas at Dallas, United States
Co-authors: Uditha Wijesuriya
The spatial depth approach toward multivariate quantiles has been very successful for its tractability, computational ease, and convenient asymptotics. Here its extension to the setting of functional data is treated. For a data set of real-valued curves, methods are described for useful display of the sample median curve, 50% and other central regions of sample curves, and sample 'outlier' curves. Computations may be carried out in the Hilbert space of curves or in a corresponding Euclidean space obtained by discretization. Population quantile curves for any selected rank can be estimated by straightforward sample versions. Very importantly, the spatial approach yields asymptotic normality of the sample quantile curves, and this is applied to develop confidence bands for population quantile curves. The techniques are illustrated with several data sets, and comparisons are made with some competing methods, with emphasis on performance in outlier detection.

E350: **Boosting functional regression models**
Presenter: **Sonja Greven**, Ludwig-Maximilians-University Munich, Germany
Co-authors: Sarah Brockhaus, Fabian Scheipl, Torsten Hothorn
Boosting allows the estimation of a broad class of functional regression models with functional response and multiple functional and scalar covariates. Using different base-learners in the boosting algorithm, linear effects of functional covariates can be combined with linear and smooth effects of scalar covariates for suitably flexible model specification. Additionally, correlated data from longitudinal, hierarchical or spatial designs can be accommodated by including group-, cluster- or location-specific effects. Furthermore it is possible to minimize diverse loss functions yielding e.g. robust models like quantile regression for functional data. Another benefit of boosting is its capability for variable selection. To estimate the models the fast and stable R-package mboost is extended to meet the requirements of functional variables.

E510: **Functional regression with points of impact**
Presenter: **Alois Kneip**, University of Bonn, Germany
Co-authors: Dominik Poss, Pascal Sarda
Functional linear regression is considered, where scalar responses $Y_1,\ldots,Y_n$ are modeled in dependence of i.i.d. random functions $X_1,\ldots,X_n$. A generalization of the classical functional linear regression model is studied. It is assumed that there exists an unknown number of "points of impact", i.e. discrete observation times where the corresponding functional values possess significant influences on the response variable. In addition to estimating a functional slope parameter, the problem then is to determine number and locations of points of impact as well as corresponding regression coefficients. Identifiability of the generalized model is considered in detail. It is shown that points of impact are identifiable if the underlying process generating $X_1,\ldots,X_n$ possesses "specific local variation". Examples are well-known processes like the Brownian motion, fractional Brownian motion, or the Ornstein-Uhlenbeck process. The focus then lies upon methods for estimating number and locations of points of impact. It is shown that this number can be estimated consistently. Furthermore, rates of convergence for location estimates and regression coefficients are derived. Finally, some simulation results are presented.

E750: **Functional estimation in continuous time**
Presenter: **Delphine Blanke**, Universite d Avignon et des Pays de Vaucluse, France
Various results are presented concerning nonparametric functional kernel estimators or predictors for (multidimensional) continuous-time processes, in the case of density and regression estimation as well as prediction. Rates of convergence depend on the dimension d but also on regularity of the sample paths. We derive results concerning adaptive estimators and also some illustrations of the behaviour of kernel estimators in the sampled case.

E1156: **Nonparametric checks of the effect of functional covariates**
Presenter: **Valentin Patilea**, CREST-ENSAI, France
Given a response variable taking values in a Hilbert space and covariates that could be of finite or infinite dimension, the problem of testing the effect of the functional covariates on the response variable is addressed. This problem occurs in many situations as for instance significance testing for functional regressors in nonparametric regression with hybrid covariates and scalar or functional responses, testing the effect of a functional covariate on the law of a scalar response, testing the goodness-of-fit of regression models for functional data. We propose a new test based on univariate kernel smoothing. The test statistic is asymptotically standard normal under the null hypothesis provided the smoothing parameter tends to zero at a suitable rate. The one-sided test is consistent against any fixed alternative and detects local alternatives approaching the null hypothesis at a suitable rate. In particular, we show that neither the dimension of the outcome nor the dimension of the functional covariates influences the theoretical power of the test against such local alternatives.

ES75 Room G15 METHODS FOR HANDLING MISSING DATA Chair: Shaun Seaman

E110: **Using multiple imputation to improve analysis of case-control studies nested within cohorts**
Presenter: **Ruth Keogh**, London School of Hygiene and Tropical Medicine, United Kingdom
Co-authors: Ian White
In many large prospective cohorts, expensive exposure measurements cannot be obtained for all individuals. Exposure-disease association studies are therefore often based on sub-studies within the cohort, in which complete information is obtained only for sampled individuals. Two main ways of sampling within cohorts are the nested case-control and case-cohort designs. The traditional analysis of such studies is based only on sampled individuals. However, in the full cohort there may be a large amount of information on cheaply available covariates and possibly a surrogate of the main exposure, which typically goes unused. We can view the sub-study plus the remainder of the cohort as a full cohort study with missing data. Hence, we have proposed using multiple imputation (MI) to utilise information in the full cohort when data from the sub-studies are analysed. We will discuss using approximate imputation models, and using rejection sampling to impute values from the true distribution of the missing

values given the observed data. Simulation studies show that using MI to utilise full cohort information in the analysis of nested case-control and case-cohort studies can result in important gains in efficiency relative to traditional analysis. The methods will be illustrated using an example from nutritional epidemiology.

E656: **Covariance modelling in longitudinal data with informative observation**
Presenter: **Daniel Farewell**, Cardiff University, United Kingdom
Co-authors: Chao Huang
When using generalized estimating equations to model longitudinal data, both inconsistency (due to informative observation) and inefficiency (due to misspecified working covariances) are often of concern. A class of generalized inverses of singular working correlation matrices is described that allows flexible modelling of covariance within a subject's responses while offering robustness to certain kinds of informative observation. It is demonstrated how this class corresponds to dynamic models on the increments in the longitudinal responses, and illustrated its application to a study of peritoneal dialysis.

E117: **Multiple imputation of missing covariates with non-linear effects and interactions: An evaluation of statistical methods**
Presenter: **Shaun Seaman**, Medical Research Council, United Kingdom
Co-authors: Jonathan Bartlett, Ian White
When a model contains as covariates more than one function of a variable, it is not obvious how to impute missing values. Consider regression with outcome Y and covariates X and X^2. In 'passive imputation' a value X^* is imputed for X and X^2 is imputed as $(X^*)^2$. A recent proposal treats X^2 as 'just another variable' (JAV) and imputes X and X^2 under multivariate normality. We investigate three methods: 1) linear regression of X on Y to impute X and passive imputation of X^2; 2) same regression but with predictive mean matching (PMM); and 3) JAV. We also consider analyses involving an interaction. We explore the bias of the three methods, and illustrate methods using the EPIC study. When the analysis is linear regression and X is missing completely at random, JAV is consistent. JAV is biased when X is missing at random, but generally less so than passive imputation and PMM. Coverage for JAV was usually good when bias was small. However, with pronounced quadratic effects, bias can be large and coverage poor. When the analysis is logistic regression, JAV's performance can be very poor. PMM improved on passive imputation, but did not eliminate bias.

E203: **Multiple imputation of covariates by fully conditional specification: Accommodating the substantive model**
Presenter: **Jonathan Bartlett**, London School of Hygiene and Tropical Medicine, United Kingdom
Co-authors: Shaun Seaman, Ian White, James Carpenter
Missing covariate data commonly occur in observational and experimental research, and are often dealt with using multiple imputation (MI). Imputation of partially observed covariates is complicated if the substantive model is non-linear (e.g. Cox proportional hazards model), or contains non-linear (e.g. squared) or interaction terms, and standard software implementations of MI may impute covariates from models that are incompatible with such substantive models. We show how imputation by fully conditional specification, a popular approach for performing MI, can be modified so that covariates are imputed from models which are compatible with the substantive model. We investigate through simulation the performance of this proposal, and compare it to existing approaches. Simulation results suggest our approach gives less biased estimates than existing commonly used MI approaches across a range of settings.

E269: **Using the number of contact attempts to model non-ignorable missing outcome data**
Presenter: **Dan Jackson**, Medical Research Council, United Kingdom
Selection models are an established way of modelling missing outcome data. However using these models in practice is fraught with difficulties. In order to overcome these problems, Repeated Attempts Models (RAMs) have more recently been proposed. These computationally quite intensive models describe the probability that each contact attempt to obtain outcome data is successful. If only one contact attempt is made for each subject, then the RAM reduces to a conventional selection model. The types of dataset that motivate the use of RAMs will be described. It will be explained why we can generally expect RAMs to perform better than selection models, whilst emphasising the additional assumptions made by the RAM. It will be however argued that each dataset with information about failed contact attempts can be expected to provide its own quirks and difficulties, and hence providing generic software for RAMs is a very challenging task. Despite this the software that is available will be described and some results for a dataset where a RAM was used will be presented.

ES107 Room G16 APPLIED STATISTICS AND DATA ANALYSIS Chair: John Aston

E907: **Functional compositional Kriging of particle-size curves in heterogeneous aquifers**
Presenter: **Alessandra Menafoglio**, Politecnico di Milano, Italy
Co-authors: Alberto Guadagnini, Piercesare Secchi
The problem of predicting the spatial field of particle-size curves (PSCs) from sieve analysis data of soil samples collected within a heterogeneous aquifer is considered. We model PSCs as cumulative distribution functions and analyze their densities as functional compositional data, embedded into the Hilbert space of compositional functions endowed with the Aitchison geometry. We develop an original geostatistical methodology, grounded on a functional compositional kriging (FCK) approach, for the analysis of these types of georeferenced functional compositional data. Our approach allows providing predictions of the entire particle-size curve at unsampled locations, together with a quantification of the associated uncertainty, by fully exploiting both the functional form of the data and their compositional nature. This enables one to provide a complete characterization of the spatial distribution of soil textural properties of the reservoir. We test our approach on PSCs sampled along a borehole located within a field scale alluvial aquifer. A comparison between hydraulic conductivity estimates obtained via FCK approach and those predicted by classical kriging of effective particle diameters (i.e., quantiles of the PSCs) is performed and the results are discussed on the basis of the application of classical empirical formulations relating hydraulic conductivity to PSC quantiles.

E1017: **Sloshing in the shipping industry: Risk modeling through multivariate heavy-tail analysis**
Presenter: **Antoine Dematteo**, Telecom ParisTech, France
Co-authors: Stephan Clemencon, Nicolas Vayatis, Mathilde Mougeot
In the liquefied natural gas shipping industry, the *sloshing* phenomenon can lead to the occurrence of very high pressures in the tanks of the vessel. The issue of modelling the probability of the simultaneous occurrence of such extreme pressures is crucial for risk assessment. For d-dimensional heavy-tailed distributions, the extreme dependence structure is entirely characterised by the angular measure, a positive measure on the intersection of a sphere with the positive orthant in $\mathbb{R}^d$. As d increases, the mutual extreme dependence between variables becomes difficult to assess. Introducing an original latent variable technique, we show here how to find a low dimensional approximation of the angular measure which makes its estimation tractable, even for high-dimensional data set. The promoted approach is non-parametric as there is no assumption of any structure besides the heavy-tailed marginal distributions and has been applied successfully to pressure data to model sloshing. We use this representation to simulate heavy-tail distributions and perform Monte-Carlo simulation to estimate the probability of simultaneous occurrences of

extreme values. Our methodology is proved to be very efficient and relevant on simulated data. It is then used to assess the risk induced by the sloshing phenomenon is the LNG shipping industry.

E1162: **Sparse count data regression for predicting bicycle usage in a large-scale bicycle sharing system**
Presenter: **Yufei Han**, IFSTTAR, France
Co-authors: Etienne Come, Latifa Ouhekllou

Recent years witnessed arising interests in analyzing bicycle usage fluctuation over time in urban bicycle sharing systems. Due to social-economical factors of urban areas, bicycle usage at different stations usually presents spatio-temporal correlation. Investigating the correlation structure is helpful for predicting bicycle usage efficiently. For each station, we construct two L_1-norm regularized negative binomial regression models. The regression targets are the numbers of leaving and arriving bicycles at the station respectively at the specific time. The covariates of the regression models are the bicycle usage counts of all stations in the precedent hours and the context information, such as weather, hours of day and day type (holiday/working day). Negative binomial regression has superior performances in fitting over-dispersed bicycle counts. The L_1-norm regularization enforces sparsity of the regression coefficients. Only covariates with the non-zero regression coefficients are selected for prediction. This way we identify the local bicycle usage records and the global contexts affecting the bicycle usage of the station the most. The selected local bicycle usage records indicate the spatio-temporal correlation structure between the concerned station and the left. Experiments on practical usage data of Velib system in Paris verify our method.

E729: **Design alternatives for experiments with hard-to-change mixed-level factors**
Presenter: **Carla Almeida Vivacqua**, Universidade Federal do Rio Grande do Norte, Brasil
Co-authors: Linda Lee Ho, Andre Pinho

Many industrial experiments involve factors which have levels more difficult to change than others. This naturally leads to the hard-to-change factors being less reset than the easy-to-change factors during the execution of the experiment. The most common structure generated by these randomization restrictions are the so-called split-plot type designs. Methodologies for designing and analyzing such experiments have advanced lately, especially for two-level designs. However, practical needs may require the inclusion of factors with more than two levels. Guidelines for practitioners on the selection of a design involving mixed-level factors using different criteria are presented. An experiment to evaluate the quality of tyres is discussed as an illustration. It involves four factors: temperature, type of lubricant, type of pavement, and type of rubber. The described factors are arranged in decreasing order according to the degree of difficulty for changing the factor levels. Due to lack of previous planning, the actual experiment was executed as an unreplicated fractional factorial split-split-split-plot design. As a consequence of the inadvertent split-plotting, the evaluation of the significance of the main effects and the interaction of temperature and type of lubricant are compromised. Other alternatives, along with advantages, disadvantages and properties of each design are shown.

E918: **Prediction of patient reported outcome measures via multivariate ordered probit models**
Presenter: **Caterina Conigliani**, Universita Roma Tre, Italy
Co-authors: Andrea Manca, Andrea Tancredi

The assessment of health-related quality of life, and more generally of patient reported outcome measures (PROMs), is of central importance in many areas of research and public policy. Unfortunately, it is not uncommon for clinical studies to employ different PROMs, thus limiting the comparability of the evidence-base they contribute to. This issue is exacerbated by the fact that some national agencies are now explicit about which PROMs must be used to generate evidence in support of reimbursement claims. The National Institute for Health and Clinical Excellence for England and Wales, for instance, has identified in the EuroQol-5D (EQ-5D) the PROM to be used, and recognises the possibility to use a mapping approach to predict EQ-5D from other PROMs when EQ-5D have not been collected. Here we consider the problem of directly predicting EQ-5D responses from the Short Form 12 (SF-12), while recognising both the likely dependence between the five dimensions of the EQ-5D responses at the patient level, and the fact that the levels of each health dimension are naturally ordered. We also address the key problem of choosing an appropriate summary measure of agreement between predicted and actual results when analising PROMs. We carry out the analysis within a Bayesian framework, employing Markov chain Monte Carlo techniques.

E1217: **Social media data mining using statistical learning**
Presenter: **Wenqian Cheng**, London School of Economics, United Kingdom

There are various statistical learning methods can be used for analyzing social media's data. Initially, the focus is how to build statistical models from the data resources in the micro-blogs to predict changes in the stock market. The mood states which can be tracked and analyzed by statistical analysis might be a good forecast for the future trend of stock market. The aim of using text mining is to understand general public's perspectives towards certain keywords (e.g. specific companies or hottest topics). High-quality information is typically derived through the devising of patterns and trends through statistical pattern learning. Machine learning methods such as Clustering and Support Vector Machine are used for sentiment analysis. In addition, to discover the abstract "topics" that occur in a collection of posts, topic modeling was applied in the simulation study (mainly focusing on Chinese Micro-blog: Weibo).

ES77 Room B20 MATHEMATICAL ASPECTS OF COPULAS II Chair: Wolfgang Trutschnig

E1054: **Some new properties of the Bernstein estimator of a copula density**
Presenter: **Jan Swanepoel**, North West University Potchefstroom Campus, South Africa
Co-authors: Paul Janssen, Noel Veraverbeke

Copulas are functions that couple multivariate distribution functions to their one-dimensional marginal distribution functions. They are useful tools to model multivariate data and thus have found applications in many fields such as financial economics and in the analysis of multivariate survival data. For many problems it is not evident how to choose a parametric family of copulas to describe the data at hand and nonparametric estimation of the copula is therefore an attractive option to proceed with the analysis. Kernel and Bernstein estimators have been used as smooth nonparametric estimators for copulas. A general theorem that describes the asymptotic distributional behaviour of the Bernstein estimator of a copula density is discussed. Compared to existing results, the theorem does not assume known marginals, which makes the estimator applicable to real data. Furthermore, an easily calculable form of the estimator is derived for small samples and the results of a simulation study are reported which compare the behaviour of this estimator to other existing copula density estimators.

E231: **Sample d-copula of order m**
Presenter: **Jose Maria Gonzalez-Barrios**, Universidad Nacional Autonoma de Mexico, Mexico
Co-authors: Maria Magdalena Hernandez-Cedillo

Using the ideas of the construction of fractal copulas, we introduce the sample d-copula of order m with $d, m \geq 2$. This new copula allows us to estimate the multivariate copula based on a sample from a continuous distribution function in R^d or from a d-copula. The greatest advantage of the sample d-copula is the fact that it is already an approximating d-copula and that is easily obtained. We will see that these new copulas provide a nice way to study multivariate continuous data with an approximating copula which is simpler that the empirical multivariate copula, and that the

empirical copula is the restriction to a grid of a sample d-copula of order n, where n is the sample size. The sample d-copulas can be used to make statistical inference about the distribution of the data, as shown by the new statistical methodologies presented.

E720: **Inference on copula-based correlation structure models**

Presenter: **Enrico Foscolo**, Free University of Bozen-Bolzano, Italy

Modern data analysis calls for an understanding of stochastic dependence going beyond the Gaussian distribution and high-dimensionality. Inferential methods are developed for correlation structure models that combine second-order structure moments aiming at reducing dimensionality with less restrictive distributional assumptions; i.e., same marginal distributions and Normality among others. The method is grounded on copula functions and pseudo-likelihood. The asymptotic properties of the copula-based estimators are investigated. Monte Carlo simulations are carried out across varying data distributions and sample sizes in order to asses the robustness of the proposed procedure. Practical issues related to these findings are finally discussed.

E823: **Some classes of copulas related with target-based utilities**

Presenter: **Fabio Spizzichino**, University La Sapienza, Rome, Italy

Some issues of stochastic dependence, that can arise in the Target-Based Approach to the theory of utility and decisions under risk, will be considered. First, consider one-attribute decision problems, where X is a scalar random variable (a risky prospect) and U a utility function. U is assumed to be bounded and right-continuous, besides being increasing. Thus, by means of normalization, one can look at U as at a probability distribution function on the real line. A real random variable T, distributed according to U and such that T, X are stochastically independent, can be considered as a target in the decision problem: the expected utility $E(U(X))$ coincides in this case with the probability $P(T < X)$. We are interested in the extension of such an approach to cases where independence between the target and each prospect X is replaced by a condition of stochastic dependence, described by fixing a connecting copula C. The question about the reasonable properties which should be imposed on C leads to some special classes of copulas, which will be analyzed. The extension to multi-attribuite decision problem will also be considered.

E667: **A product of multivariate copulas**

Presenter: **Songkiat Sumetkijakan**, Chulalongkorn University, Thailand

Co-authors: Tippawan Santiwipanont, Korapin Luangsomboon

Based on a previous approach to multivariate shuffles, we give a definition of multivariate shuffles of copulas which is equivalent to the push-forward definition. It is well known that shuffling a bivariate copula amounts to $*$-multiplying by a shuffle of Min. In an attempt to find an analogous fact for shuffling a multivariate copula, we define a binary operation on the class of d-dimensional copulas and investigate its fundamental algebraic properties.

Monday 16.12.2013 14:30 - 15:50 Parallel Session Q – CFE

CS11 Room Woburn BAYESIAN METHODS IN MACROECONOMICS AND FINANCE Chair: Andreas Carriero

C864: **Granger-causal-priority and choice of variables in vector autoregressions**
Presenter: **Marek Jarocinski**, European Central Bank, Germany
Co-authors: Bartosz Mackowiak

A researcher is interested in a set of variables that he wants to model with a vector autoregression and he has a dataset with more variables. This question about which additional variables from the dataset to include in the VAR arises in many applications of VARs, in prediction and impulse response analysis. We develop a Bayesian methodology to answer this question. We rely on the idea of Granger-causal-priority, related to the well-known concept of Granger-noncausality. The methodology is simple to use, because we provide closed-form expressions for the relevant posterior probabilities. Applying the methodology to the case when the variables of interest are output, the price level, and the short-term interest rate, we find remarkably similar results for the United States and the euro area.

C1012: **Bayesian factor augmented vector autoregressive models**
Presenter: **Alessia Paccagnini**, Bicocca University, Italy

The use of Bayesian Factor Augmented Vector Autoregressive Models (Bayesian FAVAR) to explain macroeconomic fluctuations is discussed. The implementation of the Factor Models is focused on the recent economic crises, taking into account the different regimes over the last 40 years in the US economy. Different Bayesian FAVAR models are presented and compared in forecasting terms. Besides exercise in US economy data, the Bayesian FAVAR models are estimated using artificial data. Small and Medium scale Dynamic Stochastic General Equilibrium Models are the data generation process for the pseudo-data experiments. A special interest is dedicated to the use of financial factors to improve the forecasting ability of the macroeconomic variables.

C936: **Bayesian graphical vector auto regression**
Presenter: **Daniel Felix Ahelegbey**, University of Venice, Italy

A Bayesian graphical inference approach for the identification problem in vector autoregressive (VAR) models is proposed. The methodology decomposes the structural model into contemporaneous and lagged dependence. We then apply a stochastic search based on Markov chain Monte Carlo (MCMC) algorithm to sample the conditional independent relations in a form of directed acyclic graphs (DAGs). With the samples DAGs, we estimate a seemingly unrelated reduced form VAR. We present a synthetic dataset illustration and an empirical application to forecast US macroeconomic time series.

CS43 Room Beveridge EARLY WARNINGS INDICATORS AND MACRO-PRUDENTIAL POLICY II Chair: Gianluigi Mazzi

C558: **Evaluating early warning indicators of banking crises: Satisfying policy requirements**
Presenter: **Mikael Juselius**, Bank for International Settlements, Switzerland
Co-authors: Mathias Drehmann

Early warning indicators (EWIs) of banking crises should ideally be evaluated relative to their performance in terms of the macroprudential policy maker's decision problem. Several practical aspects of this problem are translated – such as difficulties in assessing the costs and benefits of various policy measures as well as requirements for the timing and stability of EWIs – into statistical evaluation criteria. Applying the criteria to a set of potential EWIs, it is found that the credit-to-GDP gap and a new indicator, the debt service ratio (DSR), consistently outperform other measures. The credit-to-GDP gap is the best indicator at longer horizons, whereas the DSR dominates at shorter horizons.

C738: **A regime switching skew-normal model for measuring financial crisis and contagion**
Presenter: **Cody Yu-Ling Hsiao**, Australian National University, Australia
Co-authors: Joshua Chan, Renee Fry-McKibbin

A regime switching skew-normal model for financial crisis and contagion is proposed in which we develop a new class of multiple-channel crisis and contagion tests. Crisis channels are measured through changes in "own" moments of the mean, variance and skewness, while contagion is through changes in the covariance and co-skewness of the joint distribution of asset returns. In this framework: i) linear and non-linear dependence is allowed; ii) transmission channels are simultaneously examined; iii) crisis and contagion are distinguished and individually modeled; iv) the market that a crisis originates is endogenous; and v) the timing of a crisis is endogenous. In an empirical application, we apply the proposed model to equity markets during the Great Recession using Bayesian model comparison techniques to assess the multiple channels of crisis and contagion. The results generally show that crisis and contagion are pervasive across Europe and the US. The second moment channels of crisis and contagion are systematically more evident than the first and third moment channels.

CS104 Room 349 CONTRIBUTIONS IN BAYESIAN ECONOMETRICS I Chair: Ghislaine Gayraud

C945: **Bayesian cointegration of multivariate time series**
Presenter: **Basile Marquier**, University of Sheffield, United Kingdom
Co-authors: Kostas Triantafyllopoulos, Miguel Juarez

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. Continuing on this trend, we discuss Bayesian analysis of the VEC model, using Markov chain Monte Carlo methods. We propose a Gibbs sampling scheme for estimation, while the model hyperparameters are estimated using historical data and the cointegration order is identified using Johansen's cointegration test. Prior specification is discussed in some detail. The methodology is illustrated with a data set comprising of 10 hedge fund indices, measuring the performance of the hedge fund industry. Some financial implications of the underlying cointegration relationships are discussed. A number of possible extensions of the proposed methodology is provided.

C1057: **An application of bayesian VAR-copula to the effect of macroeconomic risk appetite on growth**
Presenter: **Fuyu Yang**, University of East Anglia, United Kingdom
Co-authors: Roberto Leon-Gonzalez

GDP growth and monetary policy are closely intertwined with the size of bank balance sheets. Evidence suggests that bank balance sheets grow during booms, and shrink during recessions. We propose a flexible VAR-copula model to evaluate a nonlinear association amongst the looseness of bank balance sheet constraints, GDP growth, and macroeconomic risk premium variables. Compared with the conventional VAR model, the relaxation of the distributional assumption in the error term is more realistic in terms of model fitting, which allows for an asymmetric association

amongst the key variables in economic downturns. To serve the purpose of model comparisons, we evaluate the impulse response function and log-likelihood using both the VAR-copula and the VAR models.

C1182: **Semi-parametric GARCH via Bayesian model averaging**
Presenter: **Wilson Ye Chen**, University of Sydney, Australia
Co-authors: Richard Gerlach
In a standard GARCH model, the next period variance is driven by a quadratic function of the current shock. A semi-parametric GARCH model that makes fewer assumptions on the functional form of the news impact function is proposed by the authors, where the relationship between the next period variance and the current shock is modelled by a spline. The knots of the spline are determined using a Bayesian variable selection approach, where a Metropolis-Hastings algorithm is used to generate samples from the joint posterior distribution of the model parameters and the knots. The next period variance is then obtained by model averaging using the generated samples. In a simulation study, the performance of the proposed approach is compared to parametric GARCH models in cases where the parametric models are correctly specified and where they are misspecified. In an empirical study, the proposed approach is applied to the returns of stock indices and individual stocks, where the accuracy of the one-step-ahead volatility forecasts are assessed.

C1034: **Markov-switching quantile autoregressions: A Gibbs sampling approach**
Presenter: **Xiaochun Liu**, Emory University, United States
Co-authors: Richard Luger
Quantile location-scale models with autoregressive errors are considered in which the location and scale parameters are subject to regime switching. The regime changes are determined by the outcome of a latent, discrete-state Markov process. We propose a Gibbs sampling algorithm for posterior inference on the resulting Markov-switching quantile autoregressive model by using data augmentation and the mixture representation of the asymmetric Laplace distribution. An estimate of the marginal likelihood, and hence the Bayes factor for model comparison, is available as a by-product of the simulation procedure since all complete conditional densities used in the Gibbs sampler have closed-form expressions. A simulation study shows the usefulness of the proposed methodology and the new dynamic quantile model is further illustrated with an empirical application to the U.S. real interest rate by allowing for three possible regimes in the quantile process.

CS105 Room 104 ECONOMETRIC THEORY Chair: Stephen Pollock

C926: **To smooth or not to smooth: The case of discrete variables in nonparametric regression**
Presenter: **Degui Li**, University of York, United Kingdom
Co-authors: Leopold Simar, Valentin Zelenyuk
The nonparametric smoothing technique with both discrete and categorical regressors is considered. In the existing literature, it is generally admitted that it is better to smooth the discrete variables, which is similar to the smoothing technique for continuous regressors but using discrete kernels. However, such an approach might lead to a potential problem which is linked to the bandwidth selection for the continuous regressors due to the presence of the discrete regressors. Through the numerical study, we find that in many cases, the performance of the resulting nonparametric regression estimates may deteriorate if the discrete variables are smoothed in the way addressed so far in the literature, and that a fully separate estimation without any smoothing of the discrete variables may provide significantly better results. As a solution, we suggest a simple generalization of the popular local linear simple smoothing approach to address this problem and to provide estimates with more robust performance. We analyze the problem theoretically, develop the asymptotic theory for the new nonparametric smoothing method and study the finite sample behavior of the proposed generalized approach through extensive Monte-Carlo experiments as well present an empirical illustration.

C1096: **On overidentifying restrictions tests and their incremental versions**
Presenter: **Milan Pleus**, University of Amsterdam, Netherlands
Overidentifying restrictions tests are examined in the context of IV regression in a single adequately specified structural equation, where the focus is on the incremental versions of these tests. Well-known tests such as the Sargan test , Basmann's test and a Hausman-type test are examined. Alternative formulations are discussed and clarified, such as testing parameters in an auxiliary regression. We derive non-local power properties of these tests and clarify recent papers that state that these tests provide little comfort in testing the actual moment restrictions. A Cornish Fisher correction for the Sargan test is re-examined. Additionally, simulation experiments are designed by choosing econometrically relevant design parameters which are functions of the DGP parameters. We compare the different test statistics and conclude that in small samples the Cornish Fisher corrected test outperforms the others in terms of size control. However, with respect to the power properties of the test we find that the corrected Sargan test performs less than the other statistics.

C1066: **On distributions of ratios**
Presenter: **Simon Broda**, University of Amsterdam, Netherlands
Co-authors: Raymond Kan
A large number of exact inferential procedures in statistics and econometrics involve the sampling distribution of ratios of random variables. If the denominator variable is positive, then tail probabilities of the ratio can be expressed as those of a suitably defined difference of random variables. If in addition, the joint characteristic function of numerator and denominator is known, then standard Fourier inversion techniques can be used to reconstruct the distribution function from it. Most research in this field has been based on this correspondence, but which breaks down when both numerator and denominator are supported on the entire real line. Inversion formulae and saddlepoint approximations that remain valid in this case, and reduce to known results when the denominator is almost surely positive, are derived. Applications include the IV estimator of a structural parameter in a just identified equation.

C1131: **On the need of regularization: Wald-type test under nonregular condidtions**
Presenter: **Artem Duplinskiy**, Maastricht University, Netherlands
Hypotheses testing in presence of a possibly singular covariance matrix is studied. An alternative way to handle possible nonregularity in a covariance matrix of a Wald test is proposed. A Wald type test $\lambda_1 = (R(\beta) - c)'(R(\beta) - c)$ is considered, where the inverse of the matrix of the second moments is substituted by the identity matrix in calculation of the quadratic form. It uses the fact that the square root of a positive semi-definite matrix is a continuous transformation of the matrix. The test statistic is no longer pivotal, however continuity of the square root allows to approximate the asymptotic distribution of this test, using bootstrap. Two types of approximation are used. The first relies on the asymptotic distribution of the elements of the vector $(R(\beta) - c)$ in the quadratic form and uses parametric bootstrap to approximate the asymptotic distribution of λ_1 by drawing from the asymptotic distribution of $(R(\beta) - c)$ with estimated covariance matrix. The second applies residual bootstrap to obtain the distribution of λ_1. Despite that λ_1 is not pivotal and one can not expect it to provide asymptotic refinements in this case, residual bootstrap delivers critical values, which control empirical size and show good empirical power even in small samples.

CS126 Room Senate FINITE SAMPLE AND IDENTIFICATION-ROBUST METHODS Chair: Bertille Antoine

C1008: **Inference with mixed frequency data**
Presenter: **Bertille Antoine**, Simon Fraser University, Canada
Consider a model where relevant variables are observed at different frequencies, say monthly (high frequency) and quarterly (low frequency). For the applied researcher who wants to use both pieces of information together, two common solutions are available: (i) select the lowest frequency (say quarterly) and aggregate the high-frequency (say monthly) observations; (ii) select the highest frequency (say monthly) and use some interpolation techniques to make-up the 'missing' observations. Both methods are quite simple, and somewhat flexible in the aggregation/interpolation techniques one can use. However, none is completely satisfactory as valuable information disappears in the first case, while potentially misleading information is used in the second one. We propose here a GMM-type inference procedure that relies on a new interpolation scheme that makes use of all available information. We show that our estimator is consistent and asymptotically normally distributed. Its small sample properties are investigated by simulations. We find that our estimator performs well compared to existing ones, especially with small sample sizes.

C1064: **Exact moment-based tests of linearity in Markov-switching models**
Presenter: **Richard Luger**, Georgia State University, United States
Co-authors: Jean-Marie Dufour
A fundamental question that arises when using Markov-switching models is whether the data-generating process in indeed characterized by regime changes. Under the null hypothesis of linearity, some parameters are not identified and scores are identically zero thereby invalidating standard likelihood-based inference. We propose new Lagrange multiplier tests of linearity that exploit the moment structures that these models can produce as a function of the transition probabilities and parameters of the underlying state densities entering the switching process. In particular, we show that maximized Monte Carlo tests provide provably exact tests for such models despite the problems that undermine conventional methods. The new test procedure is illustrated with an application to an autoregressive model of quarterly U.S. output growth.

C1045: **Necessary and sufficient conditions for nonlinear parametric function identification**
Presenter: **Xin Liang**, McGill University, Canada
Co-authors: Jean-Marie Dufour
The conditions for both global and local parametric function identifications are studied and extensions to some of the well-known classical rank conditions are made. We make contributions in the following aspects. First, we introduce the concept of local identification around a parameter value and distinguish it from the classical definition of local identification at a parameter value. This concept will help achieve "global" parametric function identification in the reduced parameter space. Second, we propose both necessary and sufficient conditions for global and local parametric function identifications when the classical assumption of full rank Jacobian matrix fails. It is noted that such necessary and sufficient conditions for parametric function identification are valid under rather weak assumptions and they hold with or without linear and nonlinear restrictions. Third, we provide a group of equivalent conditions for parametric function identification that are intuitive and easy to check in practice. The useful rank condition for parametric function identification is just one of these equivalences. Fourth, we verify that the established identification conditions can be readily applied to the likelihood model setups through Kullback-Leibler divergence and local parametric function identification can be achieved by higher order moments. Fifth, the results can help empirical research for identification since the general parametric function identification conditions proposed are easily applicable to the simultaneous models and the DSGE models. It shows that most of the classical identification conditions in the literature can be generalized as special cases.

CS24 Room Montague CONTRIBUTIONS TO FACTOR MODELS AND FORECASTING APPLICATIONS Chair: Rodney Strachan

C116: **Real-time nowcasting of nominal GDP**
Presenter: **Danilo Leiva-Leon**, University of Alicante, Spain
Co-authors: Marcelle Chauvet, William Barnett
Due to the zero lower bound interest rate, some non-conventional policies, such as Nominal GDP targeting, start to become sounded among policy makers. We focuse on providing early assessments of current quarterly Nominal GDP growth for the US economy. These nowcasts are computed by using the exact amount of information that policy makers have at hand at the time that predictions are done. We explore the predictive ability of several univariate and multivariate specifications, by also looking for the most helpful indicators in performing this task. The results show that, among the proposed candidates, a small scale dynamic factor model that contains information of real economic activity, inflation dynamics and Divisia monetary aggregates, produces the most accurate nowcasts of Nominal GDP.

C1138: **Comparing linear and non-linear dynamic factor models for large macroeconomic datasets**
Presenter: **Alessandro Giovannelli**, University of Tor Vergata - School of Economics, Italy
A non-linear extensions for macroeconomic forecasting is provided using large datasets based on dynamic factor model (DFM). The idea is to allow the factors to have a non-linear relationship to the input variables using the methods of squared, kernel and neural networks principal components analysis (PCA). We compare the empirical performances of these methods with DFM, conducting a pseudo forecasting exercises based on a U.S. macroeconomic dataset with 115 monthly variables spanning 1960 - 2012. The results obtained from the empirical study suggest that, on average for short horizons, nonlinear DFM improves DFM in almost 20% of cases, whereas for longer horizons the number of series increases to 40% of cases. Despite these results, there are categories of series that appear to be more prone to non-linear forecasts, among them include employees, prices and money categories.

C1077: **A criterion for the number of factors in a data-rich environment**
Presenter: **Ard den Reijer**, Sveriges Riksbank, Sweden
Co-authors: Pieter Otter, Jan Jacobs
Recently, a formal justification has been provided for the use of the scree test in determining the number of factors in large factor models. We derive a criterion for the determination of the number of factors, that is associated with the previous test statistic but does not rely on frequency domain analysis. Monte Carlo experiments compare the finite-sample properties of our criterion to previous ones. The simulations show that our criterion and another recent one are superior. An application with an U.S. macroeconomic data set illustrates our procedure.

C921: **Invariant inference and efficient computation in the static factor model**
Presenter: **Rodney Strachan**, The Australian National University, Australia
Co-authors: Joshua Chan, Roberto Leon-Gonzalez
Factor models are used in a wide range of areas. Two issues with Bayesian versions of these models are a lack of invariance to ordering of and scaling of the variables and computational inefficiency. Invariant and efficient Bayesian methods for estimating static factor models are developed. This approach leads to inference on the number of factors that does not depend upon the ordering or scaling of the variables, and we provide arguments to explain this invariance. Beginning from identified parameters which have nonstandard forms, we use parameter expansions to obtain a specification with standard conditional posteriors. We show significant gains in computational efficiency. Identifying restrictions that are

commonly employed result in interpretable factors or loadings and, using our approach, these can be imposed ex-post. This allows us to investigate several alternative identifying (non-invariant) schemes without the need to respecify and resample the model. We apply our methods to a simple example using a macroeconomic dataset.

CS37 Room Bedford TIME-VARYING PARAMETERS IN FINANCE AND ECONOMICS I Chair: Jim Griffin

C951: **Non-parametric estimation of intraday spot volatility: Disentangling instantaneous trend and seasonality**
Presenter: **Thibault Vatter**, University of Lausanne, Switzerland
Co-authors: Valerie Chavez-Demoulin, Hau-Tieng Wu, Bin Yu

A new framework is provided to model trends and periodic patterns in high-frequency financial data. Seeking adaptivity to ever changing market conditions, we relax assumptions from usual intraday models in two directions. First, the realized volatility (i.e. the slowly time-varying component) is modelled as an instantaneous trend which may evolve in real time over a given day. Second and more importantly, the seasonality is no longer assumed constant over the sample. We define an instantaneous version of the popular Fourier flexible form and we provide associated estimators. In a large-scale simulation study, we analyze the properties of those estimators and develop resampling procedures to obtain confidence intervals. As an empirical application, we use foreign exchange data to compare our framework to other seasonality estimators.

C997: **Stochastic volatility and Kalman filter in financial time series: Evidence from state space models**
Presenter: **Dahiru Bala**, Kyushu University, Japan
Co-authors: Taro Takimoto

State space models and Kalman filter (KF) algorithms are at the centre of recent advances in financial econometrics in modeling unobserved variables including rational expectations, measurement errors, and missing observations. Specifically, they are applied to compute exact, finite sample forecasts for Gaussian and non-Gaussian autoregressive moving average, Markov switching, and time varying parameter (TVP) models. During periods of financial crisis or market crashes, asset price volatilities often exhibit jumps and breaks leading to extreme values and distributions with fat tails. We estimate stochastic volatility (SV) parameters for Japan's financial time series using state space based approaches within the context of volatility stylised facts. We reviewed recent techniques and examined the rapid growth of state space models in finance from both the classical and Bayesian perspectives. The KF algorithm and SV techniques are used to analyse daily exchange rates, Nikkei 225 stock returns and bond yields. We find that volatility persistence parameter in SV models which indicates volatility clustering is comparable with those of GARCH-type models. In examining dynamics of these series, the TVP models are compared with other state space and conditionally heteroscedastic based models within the framework of efficient estimation of hyperparameters (variances) and forecasting performance of these models.

C1005: **Assessing macroeconomic effects of nontraditional monetary policy in Japan with TVP-SVAR model**
Presenter: **Jianhui Xu**, Kyushu University, Japan
Co-authors: Taro Takimoto

The macroeconomic impact of non-traditional monetary policy or quantitative easing (QE) implemented by the Bank of Japan (BOJ) in the last two decades is examined. We quantify the effects of these monetary policies by focusing on the impact of lower long-term interest rates on the overall economy. We conduct counterfactual analysis to estimate the impact of QE on output and inflation using the time-varying parameter VAR (TVP-SVAR) model. Results reveal that QE has significant impact on Japan's key macroeconomic variables. Based on this model, we forecast the trends of inflation and output. Finally, implications and conclusions for recent Japanese economic policies "Abenomics" were provided.

C996: **Testing the preferred-habitat theory: The role of time-varying risk aversion**
Presenter: **Till Strohsal**, Free University Berlin, Germany

Th purpose is to test the preferred-habitat theory under time-varying risk aversion. The predicted positive relation between the term spread and relative supply of longer-term debt is stronger, when risk aversion is high. To capture this effect, a time-varying coefficient model is introduced and applied to German bond data. The results support the theoretical predictions and indicate substantial time variation: Under high risk aversion, yield spreads react by about 3 times as much as when risk aversion is low. The accumulated response of term spreads to a one standard deviation change in debt supply ranges between 5 and 33 basis points.

CS81 Room Gordon CONTRIBUTIONS IN NON-LINEAR DYNAMIC MODELS Chair: Haroon Mumtaz

C1042: **European debt crisis: Contagion or interdependence of the main euro zone indices**
Presenter: **Rachida Hennani**, Montpellier - Lameta, France
Co-authors: Michel Terraza

Recent events have shaken successively in European financial markets suggest a phenomena of interdependence between stock market indices. The relationships that exist between these countries may intensify in times of crisis. It is more appropriate to use the term of contagion. This difference in terminology translates different realities that affect the choice of economic policies. The presence of contagion in times of crisis between several indices is highlighted and confirms the leading role played by the German index is confirmed. Our empirical study shows the outperformance of Mackey-Glass-DCC-GARCH model over the DCC-GARCH model, indicating that the highly complex chaotic structures are taken into account.

C1103: **A test for asymmetries in Markov-switching structural VAR models**
Presenter: **Frederic Karame**, University of Maine, France

The development of nonlinear representations and of generalized IRFs favored the study of the variables behavior in response to an economically-identified shock as regards (i) the state of the system when the shock occurs, (ii) the size of the shock and (iii) the sign of the shock. Generalized IRFs are widely used in threshold representations but have known no comparable development in Markov-switching VAR. We show that the direct transposition of linear IRF displays poor properties in terms of asymmetries. We propose a Generalized impulse-response functions for Markov-switching structural VAR and use it to develop a real statistical test for all types of asymmetries. At last, we implement this new tool on three issues in labor economics, monetary policy and finance.

C1165: **Stability analysis of an implicitly-defined labor market model**
Presenter: **Vivaldo Mendes**, ISCTE-IUL, Portugal
Co-authors: Diana Mendes

Until very recently, the pervasive existence of models exhibiting well-defined backward dynamics but ill-defined forward dynamics in economics and finance has apparently posed no serious obstacles to the analysis of their dynamics and stability, despite the problems that may arise from possible erroneous conclusions regarding theoretical considerations and policy prescriptions from such models. This problem have dealt with extensively in the past by assuming the existence of symmetry between forward and backward dynamics, even in the case when the map can not be invertible either forward or backwards. However, this procedure has been seriously questioned over the last few years in a series of works dealing with implicit difference equations and inverse limit spaces. The aim is to explore a previous search and matching labor market model in order to: (i) show that chaotic dynamics may still be present in the model for acceptable parameter values; (ii) clarify some open questions related with

the admissible dynamics in the forward looking setting, by providing a rigorous proof of the existence of cyclic and chaotic dynamics through the application of tools from symbolic dynamics and inverse limit theory

C1183: **Nonlinear forecasting with many predictors under stochastic volatility**
Presenter: **Peter Exterkate**, Aarhus University, Denmark
Kernel ridge regression is a technique to perform ridge regression with a potentially infinite number of nonlinear transformations of the independent variables as regressors. This makes it a powerful forecasting tool, which is applicable in many different contexts. However, it is usually applied only to independent and identically distributed observations. A variant of kernel ridge regression for time series with stochastic volatility is introduced. We set up the estimation problem in a Bayesian manner and derive a Gibbs sampler to obtain draws from the predictive distribution. A simulation study and an application to forecasting the distribution of returns on the S&P500 index are presented, and we find that kernel ridge regression is a promising forecasting tool for time series with stochastic volatility.

CS120 Room Bloomsbury CONTRIBUTIONS ON LONG MEMORY IN ECONOMIC AND FINANCIAL TIME SERIES Chair: Tommaso Proietti

C1090: **Long memory and regime switching beta model for hedge fund dynamics**
Presenter: **Mohamed-Ali Limam**, University of Montpellier, France
Co-authors: Virginie Terraza, Michel Terraza
Long memory and regime switching models are combined to investigate nonlinear dynamics of hedge fund returns and their risk exposure captured by market risk factor. The long memory component is taken into account both in the conditional mean using ARFIMA model and in the conditional variance using FIGARCH, FIAPARCH and HYGARCH models. Major results from a sample of European Hedge funds, show that hedge funds exhibit significant non-linear dynamics in mean and in conditional variance as well as two levels of persistence. We find a non-linear exposure to equity market risk factor since for most of hedge fund series the exposure to the MSCI during periods of bear market is smaller compared to crisis periods. The analysis of conditional variance reveals the risky part of hedge funds which are characterized by a high persistent volatility in both regimes. Long memory parameters are below 1 in low volatility regime which means that shock effects on variance will persist for a long time and that they converge slowly to a steady state. Henceforth, contrary to the latter mean-reverting behavior, conditional variance in high volatility regime is a non-stationary process. Even though remote shocks affect the present value of the series, this will tend to the value of its conditional variance in the long run.

C1115: **Horizon dependent portfolio allocation**
Presenter: **Kai Ming Lee**, Ortec Finance Research Center, Netherlands
In conventional investment wisdom, stocks investments, while volatile at short horizons, are often considered to be less risky when held for the long run. Academic debate however, shows little consensus on this issue. We present new results on short and long horizon returns for both stock and bond investments. Using separate VAR models on different frequency bands, rich long range dependence and comovements can be modeled without the tight parameterisations imposed by low-order VAR models defined on the entire spectrum. Optimal portfolio allocation results are shown to depend on the investment horizon, and contrasted with results from traditional VAR approaches.

C930: **How dynamic networking drives stock market volatility, and what it matters**
Presenter: **Hongwei Chuang**, Institute of Statistical Science Academia Sinica, Taiwan
Recently, there has been an explosion of research interest in networks, especially in the fields of biology, computer science, and sociology. This study aims to explore applications of network analysis to study the investors' trading behavior and profits. We find an agent's profitability is determined by the network centrality measure which acts in cooperation with information diffusion hypothesis. Moreover, we also investigate the relationship between the dynamic networking and market volatility by using an unique data of brokers' all trades on the Taiwan Stock Exchange (TSE) over a long period from 2001 to 2011. A financial linkage is specifically determined by using Vector Autoregression (VAR) model to measure the multivariate Granger causality. Our empirical findings are not only consistent with several predictions from the theory of financial networks, but also provide an alternative explanation for the stylized facts of stock returns about volatility exhibited by the market returns such as the clustering and the correlation persistence.

C1112: **Moment and memory properties of exponential ARCH models**
Presenter: **Xiaoyu Li**, University of Exeter, United Kingdom
The moment and memory properties of the exponential ARCH (EARCH) models are investigated. It derives the conditions for the second, fourth and higher order moments and the near-epoch dependence (NED) properties of the EARCH processes. The properties of the NED are used to determine the memory properties of the EARCH models, because they might be easier to be verified than the mixing processes. Moreover, the class of the NED process covers a large number of time series processes which do not satisfy the conditions of mixing processes. This study shows that the shocks of EARCH infinite processes can be L1-NED and L2-NED on an independent process and they can also be geometrically NED on an independent process under mild conditions. This means that the EARCH infinite models are able to be of both hyperbolic memory and geometric memory. The EGARCH(1,1) model is geometric memory.

Monday 16.12.2013 14:30 - 15:50 Parallel Session Q – ERCIM

ES112 Room B35 COMPUTATIONAL STATISTICS II **Chair: Michael Berry**

E1085: **Tree-based sparse estimation of varying coefficients in ordinal mixed-effect models**
Presenter: **Reto Buergin**, University of Geneva, Switzerland
Co-authors: Gilbert Ritschard
Varying-coefficients models have found widespread use in analysis of panel data. They allow coefficients of predictors to be functions of other covariates, such as time or confounding variables. In high-dimensional settings, the specification of such coefficient-functions is challenging and non-parametric smoothing methods often fail. We introduce a new method based on linear mixed-effect models for clustered or repeated ordinal response data, including ordinal panel data. The method relies upon a tree-based algorithm that estimates varying-coefficients by a piecewise constant function. The algorithm can handle non-linearity, high-dimensionality, mixed measurement scales, time-varying and time-invariant covariates and statistically controls the selection of noisy covariates. The potential of the method is demonstrated with an application, using data from the British Household Panel Study, that examines how the effect of unemployment on individual well-being varies across individual characteristics and life circumstances. Performance of approximating coefficient-functions is evaluated by simulation and compared with non-parametric smoothing methods. An R package is provided for general use.

E1224: **An efficient algorithm for estimating block recursive simultaneous equations model**
Presenter: **Mircea Ioan Cosbuc**, Alexandru Ioan Cuza University of Iasi, Romania
Co-authors: Cristian Gatu, Erricos J. Kontoghiorghes
A new efficient strategy for estimating the block recursive Simultaneous Equations Model with non-full rank variance-covariance matrix is proposed. The estimation procedure uses the Generalized Singular Value Decomposition and derives a numerically-stable three-stage least squares estimator. The algorithm exploits the block-diagonal and banded structures of the matrices involved in the factorization as well as the block recursive properties of the model. Experimental results illustrate the computational efficiency of the new estimation algorithm.

E993: **A framework for measuring the stability of recursive partitioning results**
Presenter: **Michel Philipp**, University of Zurich, Switzerland
Co-authors: Thomas Rusch, Kurt Hornik, Carolin Strobl
Recursive partitioning approaches, such as classification and regression trees and model-based recursive partitioning, have become established and frequently applied methods for exploring unknown structures in complex data. Despite their popularity, a major drawback of these methods is their instability, since small random changes in the data can cause large changes in the results. For the interpretation in practical applications, however, stability is a crucial requirement to draw consistent conclusions - but currently recursive partitioning methods provide no statistical theory for judging the confidence of their results. We therefore present a new framework for assessing the stability of recursive partitioning results, that is based on a family of distance measures for partitions. The new approach is motivated, illustrated and compared to existing distance measures by means of real and simulated examples.

E1210: **HMM-based learning with generalised hidden path estimators**
Presenter: **Alexey Koloydenko**, Royal Holloway University of London, United Kingdom
Co-authors: Juri Lember, Ufuk Mat
Despite the fact that hidden Markov models have long been known in both statistics and computer science, they have recently seen new surges of methodological developments. These include several families of path estimators, or decoders, based on generalised risks, as well as families of algorithmically defined decoders. The newly proposed families naturally include the most popular Viterbi (MAP) and posterior (MPM) decoders, as well as virtually all of their variants employed in practice, as special cases. Moreover, the new families offer the flexibility to approximate risk functions that may otherwise be too complex to allow for closed form algorithmic minimisers. However, the underlying assumption is that all of the model parameters are known, which is rarely the case in practice. We embed these generalised path estimators in the full framework wherein the hidden path and any of the unknown parameters need to be learned simultaneously, sometimes from scarce amounts of data. In particular, we study the Viterbi Training (Classification EM) and other segmentation-based training protocols, as well as the Bayesian approach, also allowing for partially labelled training data. Extensions to hidden Markov random fields and applications with tensor-valued observations are also discussed. Some asymptotic considerations are also presented.

ES49 Room B30 CONTRIBUTIONS TO NONPARAMETRIC AND ROBUST STATISTICS **Chair: Peter Rousseeuw**

E1129: **The advantages of utilizing group means in estimating the Lorenz curve and Gini index from grouped data**
Presenter: **Joseph Gastwirth**, George Washington University, United States
Co-authors: Merritt Lyon, Li Cheung, Brian Dumbacher
A histogram based method has been recently proposed for estimating the Lorenz curve and Gini index from grouped data that did not include the group means that are typically reported by government agencies. When comparing this method with existing ones utilizing the group means, it is assumed that the mid-points of each group were the group means. As the density of the income distribution tends to increase and then decrease, especially at the upper incomes, this approximation is questionable. After reviewing the theoretical basis for the additional information, it is shown that as the number of groups increase, the distance between the upper and lower bounds on the Gini index obtained from methods using the group means decreases, while this is not necessarily true for the histogram method. These results are illustrated on the 2010 Census data, which is reported in 42 groups. A simple interpolation method incorporating the group means is also provided, which yields an estimated Gini index that lies between the upper and lower bounds.

E1130: **Generalized kernel estimates**
Presenter: **Jiri Zelinka**, Masaryk University, Czech Republic
The kernel estimation of a probability density function f (briefly density) plays a very important role in non-parametric statistics. The basic estimate is $\hat{f}_h(x) = 1/(nh)\sum_{i=1}^{n} K((x-X_i)/h)$ for a random sample $X_1,\ldots,X_n$, a kernel K and a bandwidth h. We can easily obtain the kernel estimates of the derivative of the density or of the cumulative distribution function by derivation or integration of this relationship. The aim is to generalize these estimates for any linear operator L applicable to the density, i.e. the estimates of the form $\widehat{Lf}_h(x) = 1/(nh)\sum_{i=1}^{n} L(K((x-X_i)/h))$ will be the subject of our interest. Basic statistical properties of these estimates are deduced and several applications are presented.

E1132: **Robust procedures for detection changes in linear models**
Presenter: **Zuzana Praskova**, Charles University in Prague, Czech Republic
Co-authors: Ondrej Chochola
Robust procedures for detecting changes in parameters of linear models will be considered. They have been developed to reduce some sensitivity

of parameter estimators with respect to outlying observations and heavy-tailed distributions. A class of CUSUM-type test statistics based on M-estimators and weighted M-residuals is proposed and limit properties of test statistics are studied both under the null hypothesis of no change and under local and fixed alternatives in case that both regressors and errors are sequences of weakly dependent random variables or vectors. Properties of estimators of a long run variance matrix will be also discussed.

E1271: **Robust and nonparametric procedures in change point detection**
Presenter: **Jaromir Antoch**, Charles University, Czech Republic
We will consider model $Y_{in} = x_i'\beta + x_i'\delta I_{[i>m_n]} + e_i$, $i = 1,\ldots,n$ where $m_n(\leq n)$, β and δ are unknown parameters, x_i, $i = 1,\ldots,n$, are known design points and $e_1,\ldots,e_n$ are iid random errors. It describes the situation where the first m_n observations follow the linear model with the parameter β and the remaining $n - m_n$ observations follow the linear model with the parameter $\beta + \delta$. Of prime interest will be the estimation of the parameter m_n, usually called the *change point*. We will concentrate on robust *M*-estimators of m_n. The sensitivity of the results and a comparison with classical and Bayesian methods will be demonstrated on both simulated and real data sets. We will also shortly comment the situation with more than one change point.

ES113 Room B34 STATISTICAL INFERENCE I Chair: Michael Falk

E985: **Empirical and simulated adjustments of composite likelihood ratio statistics**
Presenter: **Manuela Cattelan**, University of Padova, Italy
Co-authors: Nicola Sartori
Composite likelihood is a type of pseudo-likelihood which has gained much popularity in the recent years and it is often used when the full likelihood is too complex to deal with. It is also possible to define test statistics based on the composite likelihood which are analogous to those based on the full likelihood. However, Wald type statistics based on the composite likelihood are not invariant to reparameterisation, while score type statistics based on composite likelihoods seem to suffer from numerical instability. Finally, composite likelihood ratio statistics have an awkward asymptotic distribution which is a weighted sum of independent chi-squared random variables with one degree of freedom with weights given by the eigenvalues of a matrix related to the Godambe information matrix. Various adjustments of the composite likelihood ratio statistics have been proposed to overcome this problem. Some of these adjustments require the computation of matrixes involved in the calculation of the Godambe information matrix. We investigate the performance of the adjusted composite likelihood ratio statistics when those matrixes are not available analytically, but have to be computed either empirically or through simulation based methods.

E1171: **Inference in step-stress models: Cumulative exposure vs. tampered failure rate**
Presenter: **Maria Kateri**, RWTH Aachen University, Germany
Co-authors: Udo Kamps
Step-stress models are particular accelerated life testing models, commonly used in reliability assessment of products or machine elements. These models are usually considered under the cumulative exposure assumption. In this case it is well known that, for exponential lifetime distributions, explicit expressions can be obtained for maximum likelihood estimators of parameters as well as for their conditional density functions or conditional moment generation functions, given their existence. Applying a failure rate based approach instead, similar results can be also obtained for underlying lifetime distributions out of a general scale family of distributions, which allows for a flexible modeling. The two approaches are compared in Type-I and Type-II censored experiments.

E1087: **Partial mean processes with generated regressors: Continuous treatment effects and nonseparable models**
Presenter: **Ying-Ying Lee**, University of Oxford, United Kingdom
The unconditional distribution of potential outcomes with continuous treatments and the quantile structural function in a nonseparable triangular model can both be expressed as a partial mean process with generated regressors. We propose a multi-step nonparametric kernel-based estimator for this partial mean process. A uniform expansion reveals the influence of estimating the generated regressors on the final estimator. In the case of continuous treatment effects, an unconfoundedness assumption leads to regression on the generalized propensity score, which serves as the generated regressor in the partial mean process. Nonseparable triangular models commonly include a conditional independence assumption that yields a control function approach to deal with endogeneity. In a preliminary step, the control variable is estimated nonparametrically as a generated regressor. Our general partial mean process results can then be applied to provide the asymptotic distribution of the nonparametric estimator for the average and quantile structural functions. By extending the results to Hadamard-differentiable functionals of the partial mean process, it is possible to provide the limit distribution for estimating common inequality measures and various distributional features of the outcome variable, such as the Gini coefficient.

E1071: **Modelling bimodal discrete data using COM-Poisson mixtures**
Presenter: **Pragya Sur**, Indian Statistical Institute, India
Co-authors: Galit Shmueli, Smarajit Bose, Paromita Dubey
Bimodal truncated discrete distributions are frequently observed in aggregate survey data and in user ratings when respondents are mixed in their opinion. They also arise in censored count data, where the highest category might create an additional mode. Usually the Poisson distribution is used for fitting count data. In cases of bimodal data often mixtures of Poisson or truncated Poisson distributions serve the purpose. However, this method is suitable only for modeling equi-dispersed distributions and is limited in its ability to capture bimodality. The Conway-Maxwell-Poisson (CMP) distribution is a two-parameter generalization of the Poisson distribution that allows for over-and under-dispersion. We propose a mixture of two CMPs for capturing a wide range of truncated discrete data, which can exhibit bimodal behavior. A method for estimating the parameters of a mixture of two CMPs is suggested. Our approach uses tools from the EM algorithm and introduces a special two-step optimization within the M step to estimate multiple parameters. We examine computational and theoretical issues in this regard. The methods are illustrated for modeling ordered rating data as well as truncated count data, using both simulated and real examples.

ES114 Room B33 MULTIVARIATE STATISTICS I Chair: Robert Serfling

E278: **The multisample block-diagonal equicorrelation and equivariance test: A near-exact approach**
Presenter: **Filipe Marques**, FCT - Universidade Nova de Lisboa, Portugal
The equicorrelation and equivariance test, also known as compound symmetry, or intraclass correlation test, is of great importance in different fields in multivariate statistics like in Analysis of Variance, Profile Analysis and Growth Curve analysis. We consider an extension of this test based on the composition of three tests; the equality of covariance matrices test, the independence of several groups of variables test and the equicorrelation and equivariance test. The objective is to derive a procedure that allows us to test whether in different populations we have equal covariance matrices all with a block-diagonal equicorrelation and equivariance structure, i.e. a block-diagonal matrix where each diagonal block has an equicorrelation and equivariance structure. We designate this test by the multisample block-diagonal equicorrelation and equivariance test. Taking this test as the composition of the three tests mentioned above we show that it is possible to obtain the likelihood ratio test statistic, the expression of its null moments and the characteristic function of its logarithm. This approach also allows us to write the characteristic function

of the logarithm of likelihood ratio test statistic in a way that enables the development of new and highly accurate near-exact distributions for that statistic. These distributions are easy to implement computationally and will allow us to carry out the test with a high precision.

E974: **Testing for a δ-neighborhood of a generalized Pareto copula**
Presenter: **Stefan Aulbach**, University of Wuerzburg, Germany
Co-authors: Michael Falk

It is a well-known fact that a multivariate distribution function is in the max-domain of attraction of an extreme value distribution if and only if this is true for the corresponding copula and the univariate margins. Furthermore, a copula satisfies the extreme value condition if and only if the copula is tail equivalent to a generalized Pareto copula (GPC). We consider a more specific condition, i.e., whether a copula is in a certain δ-neighborhood of a GPC. This δ-neighborhood condition can be checked by means of a suited chi-square goodness-of-fit test. Since the corresponding p-value is highly sensitive to a proper selection of a certain threshold, we also present a graphical tool that makes the decision, whether or not to reject the hypothesis, more comfortable.

E1050: **A new sparse PCA procedure via decomposing PCA loss function**
Presenter: **Kohei Adachi**, Osaka University, Japan
Co-authors: Nickolay Trendafilov

Sparse principal component analysis (PCA) refers to any modification of the classic PCA which produces loading matrices containing a huge number of zero entries. Such loading matrices are said to be sparse. We propose a new sparse PCA method for obtaining sparse loading matrix with a specified number of zero elements. In this procedure, a least squares (LS) loss function for PCA is minimized over the PC score and loading matrices. The PC score matrix is constrained to be a column-orthonormal matrix. The loading matrix is constrained to have zero entries which number equals a specified integer. The minimization is carried out by decomposing the loss function into a sum of the two LS functions. One of them is relevant to the PC score matrix only, while the other one is a simple function of the loading matrix and irrelevant to the PC scores. This decomposition allows us to form an alternating LS algorithm which consists of computationally simple steps, whose rationales are easily understood. Finally, we present versions of the algorithm adapted to produce loading matrix with specified number of zeros in each row or column.

E1187: **Maximum likelihood factor analysis with a large number of missing values**
Presenter: **Kei Hirose**, Osaka University, Japan
Co-authors: Sunyong Kim, Yutaka Kano, Miyuki Imada, Manabu Yoshida, Masato Matsuo

The problem of full information maximum likelihood (FIML) estimation in a factor analysis model when a majority of the data values are missing is considered. The expectation–maximization (EM) algorithm is often used to find the FIML estimates, in which the missing values on observed variables are included in complete data. However, the EM algorithm has an extremely high computational cost when the number of observations is large and/or plenty of missing values are involved. We propose a new algorithm that is based on the EM algorithm but that efficiently computes the FIML estimates. A significant improvement in the computational speed is realized by not treating the missing values on observed variables as a part of complete data. Our algorithm is applied to a real data set collected from a Web questionnaire that asks about first impressions of human; almost 90% of the data values are missing. When there are many missing data values, it is not clear if the FIML procedure can achieve good estimation accuracy even if the number of observations is large. In order to investigate this, we conduct Monte Carlo simulations under a wide variety of sample sizes.

ES115 Room B18 BAYESIAN METHODS Chair: Ajay Jasra

E381: **Using the Dirichlet process clustering model to assist the investigation of high order interactions and explore graphical linear models**
Presenter: **Michail Papathomas**, University of St Andrews, United Kingdom
Co-authors: Sylvia Richardson

Detecting high order interactions between covariates in a linear model framework is not straightforward, due to the difficulty in investigating an unwieldy large space of competing models. One approach for reducing the dimensionality of the problem is to create homogenous clusters created using the Dirichlet process, a Bayesian flexible clustering algorithm. We use the fact that variable selection within a clustering algorithm selects covariates that combine to form homogeneous groups of subjects, rather than covariates with a strong marginal signal. The relation between the output from the Dirichlet process and interactions within a linear modelling framework is investigated. A novel approach is proposed where model search algorithms for a large space of graphical linear models are informed by results from a Bayesian partitioning approach implemented in tandem with a variable selection procedure. Relevant theoretical and computational results are discussed.

E844: **Bayesian semiparametric linear mixed: Effects models with normal/independent distributions**
Presenter: **Mauricio Castro**, Universidad de Concepcion, Chile
Co-authors: Victor Lachos, Diana Galvis

Normal/independent distributions are considered as a tool for robust modeling of semiparametric linear mixed-effects models from a Bayesian perspective. The normal/independent distributions is an attractive class of symmetric heavy-tailed distributions that includes the normal, Student$-t$, slash and the contaminated normal distribution as special cases, providing an appealing robust alternative to the routine use of normal distributions for longitudinal data. Using a Bayesian paradigm, an efficient Markov chain Monte Carlo (MCMC) algorithm is introduced to carry out posterior inference. In order to examine the robust aspects of this flexible class against outlying and influential observations we present a Bayesian case deletion influence diagnostics based on the $q-$divergence measures. Further, some discussions on Bayesian model selection criteria are given. The new methodologies are exemplified through simulated and a real data set of AIDS/HIV infected patients.

E1104: **A comparison of variational approximations for fast inference in mixed logit models**
Presenter: **Nicolas Depraetere**, KU Leuven, Belgium
Co-authors: Martina Vandebroek

Recently, variational Bayesian methods have come into the field of statistics. These methods aim to address some of the weaknesses (computation time, storage costs and convergence monitoring) of mainstream MCMC-based inference at the cost of a biased approximation to the posterior distribution of the model parameters. We investigate the performance of a variational approximation in the context of discrete choice data. More specifically, we apply the variational machinery to mixed logit models which is arguably one of the most used models for discrete choice data. A typical treatment of the mixed logit model using the variational methodology is hindered by the fact that the expectation of the so called log-sum-exponential function (log partition function) has no closed form expression. Therefore, one has to resort to approximating or bounding this term. We compare seven different possible bounds or approximations and evaluate their performance in the estimation of the mixed logit model. We found that quadratic bounds to the expected log partition function do not perform particularly well. A recently proposed non-quadratic bound did perform quite well. We also found that the two considered approximations to the expected log partition function performed best in most settings.

E1111: **Approximate Bayesian computation for bilinear processes**
Presenter: **Patricia de Zea Bermudez**, University of Lisbon, Portugal
Co-authors: Kamil Turkman, Antonia Turkman

The class of bilinear models plays an important role in modeling nonlinearity for various reasons, such as the fact that it is a generalization of ARMA models. Under fairly general conditions, bilinear processes approximate finite order Volterra series expansions to any desired order of accuracy over finite time intervals. Volterra series expansion are a dense class within the class of nonlinear time series. Therefore, bilinear processes are also a dense class within nonlinear processes, approximating any nonlinear process to a desired level of accuracy. Due to their capacity of producing clusters of large values, bilinear models are often suggested. However, they are not frequently used in practice due to inference problems. ABC algorithms arise as ways to deal with problems associated with likelihood functions which are analytically difficult to handle or even intractable. The challenge is to find a set of statistics capable of representing the nonlinear dynamics of the system. Seven statistics are suggested, namely a portmanteau statistic that captures the linear time dynamics, a tail index estimator, which quantifies the tail heaviness and the extremal index that measures the degree of clustering of large values. These strategies are applied to several bilinear models for different i.i.d. innovation processes.

ES46 Room G16 CONTRIBUTIONS TO STATISTICAL ALGORITHMS AND SOFTWARE IN R Chair: Cristian Gatu

E222: **GEE for longitudinal ordinal data: Comparing R-repolr, R-ordgee, SAS-GENMOD, SPSS-GENLIN**
Presenter: **Nazanin Nooraee**, University Medical Center Groningen, Netherlands
Co-authors: Geert Molenberghs, Edwin van den Heuvel

To understand the relationship between covariates and outcome, generalized estimating equations (GEE) can be applied. Different approaches have been introduced to extend the GEE for ordinal outcomes and they have been implemented in statistical software. Comparisons of the implemented algorithm in the most frequently used packages (SAS-PROC GENMOD, SPSS-GENLIN, R-geepack and R-repolr) are provided using a simulation study. Multivariate logistic distributions with moderate to highly correlation are considered as the latent variable of ordinal outcomes. All the simulation is performed in the package copula in R. The performance of GEE is assessed on the bias and coverage probability of the parameter estimations and on the number of numerical convergence issues. The simulation study indicates substantial bias in the parameter estimates and numerical issues for data sets with small number of subjects. The unstructured correlation matrix requires larger numbers of subjects than independence and exchangeable correlation matrices to reduce the bias and numerical issues. The coverage probabilities of the parameters are satisfactory for the independence and exchangeable correlation matrix, but they are frequently liberal for the unstructured option. Based on the performance and the available options, preference is given to SPSS and R-repolr.

E886: **Biomarker selection for omics data**
Presenter: **Ron Wehrens**, Fondazione Edmund Mach, Italy
Co-authors: Pietro Franceschi

The R package BioMark provides several tools to define which variables are associated with class differences in data from fields like metabolomics and proteomics. The first group of tools uses Higher Criticism to define an optimal threshold between interesting and non-interesting variables. This can be applied to any statistic, be it a t value, a regression coefficient or something else, and is related to the expected distribution of p values under the null distribution. The second group of tools is based on stability selection, i.e. an assessment of how often specific variables are highlighted as interesting under perturbation of the data. This approach is especially attractive when the number of samples is larger than, say, ten per group. Also in this case, the strategy can be applied to any type of statistic. Using real and simulated data, the application and usefulness of these techniques will be shown.

E898: **NHPoisson: An R package for fitting and validating nonhomogeneous Poisson processes**
Presenter: **Ana C Cebrian**, University of Zaragoza, Spain
Co-authors: Jesus Abaurrea, Jesus Asin

NHPoisson is an R package which provides an assembly of tools for all the steps involved in data modeling using nonhomogeneous Poisson processes in one dimension. It includes functions for data preparation, maximum likelihood estimation, covariate selection and inference based on asymptotic distributions and simulation methods. It also provides specific methods for the estimation of Poisson processes resulting from a peak over threshold approach. The package also contains a wide tool kit for validation analysis, including several types of residuals and diagnostic techniques adapted from other statistical models. It supports functions for generating nonhomogenous Poisson process trajectories. This package can be useful for anyone interested in modeling data using one-dimensional nonhomogeneous Poisson processes in any field.

E1099: **Nonparametric inference for controlled branching processes: Expectation-maximization algorithm**
Presenter: **Carmen Minuesa**, University of Extremadura, Spain
Co-authors: Miguel Gonzalez, Ines del Puerto

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 behaviour of these populations is strongly related to the main parameters of the offspring distribution. In practice, these values are unknown and their estimation is necessary. Usually it must be observed the whole family tree up to a given generation in order to estimate the offspring distribution. In this work, we deal with the problem of estimating the main parameters of the model assuming that the only observable data are the total number of individuals and progenitors in each generation. We set out the problem in a nonparametric framework and obtain the maximum likelihood estimator of the offspring distribution using the expectation-maximization algorithm. Finally, we show the accuracy of the algorithm by way of a simulated example.

ES121 Room B36 CONTRIBUTIONS IN TIME SERIES ANALYSIS I Chair: Marc Hallin

E279: **Model-free detection of structural breaks in time series**
Presenter: **Alexandra Piryatinska**, San Francisco State University, United States
Co-authors: Boris Darkhovsky

Long time series are often generated by different mechanisms (probabilistic, deterministic, or mixed). In economic time series, the time when the underlying generating mechanism changes is called the moment of structural break (MSB). Traditionally, parameters of the models are used to detect the MSB. However, frequently no plausible model is known. In such cases, a novel, model-free approach for detection of the MSB is proposed. It is based on the concept of ε-complexity of continuous functions, and the non-parametric method of change-point detection. In particular, it was established that ε-complexity of a function satisfying the Hölder condition can be completely characterized by a pair of real numbers which are called ε-complexity coefficients. Under the assumption that structural breaks lead to a change of the expected values of the ε-complexity coefficients, a non-parametric method to estimate moments of such changes is proposed. Simulations results are given.

E949: **The exact quasi-likelihood estimation of time-dependent VARMA models**
Presenter: **Abdelkamel Alj**, Universite libre de Bruxelles, Belgium
Co-authors: Kristjan Jonasson, Guy Melard

The purpose is to describe an algorithm for the evaluation of the exact Gaussian likelihood of a $r-$dimensional vector autoregressive-moving average (VARMA) process of order (p, q), with time-dependent coefficients, including a time dependent innovation covariance matrix. The elements of the matrices of coefficients and those of the innovation covariance matrix are deterministic functions of time and assumed to depend on a finite number of parameters. These parameters are estimated by maximizing the Gaussian likelihood function. The advantages of that approach is that the Gaussian likelihood function can be computed exactly and efficiently. The algorithm is based on the Cholesky decomposition method for block-band matrices. We show that the number of operations as a function of p, q and n is barely doubled with respect to a VARMA model with constant coefficients. We provide a detailed description of the algorithm followed by an example using data from the literature.

E950: **Mixture autoregressive models and financial risk**
Presenter: **Mary Akinyemi**, University of Manchester, United Kingdom
Co-authors: Georgi Boshnakov

The class of mixture autoregressive (MAR) models provides a flexible way to model various features of time series data and is very well suited for exploring financial time series. The MAR models are able to capture many stylised properties of real data, such as multi-modality, asymmetry and heterogeneity. We evaluated one-step ahead out of sample VaR and ES measures for some financial time series based on MAR models with Gaussian and Student-t innovations. For comparison, we did the same using some popular approaches. The results are backtested by examining whether the approach adequately forecasts the expected number of violations, generate independent violations and consequently gives a ES whose violation residuals have zero mean behaviour. We implement both the unconditional Kupiec and conditional Christoffersen coverage tests for the correct number of exceedances and independence of these exceedances. We find that the approach based on the MAR models perform better overall. Our results suggests that the MAR models are well suited to capture the kind of data dynamics present in financial data and provide a useful alternative to other approaches.

E1020: **Detection of structural changes using an information causality measure**
Presenter: **Dimitris Kugiumtzis**, Aristotle University of Thessaloniki, Greece

Recently, an information based measure was developed for Granger causality called partial mutual information from mixed embedding (PMIME). PMIME has the advantage of being able to detect direct causal effects in the presence of many inter-related variables, and allows for the formation of complex networks from multivariate time series. We first show the appropriateness of PMIME for multi-variate time series analysis comparing it with other known Granger causality measures (conditional Granger causality index and partial transfer entropy). Then we employ PMIME for the detection of structural changes in two applications, major financial stock indices, and multi-channel scalp electroencephalograms at epileptic activity.

ES96 Room B02 CONTRIBUTIONS TO GRAPHICAL MODELS Chair: Daniel Vogel

E584: **Tree entropy for variable selection**
Presenter: **Klea Panayidou**, University of Frederick, Cyprus

The aim is to examine the problem of exploiting associations between binary variables in a set and selecting a subset of variables for subsequent inference. A well-known method for fitting tree distributions is used. Consistency of the tree estimator is investigated under correct model specification and under misspecification. Tree entropy, a measure for selecting a subset of variables with maximum predictive power, is introduced and justified within a decision theoretic framework. A greedy approach is used for the optimization problem and the solution found is evaluated. The proposed methods are illustrated with experiments on genetics data.

E1003: **Social network analysis and Bayesian networks: Taxonomy of environmental contrasts**
Presenter: **Maria Caterina Bramati**, Sapienza University of Rome, Italy

The main is to connect the Social Network Analysis descriptive measures of centrality to a probabilistic approach to networks. A latent conditional structure of the network which shapes both the structural information and uncertainty characterizing the data is assumed. This is achieved by means of a Bayesian Network approach which allows reweighting the original adjacency matrix according to the latent network structure. Therefore, traditional SNA measures for network description can be easily computed. This is illustrated using a dataset on Issue Correlates of War (ICOW) which contains claims in world politics for three dimensions of interests, land, river and sea for about 244 claims located in 5 world regions and observed from 1816 to 2001. Each node represents a country involved in single or multiple disputes over the use of natural resources. Thus each dyad is represented by an edge connecting the countries involved in the contrast. Each observed edge is a realization of random variables which are connected by a conditional latent structure. As an example, we consider the duration, the geographical location and the issue of the disputes as main variables characterizing the conflicts. Then, the statistical model provides a pictorial representation of dependence relations between variables by means of modular diagrams, namely Directed Acyclic Graphs (DAGs). The presence of an edge joining nodes in the DAG is interpreted as statistical dependence between the variables. The network is 'learned' from data on international conflicts and the arc strengths are used as scores for building a new 'observational' adjacency matrix.

E994: **Bayesian estimation of multiple graphical models**
Presenter: **Chris Oates**, University of Warwick, United Kingdom
Co-authors: Sach Mukherjee

Graphical models are widely used to make inferences concerning interplay in multivariate systems. In many applications, data are collected from multiple related but non-identical units whose graphs may differ but are likely to share many features. Several methods have recently been developed for jointly learning the structure of multiple (related) graphical models under an exchangeability assumption. Here we present a hierarchical Bayesian formulation which enjoys several advantages over existing techniques: (i) complex, non-exchangeable relationships between units are permitted; (ii) a novel nonparametric prior for regularisation between graphs is proposed; (iii) for time series data, an efficient, exact algorithm is provided. We present theoretical and empirical results which demonstrate gains with respect to (i) independent estimation; (ii) estimators based on exchangeability (in non-exchangeable settings).

ES126 Room G15 CONTRIBUTIONS IN STATISTICAL MODELLING Chair: Eric Beutner

E920: **Monte Carlo modified likelihood for panel data models**
Presenter: **Nicola Sartori**, University of Padova, Italy
Co-authors: Ruggero Bellio, Claudia Di Caterina
Modified profile likelihood methods may be effectively applied to estimate the structural parameters of econometric models for panel data, with a remarkable reduction of bias with respect to ordinary likelihood methods. The implementation of the modified profile likelihood require the computation of an expected value of derivatives of the loglikelihood with respect to the fixed effects. Although such a quantity is straightforward to obtain analytically in many important instances, it may be cumbersome to compute it in general models, in particular with dependent data. The proposal here is to approximate the required expected value by Monte Carlo simulation. Simulation studies in dynamic models, for both continuous and discrete data, show the good behavior of the inference based on this Monte Carlo modified profile likelihood. The proposed estimation method is very general and can be easily implemented in R, as done in the package panelMPL.

E965: **Approximate maximum likelihood estimation of the autologistic model**
Presenter: **Marco Bee**, University of Trento, Italy
Co-authors: Giuseppe Espa, Diego Giuliani
Approximate Maximum Likelihood Estimation (AMLE) is a simple and general method recently proposed for approximating MLEs without evaluating the likelihood function. In particular, the only requirement is the ability to simulate the model to be estimated. Thus, the method is quite appealing for spatial models because it does not require evaluation of the normalizing constant, which is often computationally intractable. In the class of Markov random fields, the autologistic model is of paramount importance because it has a wide range of applications to both spatial statistics and statistical mechanics. The most commonly used estimation technique, namely Maximum Pseudo-Likelihood (MPL) is notoriously inefficient. An AMLE-based algorithm for parameter estimation of the autologistic model is proposed. The impact of the numerical choice of the input parameters of the algorithm is studied by means of extensive simulation experiments, and the outcomes are compared to those obtained with different approaches. Results show that AMLE is much more precise, in terms of MSE, with respect to MPL, and comparable to existing ML-type methods. Although the computing time required by AMLE is non-negligible, the implementation is easy, a feature strictly related to its conceptual simplicity.

E1105: **Causal effect of the degree programs on the work path of the graduates in the multivariate latent Markov model**
Presenter: **Fulvia Pennoni**, University of Milano-Bicocca, Italy
Co-authors: Francesco Bartolucci, Giorgio Vittadini
A model is developed for the evaluation of the casual effect of the degree on the work path, when this path is seen as the manifestation of the human capital (HC) represented by a sequence of latent variables. The path is observed in terms of three response variables: type of contract, skill level of the job, and earning level. The sequence of HC latent variables gives rise to a latent process that is assumed to follow an hidden Markov chain. In order to asses an effect and following the recent developments in causal inference, we propose to integrate the inverse probability of treatment weighting in this framework. Background covariates are used to asses balances among the groups of students with different degrees. The resulting multivariate latent Markov model is fitted by a maximum likelihood procedure by using the EM algorithm. Standard errors for the parameter estimates are obtained by a nonparametric bootstrap method. The proposed approach is applied to the analysis of data deriving from administrative archives concerning the labour market. Through this application, we also show the advantages of the proposed approach to study the HC development with respect to other previous approaches.

E1011: **Using data cloning for estimating a Bradley-Terry model at the presence of informative cluster size**
Presenter: **Anna Gottard**, University of Firenze, Italy
Data Cloning is a procedure to compute maximum likelihood estimates and the Fisher information matrix, by utilizing instrumentally the Bayesian paradigm and MCMC procedures. It is useful in some cases of intractable likelihood function. In some applications, ranking sports teams and players' performance is of interest. The Bradley-Terry (BT) model has been utilized for paired comparisons in these contexts. The model assumes that the probability of a victory for a player i on a player j depends on the players' abilities, measured by specific parameters of the model. In including random effects, this model results in a multiple membership model, with a non-hierarchical structure of the random effects. Its likelihood function is analytically intractable. Tennis tournaments are knockout tournaments, with players competing head-to-head, the winners advancing to the next round and the losers being eliminated from the tournament. Consequently, the number of times a competitor plays at a tournament, related to cluster size in a random effect BT model, depends on his/her ability. We propose to jointly model the probability of victory of each contest and the number of times each contender plays at each tournament. Maximum likelihood estimates are obtained via Data Cloning.

ES108 Room B20 CONTRIBUTIONS ON HIGH-DIMENSIONAL STATISTICS Chair: Richard Samworth

E941: **Nonparametric independence screening and structural identification for ultra-high dimensional longitudinal data**
Presenter: **Toshio Honda**, Hitotsubashi Univeristy, Japan
Co-authors: Ming-Yen Cheng, Jialiang Li, Heng Peng
Ultra-high dimensional longitudinal data are increasingly common and the analysis is challenging both theoretically and methodologically. We offer a new automatic procedure in hunting for a sparse semivarying coefficient model, which has been widely accepted for modeling longitudinal data. Our proposed method first reduces the number of covariates to a moderate order by employing a screening procedure, and then identifies both the varying and non-zero constant coefficients using a group SCAD estimator, which is then refined by accounting for the within-subject correlation. The screening procedure is based on working independence and B-spline marginal varying coefficient models. Under weaker conditions than those in the literature, we show that with high probability only irrelevant variables will be screened out and the number of variables left can be bounded by a moderate order, thus the desirable sparsity and oracle properties of the subsequent variable selection step is allowed based on the group SCAD estimator. We also suggest ways to implement the method and to select the tuning parameters.

E971: **Sparse linear discriminant analysis in high-dimensions: A review**
Presenter: **Tsegay Gebrehiwot Gebru**, The Open university, United Kingdom
Co-authors: Nickolay T. Trendafilov
With the recent development of new technologies, high-dimensionality is becoming a common problem in various disciplines for data analysis. Many methods have been proposed for dimension reduction in order to analyze high dimensional data. It is well known that the classical linear discriminant analysis (LDA) is one of the dimension reduction methods when the observations are grouped. LDA performs well when the number of observations to be classified is much larger than the number of variables used for classification. However, in the high dimensional setting (i.e., when the number of variables is much larger than the number of observations) the LDA fails to perform classification effectively. This is mainly because of the existence of redundant variables in the data. As a result, many researchers proposed various methods to circumvent this difficulty. The recently proposed methods for sparse LDA are briefly reviewed. Some methods extend the classic LDA and focus on the regularization of the within-group covariance matrix. Others try to employ different classification strategies which do not involve within-group covariance matrix, e.g.

based on classification error, etc. The common feature of all methods is that they obtain sparse discriminant functions, and thus select only few of the original variables. Finally, we briefly discuss the potential problems in the existing methods, and possible alternatives for their improvement.

E1126: **GRID for variable selection in high dimensional regression**
Presenter: **Francesco Giordano**, University of Salerno, Italy
Co-authors: Soumendra Nath Lahiri, Maria Lucia Parrella

In the context of nonparametric regression, one assumes that the number of covariates tends to infinity but only some of these covariates are relevant for the model. The aim is to identify the relevant covariates and to obtain some information about the structure of the model. A new nonparametric procedure is proposed, called GRID (Gradient Relevant Identification Derivatives). The key idea is to use a modified local linear estimator in a non standard way together with Empirical Likelihood method to make variable selection without any additivity assumption. Under some regularity conditions, GRID automatically identifies the relevant covariates of the regression model, also distinguishing the nonlinear from the linear ones (a covariate is defined linear or nonlinear depending on the marginal relation between the response variable and such a covariate). Besides, the interactions between the covariates are automatically identified, without the necessity of considering some kind of stepwise selection method. In particular, GRID can identify the mixed terms of any order without increasing the computational complexity of the algorithm. Finally, the procedure is completely data-driven, so it is easily implementable for the analysis of real datasets. Some theoretical results are shown for the proposed procedure.

ES131 Room B29 CONTRIBUTIONS TO OPTIMAL DESIGNS Chair: Alexander Donev

E970: **Optimal experimental designs for nonlinear analysis: Solving the conundrum**
Presenter: **Jose M Vidal-Sanz**, Universidad Carlos III de Madrid, Spain
Co-authors: Mercedes Esteban-Bravo, Agata Leszkiewicz

To reduce the experimental cost and respondents' fatigue, optimal experimental designs maximize the information elicited from the respondent, or equivalently minimize the estimator variance. However, in many models the variance depends on the unknown regression parameters β. Therefore we cannot optimally design the experiment because its efficiency depends on parameters to be estimated from the data. Previous literature dealt with this puzzle by imposing assumptions on the unknown parameters: (1) choosing an arbitrary vector of parameters β supposedly applying 'prior knowledge'; or (2) postulating a probability distribution for β over the parametric space hopefully concentrated around the true value. Therefore, the design is efficient only if these assumptions are correct. Little is known about the robustness of the design, when the true parameters deviate from the assumed values. Moreover, if we knew the value of true parameters, there would be no reason to do the experiment in the first place. We propose a general approach to compute optimal conjoint designs in problems in which the covariance matrix depends on the unknown parameter. We solve this problem using efficient computational methods for robust optimization, and provide numerical examples for discrete-choice experiments comparing our approach and the classical methods.

E1180: **Approximate Bayesian adaptive optimal design for computer experiments**
Presenter: **Noha Youssef**, American University in Cairo, Egypt

It is very common to model the computer experiment output using a Gaussian process model. The Gaussian process model consists of two parts; a fixed part (the mean) and a random part. If the hierarchical structure is used with the Gaussian process model then a hyperprior distribution is assigned for the hyperparameters of the prior distribution of the mean parameters. The Wishart distribution is a suitable prior distribution for the covariance hyperparameters of the mean parameters however computing the posterior distribution is not tractable. In addition, sequentially selecting an optimal design such as the maximum entropy sampling design relies on obtaining the predictive distribution for the model output at every single step. Two alternatives are proposed here to tackle the intractability problem for both the posterior and the predictive distributions. The first is to create an approximate criterion that does not depend directly on the predictive distribution and the second is to choose a good approximation method such as the Approximate Bayesian Computation method to approximate the posterior and the predictive distributions.

E1241: **Optimal design for correlated processes with input-dependent noise**
Presenter: **Alexis Boukouvalas**, Aston University, United Kingdom
Co-authors: Dan Cornford, Milan Stehlik

Optimal design for parameter estimation in Gaussian process regression models with input-dependent noise is examined. The motivation stems from the area of computer experiments, where computationally demanding simulators are approximated using Gaussian process emulators to act as statistical surrogates. In the case of stochastic simulators, which produce a random output for a given set of model inputs, repeated evaluations are useful, supporting the use of replicate observations in the experimental design. The findings are also applicable to the wider context of experimental design for Gaussian process regression and kriging. Designs are proposed with the aim of minimising the variance of the Gaussian process parameter estimates. A heteroscedastic Gaussian process model is presented which allows for an experimental design technique based on an extension of Fisher information to heteroscedastic models. It is empirically shown that the error of the approximation of the parameter variance by the inverse of the Fisher information is reduced as the number of replicated points is increased. Through a series of simulation experiments on both synthetic data and a systems biology stochastic simulator, optimal designs with replicate observations are shown to outperform space-filling designs both with and without replicate observations. Guidance is provided on best practice for optimal experimental design for stochastic response models.

E977: **Maximin efficient designs for estimating the interesting part of a dose-effect curve**
Presenter: **Frank Miller**, Stockholm University, Sweden
Co-authors: Ellinor Fackle Fornius, Hans Nyquist

As the costs of clinical studies increase, the demand for more efficient designs also increases. Therefore, there is a growing interest in introducing designs that optimize precision in clinical studies. Unfortunately, optimal designs generally require knowledge of unknown parameters. We consider the maximin approach to handle this problem. A maximin efficient design maximizes the efficiency when compared to a standard design, as the parameters vary in a specified subset of the parameter space. Maximin efficient designs have shown to be numerically difficult to construct. However, a new algorithm, the H-algorithm, considerably simplifies the construction of these designs. We exemplify the maximin efficient approach by considering an Emax- sigmoid model describing a dose-response relationship and compare inferential precision with that obtained when using a uniform design. In a first approach to construct a maximin efficient design we specify a number of possible scenarios, each of which describing a possible shape of the dose-response relation. The design obtained is shown to be at least 15 percent more efficient than the uniform design. It is then shown that the obtained design is maximin efficient also for a much larger parameter set defined by parameter values between those specified by the initial scenarios.

Monday 16.12.2013 16:20 - 17:40 Parallel Session R – CFE

CS110 Room B36 CONTRIBUTIONS IN BAYESIAN ECONOMETRICS II **Chair: Marek Jarocinski**

C039: **A new measure of the output gap for the EU countries: A state space approach to productivity and efficiency measurement**
Presenter: **Camilla Mastromarco**, University of Salento - Lecce, Italy

The output gap is an important indicator of capacity utilization in an economy and represents an important measure for policy makers who try to target stability and growth. Estimating the output gap is difficult because it is not directly observable. Based on a modification of the Kalman filter prediction equations, we propose a measure for the output gap based on efficiency frontier estimation. We estimate efficiency and total factor productivity for 15 EU economies in a Bayesian framework. The state space approach allows us to estimate a stochastic frontier model from univariate time series. This is an advantage in a time series context, when data quality and comparability is an issue. Moreover, it is possible to model inefficiency as an autoregressive process. The technical efficiency measure can be interpreted as output gap. We find evidence for the well known productivity slowdown, as well as an increase in the output gap over the entire observation period.

C737: **Inference for nonlinear exchange rate dynamics**
Presenter: **Edouard Djeutem**, Simon Fraser University, Canada

Sequential Monte Carlo method is used to estimate a nonlinear monetary model of exchange rate. In our model, the exchange rate dynamics is driven by fundamentals and structural parameters of unknown form. The fundamentals are assumed to follow a time varying autoregressive process of order one. As a consequence, this model features stochastic volatility and thus particle filtering offers an advantage over other inference methods. We estimate the maximum likelihood, construct confidence interval, and perform residual analysis of our state space form model. The estimated model captures most of the nonlinear dynamics of exchange rate. Furthermore, our analysis shows that the parameter instability does not account for most of the variability in exchange rate.

C981: **Bayesian analysis of dynamic effects in inefficiency: Evidence from the Colombian banking sector**
Presenter: **Jorge Galan**, University Carlos III de Madrid, Spain
Co-authors: Helena Veiga, Michael Wiper

Firms face a continuous process of technological and environmental changes that requires them to make managerial decisions in a dynamic context. However, costs and other constraints prevent firms from making instant adjustments towards optimal conditions and may cause inefficiency to persist in time. We propose, under a Bayesian framework, a flexible dynamic model that captures differences in the adjustment costs among firms and distinguishes dynamic effects in the inefficiency from firm inefficiency heterogeneity. The new model is fitted to a ten year sample of Colombian banks. Our findings suggest that separating firm characteristics associated with size and foreign ownership from the dynamics of inefficiency improves model fit, while acquisitions are found to affect the dynamic component. Overall, Colombian banks present high inefficiency persistence but important differences between institutions are found. In particular, merged banks present low inefficiency persistence which allows them to recover rapidly the efficiency losses derived from merging processes.

C1088: **Identification and estimation of structural breaks in duration data**
Presenter: **Rasika Yatigammana**, University of Sydney, Australia
Co-authors: Richard Gerlach

Various shocks on an economy impact related financial time series, which usually results in its properties exhibiting changes over time. Structural breaks, time-varying parameters, regime switching, etc are frequently used in modelling such shifts in behaviour. It is acknowledged that structural breaks in duration data may significantly change forecasting and trading strategies. In the presence of structural breaks in the data generating process, not allowing for them in model specification could lead towards misspecification creating erroneous results. Therefore a Bayesian approach is used to detect structural breaks in adjusted duration data analysed in the form of Autoregressive Conditional Duration (ACD) models. A Gaussian mixture proposal within a Markov Chain Monte Carlo framework was used in the estimation process. Considering optimality, an alternative efficient algorithm is investigated. Here, the proposal or candidate density constructed as a mixture of Student's-t distributions is employed in an independent chain Metropolis-Hastings algorithm to draw values in sampling period. Structural break estimation is an integral part of the process. After an extensive simulation study on lengthy multiple series, structural breaks in actual trade data is estimated. Finally a series of generated forecasts using the estimated model is compared with the existing literature.

CS122 Room G16 TIME-VARYING PARAMETERS IN FINANCE AND ECONOMICS II **Chair: Cristina Amado**

C142: **A unified theory for time-varying models: Foundations and applications in the presence of breaks and heteroskedasticity**
Presenter: **Menelaos Karanasos**, Brunel University, United Kingdom
Co-authors: Alekos Paraskevopoulos, Stavros Dafnos

An elegant approach to examine the dynamics of stochastic time series models with time dependent coefficients is developed. We provide the closed form of the fundamental solutions for time-varying autoregressive moving average models which is a long standing research topic. This enable us to characterize these models by deriving i) its multistep ahead predictor; ii) the first two unconditional moments; and iii) its covariance structure. In addition, capitalizing on the connection between linear difference equations and the product of companion matrices, we employ our general methodology to obtain an explicit formula for the latter. We also apply our method to obtain results on generalized continuant matrices. To illustrate the practical significance of our results we consider autoregressive models with multiple breaks and also apply our unified approach to a variety of processes such as i) periodic, cyclical and smooth transition autoregressive models, ii) time-varying generalized autoregressive conditional heteroscedasticity specifications, and iii) generalized random coefficients autoregressive models.

C1114: **The SVaR: Modeling state space of value at risk**
Presenter: **Diogoye Faye**, University of Montpellier, France

The RiskMetrics model developed by the bank JP Morgan following the amendment of Basel accords 1988 was erected as a measure of financial risk to deal with important disturbances affecting international banking markets. Commonly known as Value at Risk, it was accepted by all bodies and financial institutions to be a coherent risk measure. Despite its popularity, it is the subject of many controversies. Indeed, the estimation parameters of RiskMetrics are assumed to be fixed over time, which is contrary to the characteristics of financial markets. Two reasons are used to justify temporal instability: a) Due to the presence of heterogenous agents the VaR is not analysed by focusing on a single temporal dimension but rather on trading frequencies; b) The structure of financial time series which is usually affected by the crash bubble phenomenons and so on. These can be considered as hidden variables that we must take into account in the risk assessment. To provide a solution to this problem, we propose respectively WVaR models and an application of the kalman filter on VaR. This filter recursively corrects by its functions the parameters of time. We define also a risk measure called SVaR: It's VaR obtained by updating estimation parameters.

C1261: **Investigation of institutional changes in the UK housing market by structural break tests and time varying parameter models**
Presenter: **Hanxiong Zhang**, Newcastle University, United Kingdom
Co-authors: Robert Hudson, Hugh Metcalf
The purpose is to investigate empirically the institutional changes in the UK housing market and to contribute to the literature from two aspects. Firstly, the dates of structural breaks appear to match remarkable market shocks rather than political events. It seems that an unexpected shock, in particular the financial crisis, caused an overreaction by the public and thereby led to a structural break. Secondly, the slow-moving institutional changes are assessed by attempting three novel Time Varying Parameter models. We find that people's expectations, in particular the expected capital gains on the housing market rather than the traditional economic factors, are playing a far more important role in driving the UK housing prices.

CS125 Room B30 CONTRIBUTIONS TO RISK MODELING AND MANAGEMENT II Chair: Rustam Ibragimov

C108: **The asymmetry of accounting information: Problems of missing data**
Presenter: **Sami Ben Jabeur**, IESEG School of management, France
Co-authors: Rabai Belhaj Hassine, Fahmi Youssef
In a context of questioning the quality of accounting and financial information published by companies, we tried to use the technique of PLS regression to overcome the problem of missing data in a database accounting in order to better predict risk of failure. This statistical approach is applied to a sample of French SMEs in the period 2006 to 2008 based on a number of accounting and financial ratios.

C672: **Implied risk exposures**
Presenter: **Sylvain Benoit**, University of Orleans, France
Co-authors: Christophe Hurlin, Christophe Perignon
It is shown how to reverse engineer banks' risk disclosures, such as Value-at-Risk, to obtain an implied measure of their exposures to equity, interest rate, foreign exchange, and commodity risks. Factor Implied Risk Exposures (FIRE) are obtained by breaking down a change in risk disclosure into an exogenous volatility component and an endogenous risk-exposure component. In a study of large US and international banks, it is shown that (1) changes in risk exposures are negatively correlated with market volatility and (2) changes in risk exposures are positively correlated across banks, which is consistent with banks exhibiting herding behavior in trading.

C856: **Using robust principal component analysis to define an early warning index of firm's over-indebtedness and insolvency**
Presenter: **Giuseppina Damiana Costanzo**, Universita della Calabria, Italy
Co-authors: Marianna Succurro, Damiano Bruno Silipo
The aim is to define an early warning index for firms' insolvency which is very important both for potential investors, management, stockholders, actual or potential firm's competitors. An early warning index could signal a critical level of over-indebtedness behind which the financial status of the firm becomes pathological, therefore very difficult to rehabilitate. The main steps in the analysis of the firms' over-indebtedness deal with: 1. The definition of a set of variables including several aspects of the indebtedness phenomenon (leverage, indebtedness capacity, form of the financial debt, net financial position, etc.); 2. The setting up of criteria which allow us to establish when a firm may be considered over-indebted. We built up a debt index which includes eleven financial ratios, while to measure a firm's debt sustainability we built up a second index including three ratios related to the capability of the firm to meet financial obligations with current income. On the base of accounting data taken from the Amadeus database, we estimate the previous indices by considering a Robust Principal Component Analysis method. We then compare our early warning index to some other well-known indices as, for example, the z-score.

C962: **New models for financial emerging market risk**
Presenter: **Wafa Snoussi**, University of Tunis, Tunisia
Co-authors: Mhamed-ali El-Aroui
The estimation of Value-at-Risk (VaR) using standard methods is poorly adapted to financial emerging markets. The specific microstructure criteria of emerging markets such as low liquidity, very pronounced asymmetric information and overpredictability affect the VaR estimation methods. We propose to adjust the McNeil and Frey model (MF) to these specifications by developping new VaR estimation methods: MF_{liq}, MF_{asy}, MF_{pred} and MF_{comb}. In MF_{liq} method a liquidity risk component is added to the AR(1)-GARCH(1,1)-GPD VaR model by adding a spread model. The backtesting empirical results show that in tunisian market the VaR estimation error has been reduced by about 53%. In the MF_{asy} method a AR(1)-EGARCH(1,1)-GPD model is introduced to take into account the asymmetric information. In average MF_{asy} reduces the VaR estimation errors by about 35%. ARFIMA(1,d,1)-GARCH(1,1)-GPD modelling is used in MF_{pred} to consider the returns overpredictability. MF_{pred} improves the VaR estimation in the tunisian market. Finally, MF_{comb} is developped to combine the previous methods and take into account all the emerging markets specifications. Using conditional and unconditional coverage tests to choose the appropriate model, empirical results show that new MF_{comb} method outperforms the standard MF model in both 'normal' and 'crises' period for 95%, 97.5% and 99% confidence levels.

CS74 Room B33 MODELLING REGIME CHANGES III Chair: Willi Semmler

C877: **The effect of wage moderation on output and employment**
Presenter: **Matthieu Charpe**, International Labour Organization, Switzerland
Co-authors: Stefan Kuhn
A two country DSGE model is presented with search and matching to assess the effects on output and employment of a decline in the bargaining power of workers. Lower real wages have ambiguous effects on domestic and foreign output in the presence of a transmission channel linking real wages with consumption and saving decisions of households. This transmission channel takes place via credit constraint/rule-of-thumb households. The main result is that lower real wages have a positive impact on the domestic economy as the competitiveness effect dominates the lower labour income/lower domestic consumption effect. Lower real wages are however beggar-thy-neighbour in the open economy case whether the exchange rate regime is fixed or flexible. Lastly, lower real wages are detrimental to both the home country and the foreign country in the presence of a lower zero bound in the nominal interest rate

C947: **Forecasting labour markets using hiring uncertainty**
Presenter: **Ekkehard Ernst**, International Labour Organization, Switzerland
Co-authors: Christian Viegelahn
A forward-looking measure is developed for the macroeconomic uncertainty that employers are confronted with when taking decisions about the size of their workforce. The model that provides the basis for the uncertainty measure interprets hires and lay-offs of workers as an investment into a project with uncertain return and employers decide when to undertake this investment. The theoretically derived measure is then taken to the data, calculating economy-wide and sectoral uncertainty indicators for selected G20 countries from data on hiring demand and unit labour costs. The factors that influence the level of uncertainty are investigated. It is shown that the inclusion of the uncertainty indicator into country-specific forecasting models for unemployment can significantly improve forecasting performance.

C954: **A qualitative response VAR model: Joint dynamics of U.S. interest rates and business cycle**
Presenter: **Henri Nyberg**, University of Helsinki, Finland

A new regime switching vector autoregressive (VAR) model is introduced where the regime switching dynamics is described by a qualitative response (QR) variable. Parameters of the QR-VAR model can conveniently be estimated by the method of maximum likelihood and multiperiod forecasts can be constructed using a simulation-based forecasting method. The model is applied to provide a new characterization of the nonlinear dynamic relationship between the U.S. interest rates and business cycle measured in terms of the NBER recession and expansion periods. A strong bidirectional predictive linkage between the variables is found, and due to the predictability of the business cycle regimes, the QR-VAR model yields superior out-of sample forecasts for the interest rate variables compared with the VAR model.

C1267: **On how many factors and shocks cause financial stress**
Presenter: **Marcus Kappler**, ZEW, Germany
Co-authors: Frauke Schleer

The number of dynamic and static factors that generate co-movement of 150 newly compiled monthly time series on financial market conditions in the euro area is tested. The aim is to assess the dimension of the shocks that drive financial data. A second aim is to construct summary indices on the conditions and level of stress in financial markets with the aid of a dynamic factor model. Our results suggest that the data respond quite differently to fundament shocks to financial markets but the dimension of these shocks is rather limited. With the aid of an exploratory analysis we give the estimated common factors an economic meaning. We find that the presence of a "periphery banking crisis factor", a "security market factor" and a "yield curve factor" explains the bulk of variation in euro area financial data.

CS18 Room B18 FINANCIAL MODELLING Chair: Giuseppe Storti

C076: **Kalman filtering and online learning algorithms for portfolio selection**
Presenter: **Alain Kabundi**, University of Johannesburg, South Africa

The problem of how to select a portfolio of stocks in order to improve the odds of beating, not only the general stock market, but also the best stock, without prior knowledge of the statistical properties of stock prices, continues to attract intensive research work in the machine learning community. Most, if not all, online portfolio selection algorithms take the price relative as the sole input into the machine learning procedure. We use an alternative measure of price relative (relative to a price trend) that is more consistent with portfolio manager's best practice. We prove theoretically that our measure of price relative has well-defined and desirable statistical properties that make them better suited for mean reversion strategies. We find that the statistical evidence of out-of-sample predictability of stock returns is stronger once predictors are adjusted for high persistence. Our approach is evaluated against benchmark online portfolio allocation techniques using six up to date real market datasets. Our methods outperform the benchmark allocation techniques, sometimes in a spectacular way in these datasets without any additional computational demand or modeling complexity.

C120: **A methodological note on the estimation of the "smart money effect"**
Presenter: **Julio Cesar Alonso**, Universidad Icesi, Colombia
Co-authors: Luis Berggrun

The "Smart Money Effect" (SME), or whether funds that receive more money subsequently outperform those that receive less money has received ample attention in the financial literature. The most common approach to estimate the SME is to form portfolios based on past flows and record its performance in a subsequent period. Then the risk-adjusted performance (the estimated intercept or alpha) of a long-short portfolio (long in high past flow funds and short in low past flow funds) is examined to assess the existence of the SME. If there is a SME, the alpha of the long-short portfolio should be positive and significant. To estimate the significance of risk-adjusted performance, an OLS regression is commonly applied along with a Newey-West variance-covariance matrix to account for auto-correlation in the residuals. This approach neglects the possibility of correlated residuals in the estimations of the two portfolios due to common external shocks (e.g. market-wide or macroeconomic shocks). These kinds of shocks could be modeled using a Seemingly Unrelated Regression (SUR) model. We construct a Monte Carlo experiment that compares the statistical size of the SME test for two approaches that control for autocorrelation: i) OLS and ii) SUR. In particular, we study the efficiency gains in the standard errors from using a SUR instead of OLS for different sample sizes. Our preliminary results suggest that the conventional OLS approximation does not perform better than the suggested SUR approach.

C434: **Investment strategies and financing constraints**
Presenter: **Takashi Shibata**, Tokyo Metropolitan University, Japan
Co-authors: Michi Nishihara

The aim is to introduce debt issuance limit constraints along with market debt and bank debt to consider how financial frictions affect investment, financing, and debt structure strategies. The model provides four important results. First, a firm is more likely to issue market debt than bank debt when its debt issuance limit increases. Second, investment strategies are nonmonotonic with respect to debt issuance limits. Third, debt issuance limits distort the relationship between a firm's equity value and investment strategy. Finally, debt issuance limit constraints lead to debt holders experiencing low risk and low returns. That is, the more severe the debt issuance limits, the lower the credit spreads and default probabilities. The theoretical results are consistent with stylized facts and empirical results.

C909: **A multilateral arbitrage-free Nelson-Siegel term structure model for the determination of currency risk premia**
Presenter: **Sarah Mouabbi**, Queen Mary University of London, United Kingdom

Empirical evidence has so far rejected the use of the uncovered interest rate parity to determine exchange rates. Its failure is often accredited to the existence of a time-varying risk premium. An attempt is made to renew the interest in determining the depreciation of exchange rates through interest rate differentials, and ultimately extract the term structure of currency risk premia. The method proposed consists of developing a Multilateral affine Arbitrage-Free class of dynamic Nelson-Siegel term structure models (MAFNS) to obtain the discount rate variations, which in turn are used to derive a representation of exchange rate depreciations. The manipulation of no-arbitrage restrictions allows to endogenously capture currency risk premia. The estimation exercise comprises of a state- space analysis through the Kalman filter. The imposition of the Dynamic Nelson-Siegel (DNS) structure allows for a tractable and robust estimation, offering significant computational benefits, whilst no-arbitrage restrictions enforce the model with theoretically appealing properties.

CS123 Room B35 CONTRIBUTIONS IN TIME SERIES AND PANEL DATA ECONOMETRICS Chair: Mauro Costantini

C211: **An asymptotically UMP test for unit roots in cross-sectionally dependent panels**
Presenter: **Gaia Becheri**, Delft University of Technology, Netherlands
Co-authors: Ramon van den Akker

Just as for univariate time series, the presence or absence of unit roots in a panel data model can have crucial economic policy implications. The asymptotic power envelope for testing the unit root hypothesis is derived for large n and large T heterogeneous panels with cross-sectional dependence. More precisely, we consider a Gaussian heterogeneous panel allowing for incidental intercepts where the cross-sectional dependence is generated by a factor structure. The factors and the idiosyncratic shocks are modeled as linear processes driven by Gaussian innovations. For this model, we derive the asymptotic power envelope by determining the limit experiment (in the Le Cam sense). This limit experiment is shown to be of the Locally Asymptotically Normal (LAN) type, which implies that uniformly most powerful tests for the unit root hypothesis can exist. The structure of the limit experiment is then exploited to construct a new panel unit root test that is asymptotically Uniformly Most Powerful (UMP). A Monte Carlo study shows that the proposed test also exhibits good finite-sample performances.

C824: **Identifying I(0) series in macro-panels: On the usefulness of sequential panel selection methods**
Presenter: **Claudio Lupi**, University of Molise, Italy
Co-authors: Mauro Costantini

Sequential panel selection methods (SPSMs) are based on the repeated application of panel unit root tests and are increasingly used to detect $I(0)$ individual time series in macro-panels. We argue that the panel unit root tests to be used in SPSMs are best based on combinations or averaging of individual Dickey-Fuller t statistics or p values. We check the reliability of SPSMs by using extensive Monte Carlo experiments. In this paper we use an unconventional approach to simulation. Rather than simulating the individual time series, we simulate the t statistics and the p values to be combined in the panel unit root tests, both under the unit root null and under selected local-to-unit root alternatives. The analysis is carried out considering both independent and cross- dependent individual Dickey-Fuller tests. The simulation of the p values under the local-to-unit-root alternative is of interest on its own and poses new computational issues which we address in the paper. Simulation results are summarized using ROC analysis, consistently with the literature on discrete classifiers. We show that SPSMs do not possess better classification performances than conventional univariate Dickey-Fuller tests.

C825: **Modeling regional disparities in the sectoral economic performance: An experimental analysis in Italy**
Presenter: **Riccardo Corradini**, ISTAT, Italy
Co-authors: Filippo Oropallo

Converge of per-capita GDP of Italian regions over the recent published ISTAT dataset starting from 1995 up to now, is still an unexplored issue. Convergence is analyzed using related literature. An updated evidence arises from new data by the notion of co-integration and common trends and the stationarity tests applied to these panel data set series. Due to presence of a short-time series subject to a national accounts constraint a Monte Carlo simulation supports our analysis. In the second part of the paper we analyze the factors that affect regional gaps in productivity growth and in economic performance. Our research reveals the existence of a slow and not very systematic convergence of labor productivity toward a common level, and of an even more uncertain convergence of productivity. For this purpose the use of dynamic panel data models will permit us to explain the factors of the disparities and will help us to foresee the regional trends according the different factors considered in each sector model. Finally, some conclusions concerning our analysis about regional GDP and sectoral analysis are drawn.

C916: **Intraday liquidity, price dynamics and uncertainty in cap-based portfolios**
Presenter: **Paolo Mazza**, Universite catholique de Louvain, Belgium
Co-authors: Mikael Petitjean

It is investigated how informative is price dynamics to estimate contemporaneous intraday liquidity on Euronext for three market capitalization classes: small, mid, and large caps. Liquidity is measured by a comprehensive set of both book-based and trade-based proxies. Price dynamics is captured by studying price movements between high, low, opening, and closing prices. While controlling for trading activity and realized volatility, we estimate the relation between liquidity and price movements by applying OLS with clustered standard errors, robust and median regression techniques. We show that liquidity and price dynamics are closely related. For example, the intensity in the price discovery process and the level of (relative) price uncertainty are clearly associated with lower liquidity. Easy-to-observe price movements are found to be particularly useful when it comes to efficiently evaluating the level of liquidity on the stock market.

CS117 Room G15 COINTEGRATION Chair: Paolo Paruolo

C818: **Forecasting multivariate time series under different short- and long-run co-movement restrictions**
Presenter: **Diogo Saraiva**, Getulio Vargas Foundation, Brazil
Co-authors: Joao Victor Issler, Osmani Guillen, Alain Hecq

It is well known that cointegration between the level of two variables (e.g. prices and dividends) is a necessary condition to assess the empirical validity of a present-value model (PVM) linking them. The work on cointegration, namely on long-run co-movements, has been so prevalent that it is often overlooked that another necessary condition for the PVM to hold is that the forecast error entailed by the model is orthogonal to the past. This amounts to investigate whether short-run co-movements steming from common cyclical feature restrictions are also present in such a system. We test for the presence of such co-movement on long- and short-term interest rates and on price and dividend for the U.S. economy. We focus on the potential improvement in forecasting accuracies when imposing those two types of restrictions coming from economic theory.

C1065: **Estimation and testing for cointegration with threshold effects**
Presenter: **Julio Angel Afonso-Rodriguez**, University of La Laguna, Spain

The asymptotically almost efficient estimation of a single-equation cointegrating regression model is considered with threshold-type nonlinearity through direct application of the recently proposed Integrated Modified-OLS (IM-OLS) estimation method for linear cointegrating regression models. Unlike the difficulties encountered in the application of the existing standard asymptotically efficient estimation methods in the linear setup, such as the FM-OLS, CCR, or DOLS methods, to this nonlinear framework we find that this new estimation method, in addition to its simplicity, allows to obtain satisfactory distributional results under a very general set of assumptions concerning serial correlation in the error correction terms and endogeneity of the stationary threshold variable. By using the first difference of the IM-OLS residuals, we propose a simple testing procedure for the null hypothesis of cointegration with threshold effects against no cointegration. We derive the limiting distribution of this estimator under cointegration with the threshold adjustment component, as well as under no cointegration, and also illustrate its performance in finite samples through a simulation experiment in terms of bias and RMSE. Finally, we present an application to a nonlinear specification of the present value model for U.S. stock prices and dividends.

C230: **Cointegration and structural breaks in the EU debt sovereign crisis**
Presenter: **Nuno Ferreira**, ISCTE-IUL, Portugal
Co-authors: Rui Menezes, Manuela Oliveira
Sights of a sovereign debt crisis spread among financial players started in late 2009 as a result of the growing private and government debt levels worldwide. In July 2010 FITCH lowered the rating of Spain and Italy and the concerns intensified leading European nations to implement a series of financial support measures such as the European Financial Stability Facility and the European Stability Mechanism. In late 2010, Trichet stated that the sovereign debt crisis in Europe reached a systemic dimension. In a context of established crisis, a full battery of unit-root tests was performed looking for structural breaks and cointegration evidence between interest rates and stock market prices using a 13 years' time-window (1999-2011), for six European markets under stress, using US, UK and Germany as benchmark. The results identified the most significant structural breaks at the end of 2010 and rejected consistently the null hypothesis of no cointegration.

C1208: **A wavelet approach to multiple cointegration testing**
Presenter: **Javier Fernandez-Macho**, University of the Basque Country, Spain
A class of cointegration tests is introduced based on estimated low-pass and high-pass regression coefficients from the same wavelet transform of the original time series data. The procedure can be applied to test the null of cointegration in a $n+k$ multivariate system with n cointegrating relationships without the need of either detrending nor differencing. The proposed non residual-based wavelet statistics are asymptotically distributed as standard chi-square with nk degrees of freedom regardless of deterministic terms or dynamic regressors, thus offering a simple way of testing for cointegration under the null without the need of special tables. Small sample quantiles for these wavelet statistics are obtained using Monte Carlo simulation in different situations including I(1) and higher order cointegration cases and it is shown that these wavelet tests exhibit appropriate size and good power when compared to other tests of the null of cointegration.

CS44 Room B34 COPULAS AND ITS APPLICATIONS Chair: Jean-David Fermanian

C875: **Reinvestigating the UIP puzzle via analysis of multivariate tail dependence in currency carry trades**
Presenter: **Matthew Ames**, University College London, United Kingdom
Co-authors: Guillaume Bagnarosa, Gareth Peters
The currency carry trade is a well-known financial puzzle to explain. Given foreign exchange market equilibrium, the interest rate parity condition implies that the expected return on domestic assets should equal the exchange rate adjusted expected return on foreign currency assets. However, it has been shown empirically, that investors can actually on average earn arbitrage profits by borrowing in a country with a lower interest rate, exchanging for foreign currency, and investing in a foreign country with a higher interest rate, whilst allowing for any losses (or gains) from exchanging back to their domestic currency at maturity. The intention is therefore to reinterpret the currency carry trade puzzle in light of heavy tailed marginal models coupled with multivariate tail dependence features. We analyse the risk-reward for currency portfolios constructed using: a) interest rate differentials, b) stochastic ordering, c) multivariate spearman's rank correlation. To achieve this analysis of the multivariate extreme tail dependence we develop several parametric models and perform detailed model comparison. It is thus demonstrated that tail dependencies among specific sets of currencies provide other justifications to the carry trade excess return and also allow us to detect construction and unwinding periods of such carry portfolios.

C1036: **Recognizing and visualizing copulas: An approach using local Gaussian approximation**
Presenter: **Baard Stoeve**, University of Bergen, Norway
Co-authors: Geir Drage Berentsen, Dag Tjostheim, Tommy Nordbo
Copulas are much used to model non-linear and non-Gaussian dependence between stochastic variables. Their functional form is determined by a few parameters, but unlike a dependence measure like the correlation, these parameters do not have a clear interpretation in terms of the dependence structure they create. We examine the relationship between a newly developed local dependence measure, the local Gaussian correlation, and standard copula theory. We are able to describe characteristics of the dependence structure in different copula models in terms of the local Gaussian correlation. In turn, these characteristics can be effectively visualized. More formally, the characteristic dependence structure can be used to construct a goodness-of-fit test for bivariate copula models by comparing the theoretical local Gaussian correlation for a specific copula and the estimated local Gaussian correlation. An essential ingredient of this test is the use of a canonical local Gaussian correlation and Gaussian pseudo-observations which make the test independent of the margins, so that it is a genuine test of the copula structure. A Monte Carlo study reveals that the test performs very well compared to a commonly used alternative test. We also propose two types of diagnostic plots which can be used to investigate the cause of a rejected null. Finally, our methods are applied to a classical insurance data set.

C1225: **Revaluation of estimated option prices using GARCH processed with most preferable properties**
Presenter: **Jegors Fjodorovs**, Riga Technical University, Latvia
Co-authors: Andrejs Matvejevs
The possibility of identifying nonlinear time series using nonparametric estimates of the conditional mean and conditional variance hs been studies previously. One of the main problems is the development of time series $\{x(t), t \in Z\}$ methods of analysis through regression models even without knowing the regression function. The aim is to deal with the estimation of copula-based semi parametric models for random processes with usual kind of information about the distribution law. These models are characterized by conditional heteroscedasticity and have often been used in modelling the variability of statistical data. The basic idea is applied to a local linear regression with squared residuals for finding the unknown function.This approach has applied characteristics in finance. For example it is option pricing taking into account nonlinear effect and asymmetry in errors of a model for simulation of an underlying.

C197: **Dynamic D-Vine copula model with applications to value-at-risk**
Presenter: **Flavio Ziegelmann**, Universidade Federal do Rio Grande do Sul, Brazil
Co-authors: Paula Tofoli, Osvaldo Silva Filho
Vine copulas constitute a very flexible class of multivariate dependence models. Built on bivariate copulas, they can match any possible dependence structure. Research on vine copulas applied to financial data is focused mostly on the case of time-homogeneous dependence structures. However it goes against evidence found in the literature which suggests that this dependence is not constant over time. We allow the dependence parameters of the pair-copulas in a D-vine decomposition to be potentially time-varying, following a nonlinear restricted ARMA(1,m) process in order to obtain a very flexible dependence model for applications to multivariate financial data. The proposed model is validated in numerical simulations and further evaluated with respect to the accuracy of Value-at-Risk (VaR) forecasts in crisis periods. The Monte Carlo experiments are quite favorable to the dynamic D-vine copula in comparison with a static D-vine. Our findings in the empirical application illustrate that time variation is present in the dependence structure of multivariate financial returns. Moreover, the dynamic D-vine copula outperforms the static D-vine in terms of predictive accuracy for our data sets.

CS30 Room B20 CONTRIBUTIONS IN BAYESIAN METHODS IN MACROECONOMICS AND FINANCE Chair: John M. Maheu

C911: **Application of Bayesian methods to financial contagion modelling using factor models**
Presenter: **Bill Sakaria**, University of Kent, United Kingdom
Co-authors: Jim Griffin
Contagion has been described as the spread of idiosyncratic shocks from one market to another in times of financial turmoil. This can be modelled using a global factor to capture the general market movements and idiosyncratic factors to capture co-movements and volatility spill-over between markets. Many previous studies have used pre-specified turmoil and calm periods to understand when contagion occurs. We introduce a time-varying parameter which indicates the volatility spill-over from one country to another and avoids the need to pre-specify particular types of period. Efficient Bayesian inference can be made using the Kalman filter in a forward filtering and backward sampling algorithm. The model is applied to market indices for Greece and Spain to understand the effect of contagion in the Euro crisis between Greece and Spain. Initial results show a positive or negative value of the volatility-spill over parameter during the volatile periods in the Euro crisis.

C955: **Bayesian inference for partially observed SDEs driven by fractional Brownian motion**
Presenter: **Konstantinos Kalogeropoulos**, London School of Economics, United Kingdom
Co-authors: Joseph Dureau, Alexandros Beskos
Continuous-time diffusion models driven by fractional Brownian Motion (fBM) are considered, with observations obtained at discrete-time instances. As a prototypical scenario we will give emphasis on a a stochastic volatility (SV) model allowing for memory in the volatility increments through an fBM specification. Due to the non-Markovianity of the model and the high-dimensionality of the latent volatility path, estimating posterior expectations is a computationally challenging task. We present novel simulation and re-parameterisation framework based on the Davies and Harte method and use it to construct a Markov chain Monte-Carlo (MCMC) algorithm that allows for computationally efficient parametric Bayesian inference upon application on such models. The algorithm is based on an advanced version of the so-called Hybrid Monte-Carlo (HMC) that allows for increased efficacy when applied on high-dimensional latent variables relevant to the models of interest in this paper. The inferential methodology is examined and illustrated in the SV models, on simulated data as well as real data from the S&P500/VIX time series. Contrary to a long range dependence attribute of the SV process (Hurst parameter $H > 1/2$) many times assumed, the posterior distribution favours $H < 1/2$ that points towards medium range dependence.

C1073: **Local Bayesian likelihood methods for estimating time-varying parameter DSGE models**
Presenter: **Katerina Petrova**, Queen Mary University London, United Kingdom
Co-authors: Ana Galvao, Liudas Giraitis, George Kapetanios
DSGE Models have recently received a lot of attention in macroeconomic analysis and forecasting. They are usually estimated using Bayesian Likelihood Methods, assuming that the deep parameters of the model remain fixed throughout the entire sample. A Local Bayesian Likelihood Method is presented which is suitable for estimation of a DSGE model that can be written in a state space form, and which can accommodate time-variation in all the model's parameters. There are two distinct advantages in allowing the structural parameters to vary over time. First, it allows us to address the question of how constant these parameters really are and, in particular, is there evidence of misspecification in the design of DSGE models or evidence that regime changes (e.g. the Great Moderation or the appointment of Paul Volcker) have had prolonged effects on some parameters. To address the question of structural change, an empirical application of a previous method with US data is presented. Second, allowing the parameters to vary and using their most recent values is expected to make the model more flexible and therefore improve its forecasting performance. A real-time pseudo out-of-sample forecasting assessment is presented and the time-varying model's performance is illustrated in comparison with the fixed parameter DSGE model, competing models and the Official Greenbook forecasts.

CS119 Room B29 CONTRIBUTIONS IN CHANGES IN VOLATILITY AND CORRELATION DYNAMICS Chair: Denis Pelletier

C986: **Sovereign markets' malfunctioning and observable indicators**
Presenter: **Michael Stein**, University of Duisburg-Essen, Germany
Co-authors: Roberto de Santis
It is uncontroversial that the dynamic correlation between sovereign yields and the monetary policy rates declines sharply if sovereign debt markets are malfunctioning. Therefore, we propose to study the problem using regime-dependent models for benchmark sovereign yields and the monetary policy rate with multivariate GARCH models of the smooth transition conditional correlation (STCC) type. Results indicate that observable indicators indeed lead to correlation changes and breakdowns in regime-dependent fashion.

C1136: **Short and long run second-order causality: Theory, measures and inference**
Presenter: **Hui Jun Zhang**, University of Cambridge, United Kingdom
Co-authors: Jean-Marie Dufour
As global economy being more volatile and integrated, understanding the causal linkage between volatilities (e.g. volatility spillover and contagion) becomes increasingly important. Most of the literature focus on characterizing causality for variance processes only at horizon one. The purpose is to extend this concept to multiple horizons. We define (cross-)second-order and (co-)variance causality between vectors at any given horizon, allowing for the presence of auxiliary variables. We then present important properties and characterizations of the causality structures so defined. Instead of studying tests of non-causality, we propose measures of (cross-)second-order causality at any given horizon. We suggest a simple estimation method and a simulation-based inference procedure to evaluate the measures in the context of stationary VAR-MGARCH models. The asymptotic properties of the estimator and the validity of bootstrap confidence intervals are provided. Finally, we apply the proposed measures of second-order causality to study volatility spillover and contagion across financial markets in the U.S., the U.K. and Japan during the period of 2000-2010.

C1137: **On the application of volatility-constrained correlation to determine directionality between financial markets**
Presenter: **Tomoshiro Ochiai**, Otsuma Womens University, Japan
Co-authors: Jose Nacher
Recent financial crises have shown the importance of identifying the origin of market instability. To address this issue, the directionality of the influence between financial assets must be determined. The standard correlation coefficient is one of the popular metrics to compute the relation between two assets. However, this classical approach is not enough to identify the direction of the influence between two assets. We present a volatility-constrained correlation metric and analyze the correlation between Japan's Nikkei stock average index (Nikkei 225) and other financial markets. The asymmetric feature of the metric reveals which asset is more influential than the other. The proposed computational method allows us to unveil the directionality of correlation effect, which could not be observed from the standard correlation analysis. Finally, we also present a theoretical model that reproduces the results observed in the empirical analysis.

C1235: **Cross-market spillovers with volatility surprise**

Presenter: **Julien Chevallier**, IPAG Business School, France
Co-authors: Sofiane Aboura

A methodological contribution, by using the 'volatility surprise' component to capture cross-market relationships, is advanced. More precisely, volatility spillovers are quantified in the asymmetric DCC with one exogeneous variable (ADCCX) framework. The dataset includes four aggregate indices representing equities, bonds, foreign exchange rates and commodities from 1983 to 2013. The results provide strong evidence of spillover effects coming from the 'volatility surprise' component across markets. In addition, asset management implications are derived.

Monday 16.12.2013 16:20 - 17:40 Parallel Session R – ERCIM

ES119 Room Gordon COMPUTATIONAL STATISTICS III Chair: Gil Gonzalez-Rodriguez

E1142: **Frequentist and Bayesian regression analysis of very large data sets in R**
Presenter: **Leo Geppert**, TU Dortmund University, Germany
Co-authors: Katja Ickstadt
R is widely used for conducting statistical analyses. R has a lot of advantages, but handling very large data sets is not one of them, unfortunately. Our aim is to be able to employ regression analysis (frequentist or Bayesian) when a very large number of observations is present. In our setting, the number of variables is typically moderately large ($n \gg d$). We suggest a method that is based on the 'merge and reduce' technique. The data is read in and analysed sequentially. By combining the resulting models in a suitable manner, we obtain results that are very close to the optimal model on the full data set. The complexity of the models is not increased in this process, which means that the number of observations can become very large without causing problems. This approach typically leads to a noticeable reduction of the run-time, especially if the analysis requires computationally demanding methods (e.g. MCMC-methods). We also present an R package that contains an implementation of the suggested method.

E339: **D-minimax second-order designs over hypercubes for extrapolation and restricted interpolation regions**
Presenter: **Shahariar Huda**, Kuwait University, Kuwait
The D-minimax criterion for estimating slopes of a response surface involving k factors is considered for situations where the experimental region and the region of interest are $k-$cubes centered at the origin but not necessarily identical. Taking the experimental region and the region of interest to be with edges of length 2 and length 2R, respectively, optimal designs under the criterion for the full second-order model are derived for various values of R and their relative performances investigated. The asymptotically optimal design as R goes to infinity is also derived and investigated. In addition, the optimal designs within the class of product designs are also obtained. In the asymptotic case it is found that the optimal product design is given by a solution of a cubic equation that reduces to a quadratic equation for $k = 3$ and 6. Relative performances of various designs obtained are examined. In particular, the optimal asymptotic product design and the traditional D-optimal design are compared and it is found that the former performs very well.

E206: **A control chart design for monitoring autocorrelated processes with multiple exogenous inputs under model uncertainty**
Presenter: **Maria Sofia Criselda Poblador**, University of the Philippines Manila, Philippines
In monitoring autocorrelated processes, an ARIMA model is fitted and the residuals are used in constructing the control limits. However, the order of the ARIMA model is a confounding issue since incorrect specification leads to inferior parameter estimates that will produce large residuals, resulting to wider control limits. Recent research shows that a design procedure which uses the AR-Sieve approximation assuming that the process allows for an autoregressive representation of order infinity has promising impact on the run length properties of the resulting charts. A control chart based on actual process behavior is investigated using the nonparametric design procedure, including multiple exogenous input factors in the process. Prediction limits are constructed via AR-Sieve bootstrap algorithm and compared with Exponentially Weighted Moving Average. Results show that the bootstrap approach produces narrower prediction limits, thus, a quicker detection of even minimal process mean shift during an out-of-control state.

E1212: **A cross-nested ordered probit model with an application to policy interest rate**
Presenter: **Andrei Sirchenko**, National Research University Higher School of Economics, Russia
The decisions to reduce, leave unchanged, or increase (the price, rating, interest rate) are often characterized by abundant no-change outcomes generated by different decision-making mechanisms. In addition, the positive and negative responses can also be driven by distinct forces. To explain the status quo inertia and unobserved heterogeneity of the data-generating process a two-level ordered probit model is developed. It estimates simultaneously three latent equations representing – in the interest rate setting context – an inclination decision that sets a policy stance (loose, neutral or tight) and amount decisions that fine-tune the rate conditionally on the stance. The cross-nested scenario creates three mechanisms generating distinct types of no-change responses, as well as separate processes for negative and positive outcomes. The model shows good small-sample performance in Monte Carlo experiments and, applied to the panel data on policymakers' votes for the interest rate, produces substantial improvements in statistical fit and qualitatively different inferences than the conventional models.

ES122 Room Montague CONTRIBUTIONS IN TIME SERIES ANALYSIS II Chair: Soumendra Lahiri

E956: **A Gini-based test for unit root**
Presenter: **Amit Shelef**, Ben-Gurion University of the Negev, Israel
A Gini-based statistical test for unit root is suggested which is based on the well-known Dickey-Fuller test, where the OLS regression is replaced by the semi-parametric Gini regression in modeling the AR process. A residual-based bootstrap is used for finding critical values. The Gini methodology is a rank-based methodology which takes into account both the variate values and the ranks. Therefore it provides robust estimators which are rank-based, while avoiding loss of information. Furthermore, the Gini methodology relies on first-order moment assumptions and, hence, is valid for a wide range of distributions. Simulation results validate the Gini-based test and indicate its superiority in some design settings compared to other available procedures. The Gini-based test opens the door for further developments such as a Gini-based cointegration test.

E1092: **A bootstrap contribution in forecasting**
Presenter: **Clara Cordeiro**, University of Algarve and CEAUL, Portugal
Co-authors: Maria Manuela Neves
Time series analysis has benefited from the use of computer-intensive procedures, to help modelling and predicting in more complex analytical situations. This field has received great deal of support from computational statistics, with the increasing use of methodologies such as the bootstrap. In time series, this methodology is frequently applied in the resampling of the residuals. In previous studies, the present authors focused on using bootstrap with a forecasting method, in order to obtain an estimate for a future period. This was a new approach to this resampling technique, instead of its traditional usage in forecast intervals. An extensive study involving exponential smoothing methods and the bootstrap revealed this to be a promising association, which resulted in an automatic algorithm for obtaining forecast intervals and point forecasts, considering time series with only one seasonal pattern. However, in the case of more than one seasonal pattern the forecasting method used in the procedure is not suitable. A new procedure was therefore derived where the appropriate forecasting method should include this type of seasonal component. The performance of this partnership will be illustrated with some well-known data sets using R software.

E1123: **Laplace moving average processes and their spectral representation**
Presenter: **Anastassia Baxevani**, University of Cyprus, Cyprus
Co-authors: Krzysztof Podgorski
For a class of processes defined as a convolution of a deterministic kernel with increments of Laplace motion we investigate its spectral repre-

sentation. The stochastic spectral measure for this non-Gaussian second order process has uncorrelated but dependent increments. We investigate its properties and present an explicit form that is useful for approximation of the process over the entire domain. Using the obtained representation of the spectral measure we prove that a Laplace moving average with proper kernels can be approximated over the entire range by a sum of trigonometric series with random coefficients. This approximation is then be used to effectively simulate the non-Gaussian process.

E1199: **A robust autoregressive forecasting method for the newsvendor problem**
Presenter: **Alba Victoria Olivares Nadal**, Universidad de Sevilla, Spain
Co-authors: Emilio Carrizosa, Pepa Ramirez-Cobo
The classic single-item newsvendor problem is explored under a novel setting which combines temporal dependence and tractable robust optimization. First, the demand is modeled as a time series which follows an autoregressive process $AR(p)$, $p \geq 1$. Second, a robust approach to maximize the worst-case revenue is proposed: a robust distribution-free autoregressive forecasting method, which copes with non-stationary time series, is formulated. A closed-form expression for the optimal solution is found for the problem for $p = 1$; for the remaining values of p, the problem is expressed as a nonlinear convex optimization program, to be solved numerically. The optimal solution under the robust method is compared with those obtained under two versions of the classic approach, in which either the demand distribution is unknown, and assumed to have no autocorrelation, or it is assumed to follow an $AR(p)$ process with normal error terms. Numerical experiments show that our proposal usually outperforms the previous benchmarks, not only with regard to robustness, but also in terms of the average revenue.

ES105 Room Bedford MULTIVARIATE STATISTICS II Chair: Piercesare Secchi

E810: **Predicting bivariate binary rare events responses using generalised extreme value regression model and copula function**
Presenter: **Raffaella Calabrese**, University of Milano-Bicocca, Italy
Co-authors: Silvia Angela Osmetti
A new bivariate Generalised Linear Model (GLM) is proposed for binary rare events, i.e. binary dependent variables with a very small number of ones. In a bivariate GLM model we suggest the quantile function of the Generalised Extreme Value (GEV) distribution. In this way, the drawback of the underestimation of the probability of the rare event in GLM models is overcome. The dependence between the response variables is modelled by the copula function. We explore different copula functions that provide a rich and flexible class of structures to derive joint distributions for bivariate binary data. Finally, we apply the proposed model to estimate the joint probability of defaults of UK and Italian small and medium enterprises.

E1061: **Assessing answer patterns in questionnaire/item response data using mixtures of Rasch models**
Presenter: **Hannah Frick**, Universitaet Innsbruck, Austria
Co-authors: Carolin Strobl, Achim Zeileis
The Rasch model is a standard model for binary item response data, e.g., as generated by respondents 'solving' items in an aptitude test or by giving yes/no answers in questionnaires. The Rasch model logistically links the probability of solving an item (or answering 'yes') to the difference between the respondent's ability and the item's difficulty. Based on observed item response data, conditional maximum likelihood estimation can then be used to infer item difficulties and respondent abilities. However, comparisons of respondents' abilities obtained this way are only fair if the same item difficulties hold for all respondents. Mixture models are flexible means to assess this assumption and uncover heterogeneity if there is any. Furthermore, if different patterns of item difficulty can be found in a mixture of Rasch models, covariates may be available to explain these: either ex-post or in a concomitant variable model within the mixture model. Here, we present a general framework for such Rasch mixture models along with its implementation in the R package 'psychomix'. It includes the possibility of concomitant variable models as well as various options for modeling the distribution of respondents' abilities. The methods are illustrated using questionnaire data on the extrinsic motivation of R package authors. In this application, the mixture model can uncover different patterns of item difficulty for groups of the R package authors which can be partially attributed to the covariates education and occupation.

E1086: **Wold approach for nonlinear PLS mode B structural models**
Presenter: **Alba Martinez-Ruiz**, Universidad Catolica de la Santisima Concepcion, Chile
Wold suggested that PLS path modeling algorithm could be modified to allow estimation of nonlinear effects in PLS structural models. In contrast to the traditional two-step approach that involves the computation of nonlinear variables outside the PLS iterative algorithm, Wold approach considers the computation of nonlinearities into the iterative algorithm. We implemented the Wold approach for nonlinear PLS Mode B structural models. Computation of nonlinear variables into the iterative algorithm allows for updating of nonlinear constructs and inner weights using the last available information at each iteration. We investigated the performance of the Wold approach using Monte Carlo simulation. The true model considered a structure with three linear effects, a nonlinear effect and an interaction effect on a Mode B endogenous construct. The experimental design included models with two, four, six and eight indicators per construct and three sample sizes (150, 250, 500). We generated 500 random data for 12 different specifications. For models with four indicators per construct, mean relative bias of a linear effect of 0.4 decreases from 22% to 16% by increasing the sample size. For nonlinear effects, Wold approach for PLS Mode B structural models performs similar to the two-step approach for nonlinearities.

E1119: **Clustering of multivariate binary data via penalized latent class analysis with dimension reduction**
Presenter: **Michio Yamamoto**, Kyoto University, Japan
Co-authors: Kenichi Hayashi
It is well-known that reducing dimensionality of data improves the performance of clustering algorithms, especially in high-dimensional data; a lot of methods to perform simultaneously clustering and dimension reduction have been developed. A new procedure is presented for simultaneously finding an optimal cluster structure of multivariate binary data and a subspace to represent the cluster structure. The method is based on a finite mixture model of multivariate Bernoulli distributions and parameters of each component are assumed to have low-dimensional representations. Our model can be considered an extension of the traditional latent class analysis model. The proposed method introduces sparsity to the loading vectors, which produce the low-dimensional subspace, for enhanced interpretability and more stable extraction of the subspace. An EM-based algorithm is developed to efficiently solve the proposed optimization problem. The effectiveness of the proposed method is illustrated by application to artificial and real data sets.

ES116 Room 349 METHODOLOGICAL STATISTICS Chair: Joseph L. Gastwirth

E827: **The penalized analytic centre estimator**
Presenter: **Keith Knight**, University of Toronto, Canada
In a linear regression model, the Dantzig selector minimizes the L_1 norm of the regression coefficients subject to a bound λ on the L_∞ norm of the covariances between the predictors and the residuals; the resulting estimator is the solution of a linear program, which may be non-unique or unstable. We propose a regularized alternative to the Dantzig selector. These estimators (which depend on λ and an additional tuning parameter r) minimize objective functions that are the sum of the L_1 norm of the regression coefficients plus r times the logarithmic potential function of the Dantzig selector constraints, and can be viewed as penalized analytic centres of the latter constraints. The tuning parameter r controls the smoothness of the estimators as functions of λ and when λ is sufficiently large, the estimators depend approximately on r and λ via r/λ^2.

E1097: **The convergence of the reverse integrated empirical process in L_p with applications**
Presenter: **Javier Carcamo**, Universidad Autonoma de Madrid, Spain
Co-authors: Sofia Nieto, Amparo Baillo
The necessary and sufficient condition on a positive random variable X so that the reverse integrated empirical process converges in distribution in L^p, for $1 \leq p \leq 2$ is found. This condition defines a Lorenz space and can be also characterized in terms of several integrability conditions related to the process $\{(X-t)_+ : t \geq 0\}$. The main ideas are applied to obtain asymptotic results related to the Zolotarev's distance between probability distributions and statistical tests for direct stochastic dominance.

E1219: **Semiparametric mixture cure transformation models for clustered interval-censored survival data**
Presenter: **Liming Xiang**, Nanyang Technological University, Singapore
Co-authors: Xiangmei Ma
A semiparametric transformation cure model approach, which allows for the intracluster correlation of survival data and avoids pre-specified model assumptions, thus providing a broad class of candidate models for the analysis of clustered interval-censored data in the presence of a cured subgroup. We develop an EM algorithm for obtaining maximum likelihood estimates of parameters in conjunction with a self-consistent iterative algorithm for estimating the nonparametric component. We conduct simulation studies to assess the performance of the proposed estimators and illustrate the methodology with two real data examples.

E1041: **Hidden mixture transition distribution (MTD) model: General framework and model selection criteria**
Presenter: **Danilo Bolano**, University of Geneva, Switzerland
Co-authors: Andre Berchtold
Modeling time series presenting non-Gaussian features play a central role in many fields like finance, seismology, biometrical and psychological studies. Hidden Mixture Transition Distribution models are an answer to the complexity of such series because the observed heterogeneity can be induced by one or several latent factors and each level of these factors is related to a different component of the observed process. The time series is then seen as a weighted mixture and the relation between successive components is governed by a Markovian latent transition process. One of the main difficulties lies in the correct specification of the structure of the model: number of components and lags, use of covariates at the hidden and/or observed levels, modeling of the standard deviation of each component and so on. We will illustrate the critical step of reducing the complexity of the model and the selection of an appropriate model specification using ad hoc inductive hierarchical procedure. Starting from the simplest model possible, at each step the more advanced elements are added once at a time and the models are compared using information criteria. We illustrate our approach on a real data set, the U.S. Panel Study of Income Dynamics.

ES98 Room Woburn CONTRIBUTIONS TO COMPUTATIONAL APPROACHES IN FINANCIAL ECONOMICS Chair: Efstathia Bura

E079: **A multivariate Hill estimator**
Presenter: **Yves Dominicy**, Universite libre de Bruxelles, Luxembourg
Co-authors: Pauliina Ilmonen, David Veredas
An estimator of the tail behaviour of elliptical distributions is proposed. The proposed estimator is a multivariate extension of the well-known Hill estimator, which is an estimate of the extreme value index, and hence of the tail index, of univariate heavy-tailed distributions. We use the Minimum Covariance Determinant method to construct our multivariate elliptical Hill estimator. This procedure supplies us, for a given threshold, with the outlying points we are interested in. Furthermore the Minimum Covariance Determinant approach provides us with robust estimators for the location vector and the scatter matrix. We show that our multivariate tail estimator inherits the same characteristics as the univariate Hill. We develop the consistency and asymptotic properties for the estimator when we have the true location vector and scatter matrix. A Monte Carlo experiment reveals the goodness of the estimator and that the multivariate elliptical Hill estimator behaves in a similar way to the univariate one. Finally, a small empirical illustration to financial data shows that a portfolio of financial returns is indeed heavy-tailed.

E946: **An agent-based model for market impact**
Presenter: **Christian Oesch**, University of Basel, Switzerland
Studying market impact will help us to understand how financial markets incorporate information into prices and how liquidity affects the price building process. For practitioners, market impact matters because if one has a better understanding of the effect, one might be able to dodge some of the costs associated with market impact. Three different kinds of agents are used on this market: Institutional agents make up the first group and decide on large orders based on portfolio considerations. When submitting the order to the market, they split them up into smaller parts to evade price impact costs. A second group of agents will act as liquidity providers which forecast order flow. To add additional depth to the order book, zero-intelligence agents represent the third group of agents. The model is able to produce permanent market impact while keeping statistical price efficiency as well as concave market impact functions.

E1060: **Optimal macroeconomic policies for nonlinear econometric models with rational expectations**
Presenter: **Reinhard Neck**, Alpen-Adria Universitaet Klagenfurt, Austria
Co-authors: Dmitri Blueschke, Viktoria Blueschke-Nikolaeva
An algorithm called OPTCONRE is presented which provides approximations of optimal control paths for nonlinear econometric models with rational (forward-looking) expectations. It can be applied to obtain approximate numerical solutions to optimal control problems with a quadratic objective function for nonlinear econometric models where some variables are forward-looking instead of having the usual 'causal' time structure of models with backward-looking or static expectations. The algorithm was programmed in MATLAB and allows for deterministic and stochastic control, the latter with open-loop and passive learning information patterns. We demonstrate the applicability of the algorithm by an application to a policy problem using a small quarterly macroeconometric model for Slovenia and investigate differences in the optimal policy design between a model version with static and with rational expectations about interest and inflation rates. This shows the convergence and the practical usefulness of the algorithm for problems of stabilization policy in small-sized macroeconometric models. In addition, we show that the optimal policies

derived by OPTCONRE are superior in terms of the values of the optimized objective functions to policies derived for the same models by a heuristic method.

E1206: **Correlation networks: Processes, time series and surrogate data**
Presenter: **Annette Witt**, Max-Planck Institute for Dynamics and Self-Organization, Germany
Co-authors: Jan Nagler, Magdalena Kersting, Theo Geisel
Correlation networks with N vertices are considered, each vertex is associated to (i) a time series (on the realization level) or (ii) to its generating stochastic process (on the process level), links between the vertices are represented by cross correlation functions, self-links stand for autocorrelation functions. On the realization level we establish conditions for the reconstruction of the complete correlation network from a subnetwork and show that these subnetworks must meet certain loop structures. A parameter free method for the construction of correlation networks on the realization level from correlation networks on the process level is given which can also be used for generating a set of Gaussian distributed time series for a given correlation network. The properties of these time series and of so-called surrogate data (a resampling method that is common in the physical time series analysis community) are compared. Finally, an application to data from finance is presented.

ES130 Room 104 COPULAS Chair: Fabio Spizzichino

E118: **Spline estimation of conditional copulas**
Presenter: **Philippe Lambert**, University of Liege, Belgium
A new family of conditional Archimedean copula is presented. It starts from a (spline based) nonparametric estimate of the generator of an Archimedean copula that extends a previous proposal. That flexible form is generalized by letting its spline coefficients change with covariates. A smooth evolution can be forced by specifying a latent Gaussian prior for these coefficients. We consider the framework of a new family of conditional Archimedean copulas that we propose to name flex-power Archimedean. The flexibility and the properties of this nonparametric conditional copula estimator are studied using simulations. The joint posterior of the model parameters are explored using MCMC. The possibility of fast approximations will also be discussed. The modelling strategy is illustrated on real data.

E156: **Semiparametric Gaussian copula models: Geometry and efficient rank-based estimation**
Presenter: **Ramon van den Akker**, Tilburg University, Netherlands
Co-authors: Johan Segers, Bas J.M. Werker
For multivariate Gaussian copula models with unknown margins and arbitrary correlation structures, a simple, rank-based, semiparametrically efficient estimator is proposed. An algebraic representation of relevant subspaces of the tangent space is constructed that allows us to easily study questions of adaptivity with respect to the unknown marginal distributions and of efficiency of, e.g., the pseudo-likelihood estimator. Some well-known correlation structures are treated explicitly: circular matrices, factor models, and Toeplitz matrices. For constructed examples, the asymptotic relative efficiency of the pseudo-likelihood estimator can be as low as 20%. For finite samples, these findings are confirmed by Monte Carlo simulations.

E1195: **Smoothing out pseudo-observations for copula models**
Presenter: **Hideatsu Tsukahara**, Seijo University, Japan
Many statistics proposed in the context of estimation and testing involving copulas are functions of the empirical probability integral transforms of the data, a form of the so-called pseudo-observations. They are discrete in nature, as can be seen from the fact that they equal simple functions of the coordinate-wise ranks. In view of practical applications, one might guess that there is an advantage of 'smoothing out' in small samples. Such a scheme, which is extremely simple, has recently been suggested. Briefly speaking, it consists of uniform random variates ordered with the same configuration as the vector of ranks for the original sample. We study its finite-sample properties and also its asymptotic behavior. We use Monte Carlo simulation to examine the effect of this new scheme in the context of maximum pseudo-likelihood estimation and various goodness-of-fit test statistics.

E988: **A copula to handle tail dependence in high dimension**
Presenter: **Gildas Mazo**, Grenoble, France
Co-authors: Florence Forbes, Stephane Girard
The concept of copula is a useful tool to model multivariate distributions but the construction of tail dependent high dimensional copulas remains a challenging problem. We propose a new copula constructed by introducing a latent factor. Conditional independence with respect to this factor and the use of a nonparametric class of bivariate copulas lead to interesting properties like explicitness, flexibility and parsimony. We propose a pairwise moment-based inference procedure and prove asymptotic normality of our estimator. Finally we illustrate our model on simulated and real data.

ES57 Room Beveridge CONTRIBUTIONS TO ROBUST STATISTICS Chair: Christophe Croux

E193: **Nonlinear mixed-effects models and diagnostics for censored HIV viral loads with CD4 measurement error**
Presenter: **Victor Hugo Lachos**, UNICAMP, Brazil
Despite the technological advancements in efficiency enhancement of quantification assays, biomedical studies on HIV RNA measures viral load responses are often subjected to some detection limits. Moreover, some related covariates such as CD4 cell count are often measured with substantial errors. Censored nonlinear mixed-effects models are routinely used to analyze this type of data and are based on normality assumptions for the random terms. However, the derived inference may be non-robust when the underlying normality assumptions are questionable (fat-tails). We address these issues simultaneously under a Bayesian paradigm through joint modeling of response and covariate processes using an attractive class of normal/independent (NI) densities. The NI family produces symmetric heavy-tailed distributions that includes the normal distribution, the Student-t, slash and the contaminated normal distributions as special cases. The methodology is illustrated using a case-study on longitudinal HIV viral loads. Both simulation and real data analysis reveal that our models are capable of providing robust inference for heavy-tailed situations commonly encountered in HIV/AIDS, or other clinical studies.

E883: **A robust version of least angle regression**
Presenter: **Shirin Shahriari**, University of Minho, Portugal
Co-authors: Susana Faria, A. Manuela Goncalves
The problem of selecting a parsimonious subset of variables from a large number of predictors in a regression model is considered when the data contains both vertical and leverage outliers. Variable selection and outlier detection are inseparable problems, therefore a robust method that can select variables and detect outliers simultaneously is needed. Least angle regression algorithm orders the variables based on their importance in the model. This algorithm is not robust in the presence of atypical observations in data. A robust version of this algorithm is studied, and the performance of this algorithm is considered by using simulated data.

E1056: **Significance of explanatory variables for the robustified least squares**
Presenter: **Jan Amos Visek**, Charles University in Prague, Czech Republic

An estimator which is a slight modification of the classical *Ordinary Least Squares* is considered. The modification, called the *Least Weighted Squares*, minimizes the sum of weighted order statistics of the squared residuals rather than the squared residuals directly. To treat the order statistics of the squared residuals instead of the squared residuals directly would be complicated. Fortunately a simple trick reveals that the estimator is one of the solutions of the normal equations in which the appropriately reordered weights are included. Recently established uniform convergence of the empirical d. f. to its theoretical counterpart hints a way how to study the asymptotic properties of the estimator. The idea of *implicit weighting the squared residuals* opens a straightforward way for robustification of many classical statistical and econometric methods, as e. g. the *Instrumental Variables* or the *Total Least Squares*. There is already a number of results on these robustified methods, but the basic problems of significance of explanatory variables had not yet been considered. Extensive simulations hint an applicability of the result.

E957: **The shooting S-estimator: Robust regression in the independent contamination model**
Presenter: **Viktoria Oellerer**, KU Leuven, Belgium
Co-authors: Andreas Alfons, Christophe Croux

Classical parametric regression assumes that the observed data follow a model $y_i = \mathbf{x}_i'\beta + e_i$ with $i = 1,\ldots,n$. In robust regression, however, a minority of the data may be affected by abnormal noise. Thus, the data actually comes from a mixture distribution where the dominant component follows the model above and the minority component ('outliers') is unspecified. Robust techniques try to minimize the influence of those outliers. This approach, however, has some limitations. In case the contamination in the different variables is independent of each other ('independent contamination model'), there may not be any observations anymore that are not contaminated in any of its variables. But each observation will mainly consist of clean values for the different variables. Using the usual approach of downweighting observations as a whole, will cause most robust methods to break down. We propose a regression estimator called *shooting S-estimator* that takes this variablewise ('cellwise') contamination into account. It combines the coordinate descent method ('shooting algorithm') with regression S-estimation. An algorithm for the computation of shooting S-estimates is provided.

ES128 Room Bloomsbury CONTRIBUTIONS IN FUNCTIONAL DATA ANALYSIS Chair: Serge Guillas

E1015: **Bayesian bandwidth estimation for a nonparametric functional regression model**
Presenter: **Han Lin Shang**, University of Southampton, United Kingdom

The issue of bandwidth estimation is investigated in a nonparametric functional regression model with function-valued, continuous real-valued and discrete-valued regressors under the framework of unknown error density. We propose to approximate the unknown error density by a location-mixture of Gaussian densities with means being the individual residuals, and variance a constant parameter. This proposed mixture error density has a form of a kernel density estimator of residuals, where the regression function is estimated by the functional Nadaraya-Watson estimator that admits mixed types of regressors. We put forward a Bayesian bandwidth estimation approach that can simultaneously estimate the bandwidths in the kernel-form error density and functional Nadaraya-Watson estimator. We derive a kernel likelihood and posterior density for the bandwidth parameters under the kernel-form error density. Simulation studies demonstrated the estimation accuracy of the regression function and error density for the proposed Bayesian approach. Illustrated by a spectroscopy data set in the food quality control, we applied the proposed Bayesian procedure to select the optimal bandwidths in a nonparametric functional regression model with mixed type of regressors, and found that it has a more superior point forecast accuracy than the same regression model without discrete-valued regressor.

E1147: **Spatial depth-based outlier detection for functional data**
Presenter: **Carlo Sguera**, Universidad Carlos III de Madrid, Spain
Co-authors: Pedro Galeano, Rosa Lillo

The functional outlier detection problem is tackled using the idea of functional depth, and in particular the kernelized functional spatial depth (KFSD). The proposed method is inspired by a multivariate outlier detection procedure which requires the availability of training data. However, in real outlier detection problems the availability of training data is a quite unusual occurence. This drawback is overcome by proposing a resampling-based solution. Let Z_n be an observed functional dataset and obtain Y_r by simply and randomly resampling from Z_n. Based on a KFSD-based analysis of Y_r, the method provides with a threshold t for the KFSD function. Then, once obtained the KFSD values of the curves in Z_n, the KFSD values of Z_n are compared with t: those curves having a KFSD value not greater than t are flagged as potential outliers. The method is evaluated by considering challenging simulation scenarios in which the outliers are not so different with respect to non-outlying curves. The performances of the proposed procedure are compared to the ones of different existing functional outlier detection methods, and the results indicate that a KFSD-based outlier detection approach lead to good results. Finally, some real functional datasets are considered.

E1198: **Functional principal fitted component regression models using penalized B-splines**
Presenter: **AhYeon Park**, University College London, United Kingdom
Co-authors: Serge Guillas

A functional linear model with a scalar response y and a functional covariate $X(t)$ in $L^2[0,T]$ is considered. Dimension reduction for this regression can be achieved either by Functional Principal Components Regression (FPCR) or through a penalized B-splines regression. In FPCR the subspace of the functional coefficient $\beta(t)$ is restricted to the span of selected functional principal components, thus the components are chosen without regard to how well they predict the response. We introduce a model that takes into account the response when choosing the components. It is an extension of the multivariate principal fitted components regression model to functional data. We first carry out an inverse regression of X on y to compute Functional Principal Fitted Components (FPFCs) using maximum likelihood. Here, we restrict the subspace of FPFCs to B-Splines. Then, $\beta(t)$ is estimated by regressing y on these FPFCs and the roughness of $\beta(t)$ is controlled via a penalized B-Splines approach. We analyse sensitivity to various tuning parameters. Simulation studies and real data applications demonstrate that FPFC regression can outperform FPCR in terms of prediction error.

E908: **On a maximal increment of partial sums**
Presenter: **Alfredas Rackauskas**, Vilnius University, Lithuania

Let B be a separable Banach space with the norm $||x||, x \in B$. Let $X, X_1, X_2, \ldots$ be a sequence of mean zero independent identically distributed random elements in B. We consider the Erdös-Rényi type maximum of partial sums $R_n(\ell) = \max_{0\le k\le n-\ell} ||X_{k+1} + \cdots + X_{k+\ell}||, \quad 1 \le \ell \le n, \quad n \ge 1$. We discuss some limit theorems for $b_n^{-1} \max_{1\le\ell\le n} \ell^{-\alpha} R_n(\ell), \quad n \ge 1$, where $0 \le \alpha < 1$ and $b_n, n \ge 1$ is a norming sequence. Some applications to functional data analysis shall be presented.

ES106 Room Senate STATISTICAL INFERENCE II Chair: Wicher Bergsma

E1141: **Checking linearity for a multivariate regression model**
Presenter: **M. Brigida Ferraro**, University La Sapienza, Rome, Italy
Co-authors: Ana Colubi, Gil Gonzalez-Rodriguez
A linearity test for a multivariate regression model is introduced and analyzed. The proposed linearity test is based on empirical processes of the regressors marked by the residuals. Taking into account some properties of the linear combinations, it can be used a linear combination of the test statistics referred to each model or a test statistic of a model in which the response is a linear combination of the responses. Since large samples are required to obtain suitable asymptotic results a bootstrap approach is investigated. Furthermore, in order to illustrate how the proposed test works in practice, some simulation and real-case studies are given.

E1016: **A test for the validity of the hnbue assumption based on an L1 distance**
Presenter: **Sofia Nieto**, Uam, Spain
Co-authors: Amparo Baillo, Javier Carcamo
We consider a test to check if a random variable X is hnbue (Harmonic New Better than Used in Expectation), an ageing class commonly used in reliability theory. We propose a test statistic based on an L_1 distance between integrated survival functions defined in terms of a random sample of X. We discuss in depth the asymptotic behavior of the test statistic: its asymptotic distribution; the universal consistency of the testing procedure; and the rates of the corresponding power. We show the performance of our proposal in practice via some simulations and the analysis of real data sets.

E937: **A test for endogeneity in conditional quantiles**
Presenter: **Christophe Muller**, Aix-Marseille School of Economics, France
Co-authors: Tae-Hwan Kim
A test to detect the presence of endogeneity in conditional quantiles is developed. Our test is a Hausman-type test based on the distance between two estimators, of which one is consistent only under no endogeneity while the other is consistent regardless of the presence of endogeneity in conditional quantile models. We derive the asymptotic distribution of the test statistic under the null hypothesis of no endogeneity. The finite sample properties of the test are investigated through Monte Carlo simulations, and it is found that the test shows good size and power properties in finite samples. Our approach does not imply an infeasible computation time. Finally, we apply our approach to test for endogeneity in conditional quantile models for estimating Engel curves using UK consumption and expenditure data. The pattern of endogeneity in the Engel curve is found to vary substantially across quantiles.

E1046: **Estimation of two ordered normal means with known covariance matrix and its numerical analysis**
Presenter: **Genso Watanabe**, Mejiro University, Japan
Co-authors: Nobuo Shinozaki
The estimator of two ordered normal means with known covariance matrix is discussed. We evaluate the restricted maximum likelihood estimator (RMLE) which satisfied the order restriction with known covariance matrix. Other types of estimators have been proposed and it has been showed that they dominate the mean that does not take into account the order restriction and covariance matrix. Here we compare those three type estimators under stochastic domination criterion and modified Pitman closeness criterion and show that RMLE is most desirable estimator. Finally, we illustrate the behaviour of those three type estimators by numerical analysis.

Authors Index