PROGRAMME AND ABSTRACTS

8th International Conference on

Computational and Financial Econometrics (CFE 2014)

http://www.cfenetwork.org/CFE2014

and

7th International Conference of the
ERCIM (European Research Consortium for Informatics and Mathematics) Working Group on

Computational and Methodological Statistics (ERCIM 2014)

http://www.cmstatistics.org/ERCIM2014

University of Pisa, Italy
6 – 8 December 2014

http://www.qmul.ac.uk

UNIVERSITÀ DI PISA

http://www.unipi.it

http://www.unisa.it

Depósito Legal: AS 03859-2014
ISBN: 978-84-937822-4-5

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

Dear Friends and Colleagues,

We warmly welcome you to Pisa, for the Eighth International Conference on *Computational and Financial Econometrics* (CFE 2014) and the Seventh International Conference of the ERCIM Working Group on *Computational and Methodological Statistics* (ERCIM 2014). 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 2014 programme consists of 275 sessions, 5 plenary talks and over 1150 presentations. There are over 1250 participants. This is the biggest meeting of the conference series in terms of number of participants and presentations. The growth of the conference in terms of size and quality makes it undoubtedly one of the most important international scientific events in the area.

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 University of Pisa, University of Salerno and Queen Mary University of London, UK.

Looking forward, the CFE-ERCIM 2015 will be held at the Senate House, University of London and Birkbeck University of London, UK, from Saturday the 12th to Monday the 14th December 2015. Tutorials will take place on Friday the 11th of December 2015. You are invited and encouraged to actively participate in these events.

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

The CFE-ERCIM 2014 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 1150 members and the following specialized teams

BM: Bayesian Methodology

CODA: Complex data structures and Object Data Analysis

CPEP: Component-based methods for Predictive and Exploratory Path modeling

DMC: Dependence Models and Copulas

DOE: Design Of Experiments

EF: Econometrics and Finance

GCS: General Computational Statistics WG CMStatistics

GMS: General Metholological Statistics WG CMStatistics

GOF: Goodness-of-Fit and Change-Point Problems

HDS: High-Dimensional Statistics

ISDA: Imprecision in Statistical Data Analysis

LVSEM: Latent Variable and Structural Equation Models

MCS: Matrix Computations and Statistics

MM: Mixture Models

MSW: Multi-Set and multi-Way models

NPS: Non-Parametric Statistics

OHEM: Optimization Heuristics in Estimation and Modelling

RACDS: Robust Analysis of Complex Data Sets

SAE: Small Area Estimation

SAET: Statistical Analysis of Event Times

SAS: Statistical Algorithms and Software

SEA: Statistics of Extremes and Applications

SFD: Statistics for Functional Data

SL: Statistical Learning

SSEF: Statistical Signal Extraction and Filtering

TSMC: Times Series Modelling and Computation

You are encouraged to become a member of the WG. For further information please contact the Chairs of the specialized groups (see the WG's 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 group 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 researchers 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 published under the CSDA Annals of CFE) and submitting research proposals.

Specialized teams

Currently the CFEnetwork has over 700 members and the following specialized teams

AE: Applied Econometrics

BE: Bayesian Econometrics

BM: Bootstrap Methods

CE: Computational Econometrics

ET: Econometric Theory

FA: Financial Applications

FE: Financial Econometrics

TSE: Time Series Econometrics

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

SCHEDULE

CFE 2014

Saturday, 6th December 2014

09:00 - 09:10	Opening CFE-ERCIM
09:10 - 10:00	Plenary Session A
10:00 - 10:30	Coffee Break
10:30 - 11:45	Parallel Sessions C
11:55 - 12:45	Plenary Session D
12:45 - 14:35	Lunch Break
14:35 - 16:15	Parallel Sessions E
16:15 - 16:45	Coffee Break
16:45 - 18:50	Parallel Sessions G
20:30 - 22:15	Reception

Sunday, 7th December 2014

08:45 - 10:25	Parallel Sessions I
10:25 - 10:55	Coffee Break
10:55 - 13:00	Parallel Sessions J
13:00 - 14:45	Lunch Break
14:45 - 16:25	Parallel Sessions K
16:25 - 16:55	Coffee Break
16:55 - 18:35	Parallel Sessions L
20:30 - 23:30	Conference Dinner

Monday, 8th December 2014

08:45 - 10:05	Parallel Sessions N
10:05 - 10:35	Coffee Break
10:35 - 12:15	Parallel Sessions P
12:25 - 13:15	Plenary Session Q
13:15 - 14:50	Lunch Break
14:50 - 16:30	Parallel Sessions R
16:30 - 16:55	Coffee Break
16:55 - 18:15	Parallel Sessions S
18:25 - 19:15	Plenary Session T
19:15 - 19:30	Closing

ERCIM 2014

Saturday, 6th December 2014

09:00 - 09:10	Opening CFE-ERCIM
09:10 - 10:00	Plenary Session A
10:00 - 10:30	Coffee Break
10:30 - 12:35	Parallel Sessions B
12:35 - 14:35	Lunch Break
14:35 - 16:15	Parallel Sessions E
16:15 - 16:45	Coffee Break
16:45 - 18:00	Parallel Sessions F
18:10 - 19:00	Plenary Session H
20:30 - 22:15	Reception

Sunday, 7th December 2014

08:45 - 10:25	Parallel Sessions I
10:25 - 10:55	Coffee Break
10:55 - 13:00	Parallel Sessions J
13:00 - 14:45	Lunch Break
14:45 - 16:25	Parallel Sessions K
16:25 - 16:55	Coffee Break
16:55 - 19:00	Parallel Sessions M
20:30 - 23:30	Conference Dinner

Monday, 8th December 2014

08:45 - 10:05	Parallel Sessions N
10:05 - 10:35	Coffee Break
10:35 - 12:15	Parallel Sessions O
12:25 - 13:15	Plenary Session Q
13:15 - 14:50	Lunch Break
14:50 - 16:30	Parallel Sessions R
16:30 - 16:55	Coffee Break
16:55 - 18:15	Parallel Sessions S
18:25 - 19:15	Plenary Session T
19:15 - 19:30	Closing

TUTORIALS, MEETINGS AND SOCIAL EVENTS

TUTORIALS

The tutorials will take place on Friday the 5th of December 2014 at the central builidng of the Scuola Superiore Sant'Anna, Piazza Martiri della Liberta' 33, Pisa (see the map at page X). The first one is given by Prof. D. Stephen G. Pollock (Band pass filtering and Wavelets analysis: Tools for the analysis of inhomogeneous time series - Aula Magna) at 9:00-13:30. The second ones are given by Prof. Mike West (Multivariate Time Series: Bayesian Dynamic Modelling & Forecasting - Aula Magna) and Prof. Vincenzo Esposito Vinzi (Component-based path modeling - Aula Magna Storica) at 15:00 - 19:30.

SPECIAL MEETINGS by invitation to group members

- The *CSDA Editorial Board* meeting will take place on Friday 5th of December 2014 from 18:00-19:30.
- The *CSDA & CFEnetwork Editorial Boards* dinner will take place on Friday 5th of December 2014 from 20:00-22:30.
- The *ERCIM Specialized Teams Co-chairs* meeting will take place on Saturday 6th of December 2014, 14:05-14:30, Lecture Room A0.
- The *CFEnetwork Editorial Board* meeting will take place on Saturday 6th of December 2014, 19:05-19:25, Lecture Room A0.
- The *IASC Assembly* meeting Sunday 7th of December, 13:00-14:30, Lecture Room A0. Open to all IASC members.
- The *COST Action CRONOS* meeting will take place on Sunday 7th of December 2014, 13:15-13:45, Lecture Room Sala Convegni. The meeting is by invitation only.

SOCIAL EVENTS

- *The coffee breaks* will take place at the halls of each floor of the venue (Polo didattico delle Piagge dell'Università di Pisa).
- *Welcome Reception, Saturday 6th of December, from 20:30 to 22:15*. The Welcome Reception is open to all registrants who had pre-registered and accompanying persons who have purchased a reception ticket. It will take place at Leopolda (Stazione Leopolda, Piazza Guerrazzi 2, 56100 Pisa, see map at page X). Conference registrants must bring their conference badge and ticket and any accompanying persons should bring their reception tickets in order to attend the reception. Preregistration is required due to health and safety reasons, and limited capacity of the venue. Entrance to the reception venue will be strictly allowed only to those who have a ticket.
- *Conference Dinner, Sunday 7th of December, from 20:30 to 23:30*. The conference dinner is optional and registration is required. It will take place at Leopolda (Stazione Leopolda, Piazza Guerrazzi 2, 56100 Pisa, see map at page X). Conference registrants and accompanying persons should bring their conference dinner tickets in order to attend the conference dinner.

Addresses of venues:

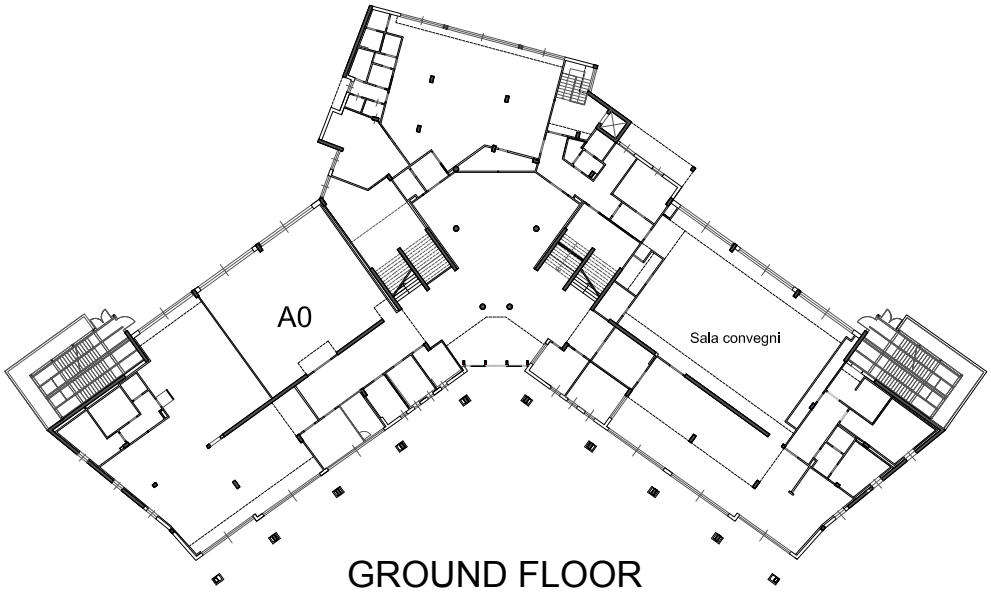
- Polo didattico delle Piagge dell'Università di Pisa, Via Giacomo Matteotti 3, 56124 Pisa.
- Conference Center, Palazzo dei Congressi di Pisa, Via Giacomo Matteotti 1, 56124 Pisa.
- Central builidng of the Scuola Superiore Sant'Anna, Piazza Martiri della Liberta' 33, Pisa.

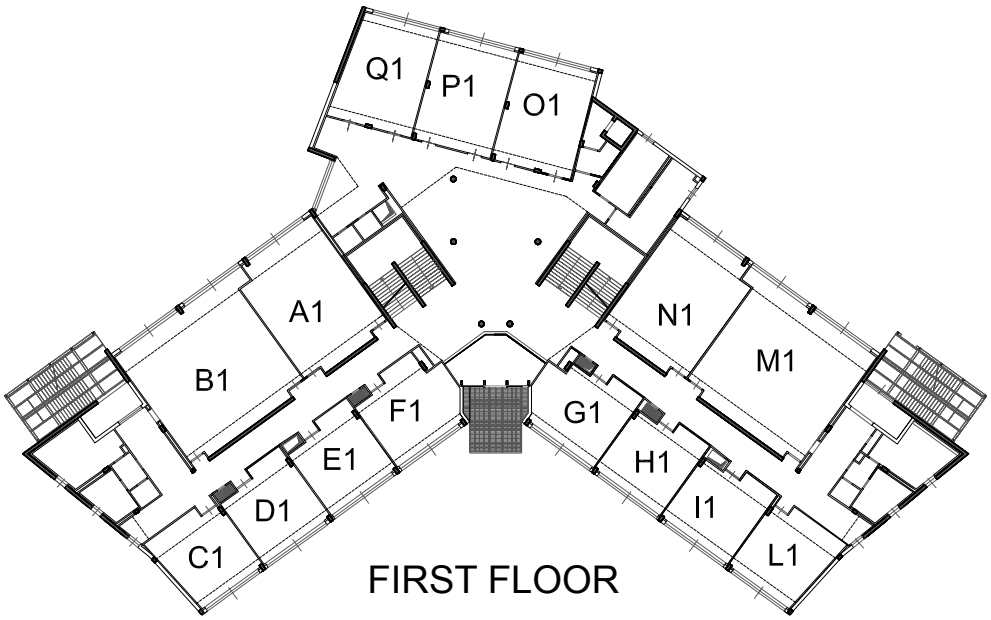
Registration, exhibitors and networking activities

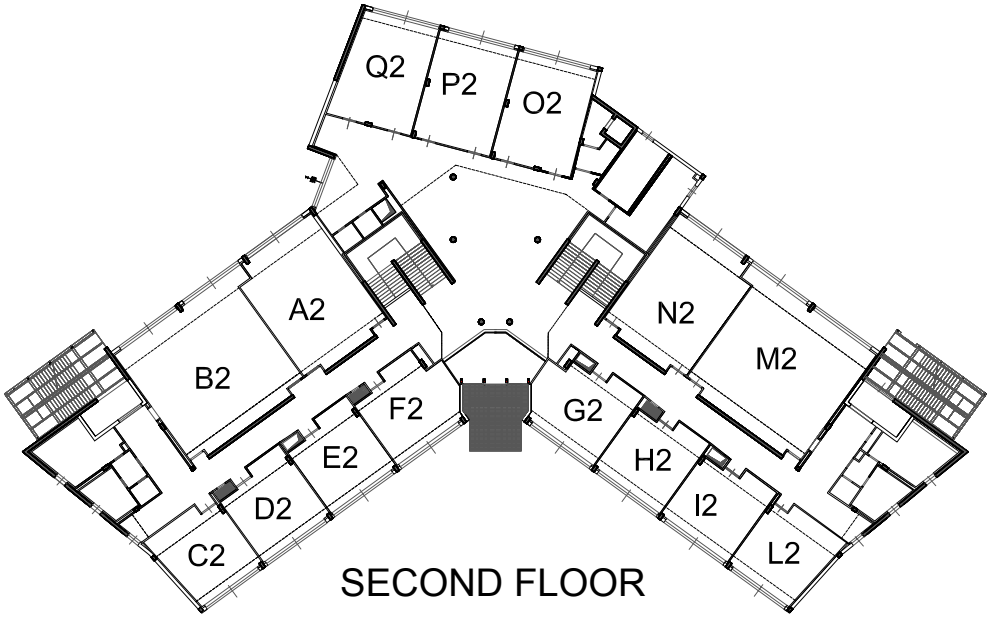
The registration and exhibitors will be located in the ground floor of the Polo didattico delle Piagge. The Lecture Room A0 is available for meetings upon request.

Lecture rooms

The paper presentations will take place at the Polo didattico delle Piagge dell'Università di Pisa (see map in page X). The different rooms are shown in the following floor plans of the venue. We advise that you visit the venue in advance. The opening, keynote and closing talks will take place at the Conference Center, Palazzo dei Congressi di Pisa. The poster sessions will take place at the first floor Hall of the Polo didattico delle Piagge dell'Università di Pisa.







Presentation instructions

The lecture rooms will be equipped with a laptop 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 laptops 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 first floor Hall of the Polo didattico delle Piagge dell'Università di Pisa. 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.

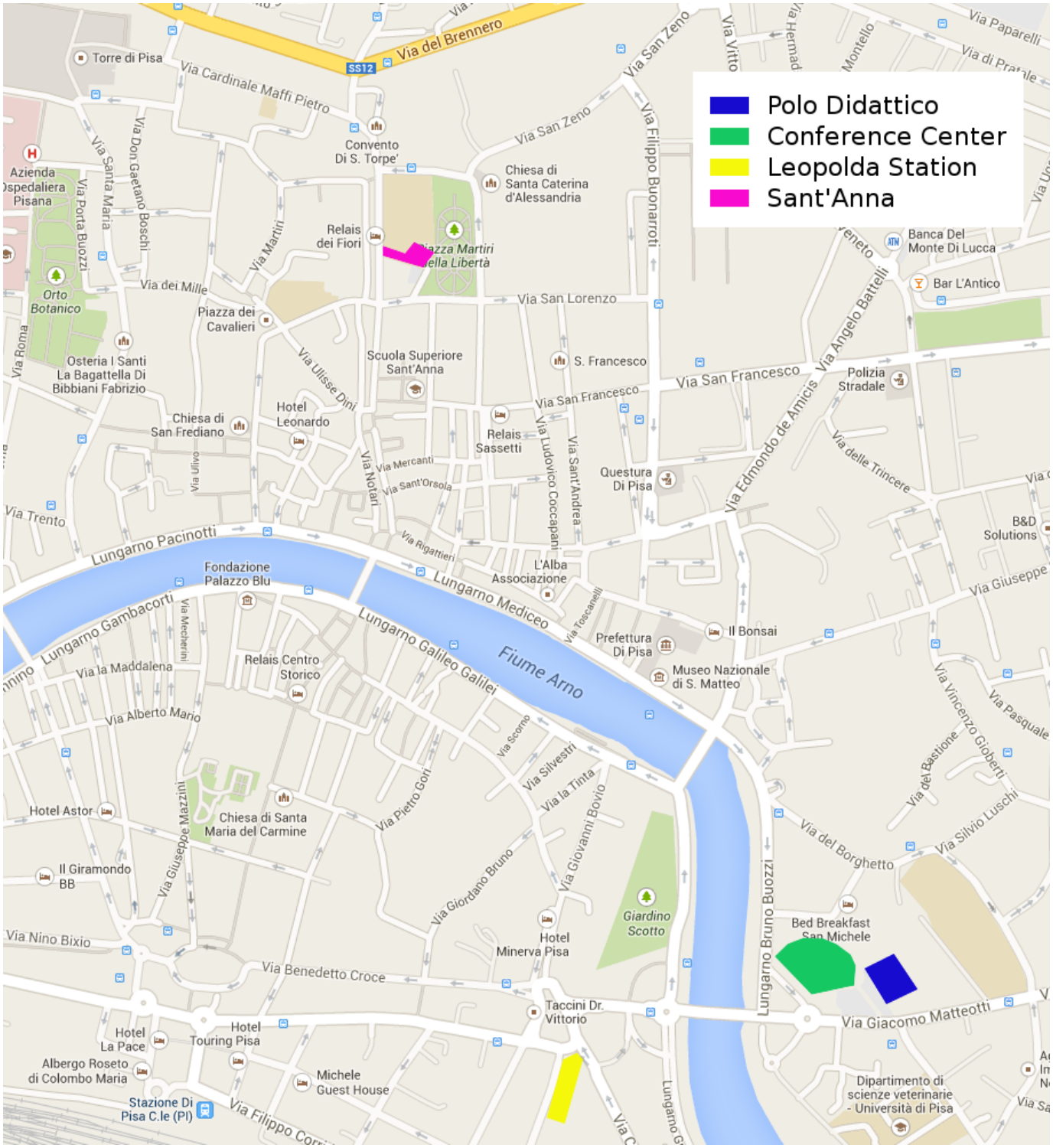
Internet

Participants from any eduroam-enabled institution should use the Eduroam service in order to obtain access to Internet. For participants without Eduroam access, there will be limited accounts for wireless Internet connection at the main venue of Polo didattico delle Piagge dell'Università di Pisa. You will need to have your own laptop in order to connect to the Internet. The username and password for the wireless Internet connection can be obtained by the IT desk which will be located next to the registration desk.

Information and messages

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

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 special issues, or in regular issues of the Journal of Computational Statistics & Data Analysis. Submissions for the CSDA should contain a strong computational statistics, or data analytic component. 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@dcs.bbk.ac.uk.

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

- Information about the CSDA special issues for the 2014-2015 can be found at:
http://www.cmstatistics.org/CSDASpecialIssues.php

SPONSORS

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

UNIVERSITÀ DI PISA
http://www.unipi.it

http://www.unisa.it

ELSEVIER
http://www.elsevier.com

http://economia.unipv.it/ric-az/misura/misura.html

http://www.sssup.it

ERCIM WG on Computational and Methodological Statistics
http://www.CMStatistics.org

ERCIM
European Research Consortium for Informatics and Mathematics
http://www.ercim.eu

Computational and Financial Econometrics
http://www.CFEnetwork.org

EXHIBITORS

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

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

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

Taylor & Francis (URL http://www.taylorandfrancis.com/)

ENDORSED SOCIETIES & GROUPS

ERCIM Working Group on *Computational and Methodological Statistics (CMStatistics)*

Computational and Financial Econometrics *(CFEnetwork)*

The Society for Computational Economics

International Statistical Institute

The Royal Statistical Society

The Italian Statistical Society

Contents

Saturday 6.12.2014 09:10-10:00 Room: Auditorium Chair: Manfred Deistler MISURA PRIN Project CFE-ERCIM Keynote talk 1

Dynamic sparsity modelling

Speaker: **Mike West, Duke University, United States** Jouchi Nakajima

The research links to some of the history of Bayesian sparsity modelling in multivariate time series analysis and forecasting, and then reviews recent modelling innovations for dynamic sparsity modelling using the concept- and resulting methodology- of dynamic latent thresholding. Recent applications of this general approach to dynamic sparsity modelling have had demonstrable success in inducing parsimony into dynamic model structures in time series analysis, improving model interpretations, forecast accuracy and precision- and, as a result, decisions reliant on forecasts- in a series of econometric and financial studies. Emerging applications in dynamic network modelling further highlight the potential and opportunity provided by the dynamic latent thresholding approach in various areas.

Saturday 6.12.2014 11:55-12:45 Room: Auditorium Chair: Jean-Marie Dufour CFE Keynote talk 2

Tests for explosive financial bubbles in the presence of non-stationary volatility

Speaker: **Robert Taylor, University of Essex, UK** David Harvey, Stephen Leybourne, Robert Sollis

The aim is to study the impact of permanent volatility shifts in the innovation process on the performance of the test for explosive financial bubbles based on recursive right-tailed Dickey-Fuller-type unit root tests proposed previously. We show that, in this situation, their supremum-based test has a non-pivotal limit distribution under the unit root null, and can be quite severely over-sized, thereby giving rise to spurious indications of explosive behaviour. We investigate the performance of a wild bootstrap implementation of their test procedure for this problem, and show it is effective in controlling size, both asymptotically and in finite samples, yet does not sacrifice power relative to an (infeasible) size-adjusted version of their test, even when the shocks are homoskedastic. We also discuss an empirical application involving commodity price time series and find considerably less emphatic evidence for the presence of speculative bubbles in these data when using our proposed wild bootstrap implementation of the test.

Saturday 6.12.2014 18:10-19:00 Room: Auditorium Chair: Erricos John Kontoghiorghes ERCIM Keynote talk 2

Maximin effects in inhomogeneous large-scale data

Speaker: **Nicolai Meinhausen, ETH, Switzerland**

Large-scale data are often characterised by some degree of inhomogeneity as data are either recorded in different time regimes or taken from multiple sources. We look at regression models and the effect of randomly changing coefficients, where the change is either smoothly in time or some other dimension or even without any such structure. Fitting varying-coefficient models or mixture models can be appropriate solutions but are computationally very demanding and often try to return more information than necessary. If we just ask for a model estimator that shows good predictive properties for all regimes of the data, then we are aiming for a simple linear model that is reliable for all possible subsets of the data. We propose a maximin effects estimator and look at its prediction accuracy from a theoretical point of view in a mixture model with known or unknown group structure. Under certain circumstances the estimator can be computed orders of magnitudes faster than standard penalised regression estimators, making computations on large-scale data feasible. We also show how a modified version of bagging can accurately estimate maximin effects.

Monday 8.12.2014 12:25-13:15 Room: Auditorium Chair: Monica Pratesi CFE-ERCIM Keynote talk 3

Recent advances in estimation of conditional distributions, densities and quantiles

Speaker: **Irene Gijbels, Katholieke Universiteit Leuven, Belgium**

Starting point is the estimation of a conditional distribution function, which is at the basis of many statistical problems. A conditional distribution function describes how the distribution of a variable of interest Y changes with the value of a covariate X (or a set of covariates). Depending on some prior knowledge on how the covariate X influences Y, one can pre-adjust the response observations for 'obvious' effects of the covariate, and as such provide opportunities to an improved estimation of the conditional distribution function. The developed techniques can be extended to a variety of other estimation settings, such as conditional quantile estimation and conditional copula estimation; as well as to more complex data settings (for example censored data).

Monday 8.12.2014 18:25-19:15 Room: Auditorium Chair: Helmut Luetkepohl University of Salerno CFE-ERCIM Keynote talk 4

Cluster-robust inference and the wild cluster bootstrap

Speaker: **James G. MacKinnon, Queens University, Canada**

Using cluster-robust standard errors is now standard in many areas of applied econometrics, because even very small levels of intra-cluster correlation can make conventional standard errors severely inaccurate when some clusters are large. However, using t statistics based on cluster-robust standard errors may not yield accurate inferences when the number of clusters is small and/or cluster sizes vary greatly. In such cases, it can be very attractive to use the wild cluster bootstrap, but just how that method is implemented matters. Studentized bootstrap intervals are easy to calculate but may not yield sufficiently reliable inferences. It is better to use restricted estimates in the bootstrap data generating process. This is easy to do when testing hypotheses, but it can be expensive to compute bootstrap confidence intervals in this way. Simulation evidence that illustrates the above points is presented. Modified wild cluster bootstrap procedures based on transforming the residuals is then discussed. These are similar to methods that are already used with the ordinary wild bootstrap for regression with heteroskedastic but independent disturbances. Unfortunately, these procedures can become computationally infeasible when some clusters are very large.

Saturday 6.12.2014 10:30 - 11:45 Parallel Session B – CFE

CS10 Room E2 QUANTILE REGRESSION APPLICATIONS IN FINANCE **Chair: Massimiliano Caporin**

C044: **Modeling and forecasting the range bipower variation conditional quantiles**
Presenter: **Giovanni Bonaccolto**, University of Padua, Italy
Co-authors: Massimiliano Caporin

The aim is focused on the Realized Range Volatility, an estimator of the quadratic variation of financial prices, taking into account the impact of microstructure noise and jumps in the context of high frequency data. In contrast to existing studies, the aim is to model and forecast the conditional quantiles of the Realized Range Volatility. Therefore, a quantile regression analysis is carried out, by using as predictors both the lagged values of the estimated volatility and some key macroeconomic and financial variables. Those give important information about the overall market trend and risk. In this way, and without distributional assumptions on the realized range innovations, it is possible to assess the entire conditional distribution of the estimated volatility. This is a critical issue for financial decision makers in terms of pricing, asset allocation and risk management. How the links among the involved variables change across the quantile levels is analyzed. Furthermore, a rolling analysis is performed in order to check how the relationships which characterize the proposed model evolve over time. The analysis is applied to 16 stocks of the U.S. market and the forecast goodness is validated by means of a suitable test statistic.

C715: **Dynamic model averaging for quantile regression**
Presenter: **Lea Petrella**, Sapienza University of Rome, Italy
Co-authors: Mauro Bernardi, Roberto Casarin

A general dynamic model averaging (DMA) approach is provided to sequential estimation of quantile regression models with time-varying parameters. We build a new sequential Markov-chain Monte Carlo (MCMC) algorithm to approximate the posterior distribution of models and parameters. The efficiency and the effectiveness of the proposed DMA approach and the MCMC algorithm is shown through simulation studies and applications to macro-economics and finance.

C412: **Spillover effect to bailout expectation: an empirical study of Denmark**
Presenter: **Massimiliano Caporin**, University of Padova, Italy
Co-authors: Francesco Ravazzolo, Paolo Santucci de Magistris

The spillover effect to bailout expectations of bond holders during the recent financial crisis is studied with a focus to Denmark. Up to the beginning of 2011, investors thought that Scandinavian governments were providing an implicit government guarantee for bank bond holders. This stems from the experience linked to the real estate bubble hitting the area in the late 1980's, where stock holders lost everything but bond holders were guaranteed. However, two small Danish banks defaulted in early 2011 and in that case bond holders were not rescued, and they were sharing losses. Starting from this evidence, we verify the spillover effect of this breaking event on the bailout expectations to Danish banks. We focus on a specific Danish institution and use non-Danish banks as control cases. In determining the change in bailout expectations we make use of different techniques, starting from cumulated abnormal return analyses on the CDS (that monitors bond holders risks), to linear models with a number of control covariates up to quantile regression methods. Results show evidence of a change in bailout expectations after 2011.

CS13 Room A2 BAYESIAN ECONOMETRICS **Chair: Richard Gerlach**

C484: **Volatility and quantile forecasts by realized stochastic volatility models with generalized hyperbolic distribution**
Presenter: **Toshiaki Watanabe**, Hitotsubashi University, Japan
Co-authors: Makoto Takahashi, Yasuhiro Omori

The realized volatility, which is the sum of squared intraday returns, is subject to the bias caused by microstructure noise and non-trading hours. The realized stochastic volatility model incorporates the stochastic volatility model with the realized volatility taking account of the bias. It is extended using the generalized hyperbolic skew Student's t-distribution as the return distribution conditional on volatility for quantile forecast such as value-at-risk and expected shortfall. It includes the Student's t and normal distributions as special cases. A Bayesian method via Markov chain Monte Carlo is developed for the analysis of the extended model, which makes it possible to estimate the parameters and forecast volatility and return quantile jointly by sampling them from their joint posterior distribution. Using daily returns and realized volatility for the S&P 500 stock index, the stochastic volatility and realized stochastic volatility models with the normal, Student's t and generalized hyperbolic skew Student's t-distributions are applied to volatility and quantile forecasts. The performance is compared using several backtesting procedures. The performance of the realized stochastic volatility model with the realized kernel calculated taking account of the bias caused by microstructure noise is also compared.

C499: **Cholesky realized stochastic volatility model with leverage**
Presenter: **Yasuhiro Omori**, University of Tokyo, Japan
Co-authors: Shinichiro Shirota, Hedibert Lopes, Hai Xiang Piao

Multivariate stochastic volatility models play important roles in financial applications, including asset allocation and risk management. However, estimation of multivariate stochastic volatility models has some difficulties such as the curse of dimension and preserving the positive definiteness of estimated covariance matrix. We consider the dynamic modelling of the Cholesky decomposition for high dimensional covariance matrices, where we incorporate stylized facts of financial markets such as leverage effects and correlations among volatilities. Further, the realized covariance is considered as another source of information for the true covariance matrix. A simple efficient estimation is proposed using MCMC algorithm.

C865: **Bayesian hierarchical spatial-temporal modeling**
Presenter: **Mike So**, The Hong Kong University of Science and Technology, China

Spatial time series are commonly observed in environmental sciences and epidemiological research where, under most circumstances, the time series are non-stationary. For better estimation and prediction, the temporal-variability must be captured in traditional spatial models. We propose a Bayesian hierarchical spatial-temporal model to describe the dependence of time series data on spatial locations while accounting for time series properties. The first layer of the hierarchical model specifies a measurement process for the observed spatial data. The second layer characterizes a latent mean process and a latent variance process. The hierarchical formulation concludes with a third layer of priors on parameters. A key idea is to model spatial and temporal dependence simultaneously. Statistical inference is performed by Markov chain Monte Carlo methods which involve spatial dependence parameters. The methodology is applied to simulated data and real environmental data for illustration.

CS22 Room Q2 DYNAMIC MODELING OF VARIANCE RISK PREMIA **Chair: Matthias Fengler**

C1281: **A dynamic partial equilibrium model for asset pricing with volatility risk premium and reference dependent preferences**
Presenter: **Maria Grith**, HU Berlin, Germany

The aim is to reconciliate the empirical findings about U-shaped and hump-shaped pricing kernels implied from index options and historical returns of the stock index by using the volatility risk premium as a state variable - proxy for the economic conditions - in a model with heterogeneous agents whose preferences are reference-dependent. In this model, a more pronounced humped-shaped pricing kernel is consistent with relatively more risk loving investors who are targeting the performance relative to the market index. In these periods the volatility risk premium is low and

it is plausible to assume that the low demand for protection against tail events makes it look less U-shaped. We try to endogenize the changes in the distribution of the reference points in a dynamic partial equilibrium model for asset pricing and link them to the dynamics of the volatility risk premium.

C1005: **An index of financial connectedness applied to variance risk premia**
Presenter: **Andrea Cipollini**, University of Palermo, Italy
Co-authors: Iolanda Lo Cascio, Silvia Muzzioli

The purpose is to construct an index of financial connectedness among France, Germany, UK, Switzerland and the Netherlands variance risk premia. The variance risk premium of each country stock market is measured by the difference between the (square) of implied volatility and expected realized variance of the stock market for next month. The total and directional indices of financial connectedness are obtained from the forecast error variance decomposition of a Vector Autoregressive Model, VAR, as recently suggested by Diebold and Yilmaz. While the authors main focus is on connectedness among financial returns, they base their analysis on a short memory stationary VAR. Given the long memory properties of the series under investigation, we base the computation of the moving average coefficients useful for the computation of variance decomposition by modeling a fractionally integrated Vector Autoregressive Model.

C1303: **Are variance risk premia affine functions in the underlying state variables?**
Presenter: **Matthias Fengler**, University of Sankt Gallen, Switzerland

In option pricing, variance risk premia are often assumed to be affine functions of the state variables. We estimate variance risk premia in the usual model-free way and use nonparametric modeling techniques to ask whether the assumption of affine variance risk premia is empirically supported. For this purpose, we build on recent theory for specification tests in additive models. Our results indicate that variance risk premia for the S&P500 may not be affine functions in the state variables.

CS34 Room G2 ECONOMETRICS OF ART MARKETS Chair: Douglas Hodgson

C586: **Innovation, experience and artists' age-valuation profiles: evidence from eighteenth-century rococo and neo-classical painters**
Presenter: **Douglas Hodgson**, UQAM, Canada
Co-authors: John Galbraith

Age-valuation profiles for artists' works are to some degree analogous to age earnings profiles for workers more generally, and are of interest for valuation of art objects as well as for what they reveal about artists' career dynamics. A number of writers have suggested that the school, or style, characterizing an artist's work may have implications for the peak of the age-valuation profile (the age of the artist at the time at which his or her most highly valued works were executed). In particular, shifts of style which change the relative valuation of innovation and originality, versus craftsmanship and experience, have been hypothesized. The second half of the 18th century provides an interesting potential test, because socioeconomic and political changes at this time were paralleled by revolutionary change in the fine arts, especially painting, witnessing major changes in subject manner and aesthetic style among prominent artists. We revisit the hypothesis using data on auction sales of works from this time, estimating hedonic age-valuation profiles using a novel data set. We pool our set of artists to estimate profiles for different birth cohorts, also using a specification where profiles shift continuously with year of birth. We also use dimension-reduction and model-averaging techniques to allow estimation of individual profiles for several artists, despite limited sample sizes.

C752: **Creative production and peer effects: evidence from the exodus of superstar painters from Paris**
Presenter: **Christiane Hellmanzik**, University of Hamburg, Germany

The purpose is to present evidence of peer effects in creative production using an IV estimation based on the exodus of artists from Paris due to the Second World War. If proximity of peers matters for the distribution of artistic knowledge and the production of high quality art, it can be expected that the departure of high quality peers from Paris has an impact on the value of artworks produced by artists who remain in Paris. The study is based on a global sample of 273 'superstars' of modern art born between 1800 and 1945. For this sample peer quality is measured using 34,141 auction results of paintings made by these artists and a historiometric measure of their prominence. The findings of this quasi-experiment confirm the OLS result that there are signifcant localised peer effects. That is, a higher quality peer group has a signifcant, positive impact on the quality of artworks produced by modern artists. This result is robust across different measures of peer quality.

C815: **Is art really a safe haven? Evidence from the French art market during WWI**
Presenter: **Geraldine David**, Universite Libre de Bruxelles, Belgium

The performance of art as an investment is a very controversial topic except in times of wars and financial distresses. During crisis art is often considered as a safe haven both by the scientific literature and the financial advisors. This vision of art is refined as a safe investment providing evidence that art has not always supplied a safe investing way during crisis. To do so it constructs an art price index for the French art market during the First World War and the postwar period in France (1911-1925). The work fills in a gap of the literature as the reactions of the art market during crisis times remain under-investigated. Most of the literature is dedicated to WWII during which art markets encountered a massive boom among occupied countries. The results show that this boom is mostly to be attributed to the particular occupation economy instaured by the Nazis rather than to the safe investment characteristic of art. Art underperformed gold, real-estate, bonds and stocks in terms of risk-return performances. Moreover several peculiarities of the market are investigated. Investors tended to prefer cheap artworks and old masters during WWI as these were less volatile at the time.

CS35 Room O2 ANALYSIS OF EXTREMES AND DEPENDENCE Chair: Artem Prokhorov

C307: **Geometric measure representations and exact distributions of extremes**
Presenter: **Wolf-Dieter Richter**, University of Rostock, Germany

Determining the probability $F(t)$ of the random event that the maximum of the components of an n-dimensional random vector X is smaller than a given value t can be considered as assigning the value $P^X(M(t))$ to the set $M(t) = \{x \in R^n : x_i < t, i = 1,...,n\}$ where P^X indicates the distribution of X, $F(t) = P^X(M(t))$. Geometric measure representations apply to the general problem of evaluating $P^X(M(t))$ with a high degree of accuracy. Evaluating the c.d.f. $F(t), t \in R$, becomes the more challenging the bigger the class of probability distributions is where the distribution P^X may come from. We will allow two types of generating dependencies among the components $X_i, i = 1,...,n$, of the vector X. A correlation structure is reflected within the class of elliptically contoured distributions by shape and orientation of the density level ellipsoids and will be generalized here in a certain sense by generalizing the density level sets when we introduce norm-contoured and star-shaped densities. It is further known from elliptically contoured distributions that uncorrelated coordinates are independent iff the density is generated by the specific Gaussian density generating function (dgf). This general relation will be proved to hold also for suitably adapted coordinates in a big family of star-shaped distributions. Deviating the dgf from the Gaussian one leads to additional dependencies.

C685: **Density forecasting of U.S. macroeconomic variables using a Gaussian copula model of cross-sectional and serial dependence**
Presenter: **Michael Smith**, University of Melbourne, Australia
Co-authors: Shaun Vahey

A Gaussian copula model with asymmetric margins is used to produce forecast densities with the scope for severe departures from symmetry. The univariate marginal models differ for each variable. A stochastic volatility model is used for variables which exhibit conditional heteroscedasticity,

and all variables are transformed using their empirical distribution functions. We link the copula data with a Gaussian copula multivariate time series model that captures both cross-sectional and (multivariate) serial dependence. Restrictions on the parameter space are outlined under which the series is strongly stationary and Markov, along with alternative representations using a latent Gaussian vector autoregression and a drawable vine copula. We examine the forecasting performance of the model for four U.S. macroeconomic variables using quarterly real-time data.

C988: **Fat tails and copulas: limits of diversification revisited**
Presenter: **Artem Prokhorov**, Concordia U and U Sydney, Australia
Co-authors: Rustam Ibragimov

The problem of portfolio risk diversification is considered in a Value-at-Risk framework with heavy tailed risks and arbitrary dependence captured by a copula function. Using power law $\mathbb{P}(X > x) \sim x^{-\alpha}$, where α captures heavy-tailedness, we explore whether benefits of diversification, represented by the ratio $\mathbb{P}(X_1 + X_2 > 2x)/\mathbb{P}(X_1 > x)$ being less than one, persist when X's are allowed to have extremely heavy tails ($\alpha < 1$) and various dependence structures. We show theoretically that for asymptotically large losses with the Farlie-Gumbel-Morgenstern copula, the threshold value of α, at which diversification stops being beneficial, is the same as for independent losses. We further show using simulations that the same is true for tail dependent Archimedean copulas such as Clayton and Gumbel but is not true for Student-t copulas, for which diversification remains optimal regardless of the degree of heavy-tailedness or strength of dependence.

CS50 Room P2 MONITORING MACRO-ECONOMIC IMBALANCES AND RISKS Chair: Gianluigi Mazzi

C175: **Assessing the anchoring of inflation expectations**
Presenter: **Till Strohsal**, Free University Berlin, Germany
Co-authors: Lars Winkelmann

A new approach to assess the degree of anchoring of inflation expectations is proposed. We extend the static setup of the predominant news regressions by introducing exponential smooth transition autoregressive dynamics. Our approach refers to globally stationary expectations and provides estimates of a market-perceived inflation target as well as the strength of the anchor that holds expectations at that target. A cross-country study based on a new data set of daily break-even inflation rates for the US, EMU, UK and Sweden shows that the degree of anchoring varies substantially across countries and expectations horizons.

C370: **Signaling twin crises: Estimating the nexus of banking and sovereign risk**
Presenter: **Gregor von Schweinitz**, Halle Institute for Economic Research, Germany
Co-authors: Peter Sarlin

Some countries in the Euro area are experiencing a twin crisis. Beyond banking crises being followed by sovereign debt crises, most recent concerns have related to repercussions of sovereign debt (un)sustainability on the risk of a banking crisis. Hence, the probability of one crisis is not only affected by the occurrence of another type of crisis, but already by the risk of them. Tapping into this feature, the aim is to build an early-warning model for twin crises. We model the probability of a banking crisis as a function of the vulnerability to sovereign debt crises and vice versa. Using different endogenous regressor models, first results suggest that there is a strong relation between the two probabilities. We also allow for estimating effects on triplet crises of banking, debt and currency calamities, which is illustrated with an application to emerging market economies at the end of the 20th century.

C724: **Melting down: systemic financial instability and the macroeconomy**
Presenter: **Manfred Kremer**, European Central Bank, Germany
Co-authors: Kirstin Hubrich, Philipp Hartmann, Robert J. Tetlow

The role of systemic financial instability in an empirical macro-financial model for the euro area is investigated, employing a richly specified Markov-Switching Vector Autoregression model to capture the dynamic relationships between a set of core macroeconomic variables and a novel indicator of systemic financial stress. We find that at times of widespread financial instability the macroeconomy functions fundamentally differently from tranquil times. Not only the variances of the shocks, but also the parameters that capture the transmission of shocks change regime, especially around times of high systemic stress in the financial system. In particular, financial shocks are larger and their effects on real activity propagate much more strongly during high-financial-stress regimes than during tranquil times. We find an economically important role of loan growth in the propagation of financial stress to the macroeconomy. We also show that prospects for detecting episodes of what we call financial fragility appear promising, although we argue that more research is required. We conclude that macroprudential and monetary policy makers are well advised to take these non-linearities into account.

CS68 Room B2 MULTIVARIATE TIME SERIES Chair: Marco Reale

C905: **Resampling and asymptotic test statistic distributions for portfolio selection**
Presenter: **Alessia Naccarato**, University of Roma Tre, Italy
Co-authors: Andrea Pierini

As investment risk can be assessed on the basis of volatility, the problem that generally arises in portfolio selection regards estimating the volatility of stock market return, which means estimating the variance and covariance matrix of returns for a number of shares. The length of the historical series becomes an aspect of particular importance in view of the fact that the volatility matrix must be estimated in order to estimate the returns and therefore that the unknown parameters for which CVAR-BEKK models are adopted are the elements of the volatility matrix. It is in fact hard to establish the minimum length T of the historical series of n shares required to ensure the asymptotic characteristics of the test statistics used for the verification of hypotheses on the parameters estimated. The purpose is to compare the theoretical distributions of the test statistics used to evaluate the significance of the parameters estimated by means of the CVAR-BEKK models with those obtained in a simulation procedure by means of resampling based on the historical series of some shares traded on the Italian market.

C908: **Fiscal foresight, limited information and the effects of government spending shocks**
Presenter: **Matteo Fragetta**, University of Salerno, Italy
Co-authors: Eamnuel Gasteiger

The aim is to quantify the impact of government spending shocks in the US. Thereby, we control for fiscal foresight, a specific limited information problem (LIP) by utilizing the narrative approach. Moreover, we surmount the generic LIP inherent in vector autoregressions (VARs) by a factor-augmented VAR (FAVAR) approach. We find that a positive deficit financed defence shock raises output by more than in a VAR (e.g. 2.61 vs. 2.04 for peak multipliers). Furthermore, our evidence suggests that consumption is crowded in. These results are robust to variants of controlling for fiscal foresight and reveal the crucial role of the LIP in fiscal VARs

C1197: **VZAR: an extension of the VAR model**
Presenter: **Marco Reale**, University of Canterbury, New Zealand
Co-authors: Granville Tunnicliffe Wilson, John Haywood

Vector autoregressive models rely on information provided by lagged values of the variables under consideration. Parsimony in the number of parameters and hence how far back we should go is an important issue in this context. By filtering the time series we can extend the information base to greater lags but still retain a small number of parameters.

CS71 Room I2 NONPARAMETRIC AND SEMIPARAMETRIC METHODS: RECENT DEVELOPMENTS Chair: Patrick Saart

C156: **Data-driven estimation of realized kernels under dependent microstructure noise and further analysis using Semi-FI-Log-ACD**
Presenter: **Chen Zhou**, University of Paderborn, Germany
Co-authors: Yuanhua Feng

Realized kernel (RK) is a model-free estimator of the integrated volatility (IV) based on high-frequency data, which is consistent in the presence of microstructure noise under given conditions. A crucial problem by applying the RK is the selection of the bandwidth. Most recently, an iterative plug-in (IPI) bandwidth selector for RK was proposed under independent microstructure noise. Our focus is to develop an IPI bandwidth selector for realized kernels under dependent microstructure noise. The proposed bandwidth selector is consistent under regularity assumptions. Its nice practical performance is illustrated by application to data of a few European companies within a period of several years. Furthermore, a non-normal extension of the recently proposed Semi-FI-Log-ACD is applied for modeling a smooth trend and long memory in the obtained RK. Forecasting of RK using bootstrap is investigated. Effect of the last financial crisis and the Euro-Zone crisis on the market volatility is also discussed briefly.

C598: **Financial applications of nonparametric methods for functional linear regression analysis**
Presenter: **Patrick Saart**, University of Canterbury, New Zealand

The objective is to investigate the use of nonparametric methods for functional linear regression (FLR) techniques in analyzing various issues, which are of empirical importance in finance. Let us recall firstly that for the FLR both the predictor and response are functions of some covariate usually but not necessarily represents a measure of time. One of the empirical problems that we intend to consider relates to the conditional correlation coefficients of financial assets. Since the beginning of the financial crisis in 2007, whether correlations between stocks and international equity markets increase during market downturns has been widely discussed by both practitioners and academics. Some researchers have shown that cross correlations between international equity markets do increase when global market becomes highly volatile. Analogously, various authors found that cross correlations only increase in strong bear market but not in bull market. This leads to modeling the conditional correlation functions conditionally upon either the market return or volatility. However, this type of analysis may fail to take into consideration the interdependence between market return and volatility. In our FLR analysis, the predictor is the market volatility, which replicates the news impact curve relationship, while the response is the conditional correlation written as function of market return. The second empirical problem we intend to study relates to the common dynamics in volatility.

C601: **Control function approach to weak instruments**
Presenter: **Namhyun Kim**, University of Konstanz, Germany

There are two well-known alternative methods to address endogeneity in the literature, Instrumental Variables (IV) estimation and Control Function (CF) approach, using instrumental variables. However, it is not an easy task to find a legitimate instrument satisfying the relevance condition where a correlation between an endogenous regressor and its instrument is strong. We propose to use the CF approach rather than IV estimation since the CF approach translates the weak IV problem into the multicollinearity one in a structural equation. The crucial underlying condition for CF approach is partitioning an endogenous explanatory variable into an exogenous (instrument) and an endogenous (disturbance term from a reduced equation) component. This partition allows us to control endogeneity by incorporating the disturbance term from the reduced equation in a structural equation. However, the weak correlation between the endogenous regressor and its instrument causes the multicollinearity problem since the variations of both endogenous regressor and disturbance term from the reduced equation are closely correlated. It is well known in the literature that the ridge type of penalised estimation method allows us to address the multicollinearity problem. Hence we aim to develop the ridge type of penalised estimation method based on the CF approach to address the weak IV issues in a linear equation case.

CS83 Room N2 ENERGY PRICE AND VOLATILITY MODELLING Chair: Helena Veiga

C147: **Do high-frequency financial data help forecast oil prices? The MIDAS touch at work**
Presenter: **Pierre Guerin**, Bank of Canada, Canada
Co-authors: Christiane Baumeister, Lutz Kilian

In recent years there has been increased interest in the link between financial markets and oil markets, including the question of whether financial market information helps forecast the real price of oil in physical markets. An obvious advantage of financial data in forecasting monthly oil prices is their availability in real time on a daily or weekly basis. We investigate the predictive content of these data using mixed-frequency models. We show that, among a range of alternative high-frequency predictors, cumulative changes in U.S. crude oil inventories in particular produce substantial and statistically significant real-time improvements in forecast accuracy. The preferred MIDAS model reduces the MSPE by as much as 28 percent compared with the no-change forecast and has statistically significant directional accuracy as high as 73 percent. This MIDAS forecast also is more accurate than a mixed-frequency real-time VAR forecast, but not systematically more accurate than the corresponding forecast based on monthly inventories. We conclude that typically not much is lost by ignoring high-frequency financial data in forecasting the monthly real price of oil.

C401: **Modelling crude oil price return volatility - level Nexus: a robust nonparametric approach**
Presenter: **Isabel Casas**, University of Southern Denmark, Australia
Co-authors: Sandy Suardi

The purpose is to document that crude oil price return volatility exhibits level dependence, that is return volatility is a function of oil price levels. Given that oil prices have undergone structural breaks, which may impact on the extent of volatility-level linkages, we adopt a robust nonparametric method of Casas and Gijbels that yields consistent volatility estimates in the presence of structural breaks in crude oil prices.Unlike parametric GARCH models, the nonparametric approach does not impose specific functional form on the volatility level dependence. Models that combine level dependence with GARCH specification or utilise the robust nonparametric method showcase higher out-of-sample volatility forecast accuracy than standalone GARCH or level dependent volatility models.

C699: **How does stock market volatility react to oil shocks?**
Presenter: **Andrea Bastianin**, Universita degli Studi di Milano, Italy
Co-authors: Matteo Manera

The impact of oil price shocks on U.S. stock market volatility is studied. The price of crude oil has traditionally been treated as exogenous with respect to macroeconomic aggregates; however, in recent years the consensus view is that the macroeconomy, the oil and the stock markets react to the same factors, such as global demand. Moreover, studies of the impact of oil shocks on real and financial variables have generally failed to distinguish the causes underlying oil price increases. We rely on the model of Kilian to derive three different oil shock variables (i.e. aggregate demand, oil-supply, and oil-demand shocks) and relate them to stock market volatility. Next, we estimate three bivariate structural VAR models, one for each oil price shock. Identification is achieved by assuming that the price of crude oil reacts to stock market volatility only with delay. This implies that innovations to the price of crude oil are not strictly exogenous, but predetermined with respect to the stock market. We show that volatility responds significantly to oil price shocks caused by sudden changes in aggregate demand, while the impact of shocks specific to the oil market is negligible.

CS88 Room C2 QUANTITIVE METHODS IN CREDIT RISK MANAGEMENT Chair: Jiri Witzany

C157: **Wrong-way risk - correlation coefficient calibration**
Presenter: **Jakub Cerny**, Charles University in Prague, Czech Republic
Co-authors: Jiri Witzany

Under the new Basel III banking regulation banks should include wrong-way risk (WWR) into the calculation of the credit valuation adjustment (CVA) of the OTC derivatives. WWR is the situation when the exposure to a counterparty is adversely correlated with the credit quality of that counterparty. Assuming the link between the interest rate swap (IRS) and the default time is represented by a Gaussian copula with a constant correlation coefficient then the WWR is expressed by this correlation coefficient. Because the observation of the default time means the dissolution of the company, the correlation can not be simply estimated using the observed data in contrast to the credit default swap (CDS) rate which directly corresponds to the intensity of default. Based on available daily Czech Republic government IRS and CDS rates we estimated the correlation using maximum likelihood method assuming that the systematic factor is governed by the AR(1) process, so we can decorrelate both time series. The results show that the correlation calibrated on the daily data is relatively high, and therefore the WWR should not be neglected in this case.

C1105: **Comparison of copulas in CDO valuation**
Presenter: **Marek Kolman**, University of Economics Prague, Czech Republic

Comparison of various copula models for CDO valuation is presented. We compare the classical approach with possibly more appropriate alternatives. These alternative models in our research scope cover double-t, NIG and stochastic Gaussian copulas. It is furthermore shown that a quasi-analytical solution in the double-t copula model exists even despite of the unfavourable properties of the t distribution. Finally, a calibration procedure is proposed and based on such a calibration comparison of results is presented.

C1030: **Default risk in agricultural lending, the effects of commodity price volatility and climate**
Presenter: **Carlos Castro**, Universidad del Rosario, Colombia
Co-authors: Karen Garcia

The aim is to propose and estimate a default risk model for agricultural lenders that explicitly accounts for two risks that are endemic to agricultural activities: commodity price volatility and climate. Results indicate that both factors are significant to explain the occurrence of default in the portfolio of a rural bank; however, their effect is smaller than expected. In addition, we design a portfolio credit risk model that serves as a quantitative tool to estimate the loss distribution and the economic capital for a rural bank. The estimated parameters of the default risk model, along with scenarios for the evolution of the risk factors are used to construct stress test with the portfolio of a rural bank.

CS90 Room M2 RISK ESTIMATION AND ESTIMATION RISK Chair: Jean-Michel Zakoian

C381: **Estimating the conditional VaR of a portfolio of multivariate GARCH returns**
Presenter: **Christian Francq**, University of Lille 3, France
Co-authors: Jean-Michel Zakoian

Developing the concept of risk in a multivariate framework is a complex task. We focus on the risk of a portfolio but, instead of considering the portfolio return as a single series, we model the joint distribution of the vector of asset returns involved in the portfolio as $\varepsilon_t = \Sigma_t(\theta_0)\eta_t$, where the volatility matrix $\Sigma_t(\theta_0) = \Sigma(\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, \theta_0)$ is specified as a general function of the past returns. Under the assumption that the distribution of η_t is spherical, the value-at-risk (VaR) of a portfolio $a'\varepsilon_t$ is equal to $\|a'\Sigma_t(\theta_0^*)\|$, where the parameter θ_0^* is expressed as function of the volatility coefficient θ_0 and a quantile of a component of the innovation process. A two-step method is proposed to successively estimate these quantities. Asymptotic confidence intervals for the VaR's of portfolios are derived. Monte-Carlo experiments and an empirical study illustrate these results.

C442: **Consistent estimation of the Value-at-Risk when the error distribution of the volatility model is misspecified**
Presenter: **Fedya Telmoudi**, University of Tunis, Tunisia
Co-authors: Christian Francq, Mohamed El Ghourabi

A two-step approach for conditional Value-at-Risk (VaR) estimation is considered. First, a generalized-quasi-maximum likelihood estimator (gQMLE) is employed to estimate the volatility parameter, then the empirical quantile of the residuals serves to estimate the theoretical quantile of the innovations. When the instrumental density h of the gQMLE is not the Gaussian density, or is not the true distribution of the innovations, both the estimations of the volatility and of the quantile are asymptotically biased. The two errors however counterbalance each other, and we finally obtain a consistent estimator of the conditional VaR. For different GARCH models, we derive the asymptotic distribution of the VaR estimation based on gQMLE. We show that the optimal instrumental density h depends neither on the GARCH parameter nor on the risk level, but only on the distribution of the innovations. A simple adaptive method based on empirical moments of the residuals allows us to infer an optimal element within a class of potential instrumental densities. Important asymptotic efficiency gains are achieved by using gQMLE when the innovations are heavy-tailed. We extended our approach to Distortion Risk Measure parameter estimation, where consistency of the gQMLE is proved. Simulation experiments and a real case study are provided.

C1113: **Risk measure inference**
Presenter: **Christophe Hurlin**, University of Orleans, France

A widely applicable bootstrap-based test of the null hypothesis of equality of two firms' Risk Measures (RMs) at a single point in time is proposed. The test can be applied to any market-based measure. In an iterative procedure, we can identify a complete grouped ranking of the RMs, with particular application to finding buckets of firms of equal systemic risk. An extensive Monte Carlo simulation shows desirable properties. We provide an application on a sample of 94 U.S. financial institutions using the CoVaR, MES and %SRISK, and conclude only the %SRISK can be estimated with enough precision to allow for a meaningful ranking.

CS61 Room D2 FILTERS WAVELETS AND SIGNALS I Chair: Stephen Pollock

C871: **Mathematical framework for pseudo-spectra of linear stochastic difference equations**
Presenter: **Marcos Bujosa**, Universidad Complutense de Madrid, Spain
Co-authors: Andres Bujosa, Antonio Garcia-Ferrer

Although spectral analysis of stationary stochastic processes has solid mathematical foundations, this is not always so for the non-stationary case. We establish a sound mathematical framework for the spectral analysis of non-stationary solutions of (time invariant) linear stochastic difference equations. To achieve it, the classical problem is embedded in a wider framework, the Rigged Hilbert space; the Fourier Transform is extended, and a new Extended Fourier Transform pair pseudo-covariance function/pseudo-spectrum is defined. Our approach is an extension proper of the classical spectral analysis, within which the Fourier Transform pair auto-covariance function/spectrum is a particular case, and consequently spectrum and pseudo-spectrum coincide when the first one is defined.

C560: **The relationship of simple sum and Divisia monetary aggregates with real GDP and inflation: a wavelet analysis for the US**
Presenter: **Michael Scharnagl**, Deutsche Bundesbank, Germany
Co-authors: Martin Mandler

Wavelet analysis is applied to compare the relationship between simple sum and Divisa monetary aggregates with real GDP and CPI inflation for

the U.S. using data from 1967 to 2013. Wavelet analysis allows to account for variations in the relationships both across the frequency spectrum and across time. While we find evidence for a weaker comovement of Divisia compared to simple sum monetary aggregates with real GDP the relationship between money growth and inflation is estimated to be much tighter between Divisia monetary aggregates and CPI inflation than for simple sum aggregates, in particular at lower frequencies. Furthermore, for the Divisia indices for broader monetary aggregates (M2, M2M, MZM) we estimate a stable lead before CPI inflation of about four to five years.

C773: **Generalized linear dynamic factor models**

Presenter: **Manfred Deistler**, Vienna University of Technology, Austria

Co-authors: Brian Anderson, Alexander Braumann, Elisabeth Felsenstein, Lukas Koelbl

Questions of identifability and estimation of generalized linear dynamic factor models (GDFM's) are considered. Our particular emphasis is on the case where the dimension of the dynamic factors is strictly smaller than the dimension of the static factors. As has been shown, in this case, generically, the static factors have an AR representation. In this case estimation via Yule-Walker equations provides a numerically effcient parameter estimation procedure, however, the Yule-Walker equations may have non-unique solutions. In such a case uniqueness and thus identifability can be obtained by prescription of appropriate column degrees. We discuss model selection procedures for estimation of the dimensions of the static and the dynamic factors, respectively, and of the appropriate column degrees of the autoregressive polynomial. Based on these integer-valued parameters, we describe a Yule-Walker based iterative estimation procedure for the autoregressive parameters. In this contribution we also derive analogous results for the ARMA or state space case. In particular we describe a realization procedure for obtaining the system parameters from the transfer function. Along such a realization procedure, a subspace estimation algorithm is discussed: We introduce an estimation procedure for the state obtained by regressing selected future outputs on selected fast outputs.

Saturday 6.12.2014 10:30 - 12:35 Parallel Session D – ERCIM

ES19 Room N1 DEPENDENCE MODELS AND COPULAS: THEORY I **Chair: Fabrizio Durante**

E476: **Estimation of a conditional copula under various settings**
Presenter: **Irene Gijbels**, Katholieke Universiteit Leuven, Belgium

When modeling the dependence structure between two random variables, given a covariate, a conditional copula function is used. We first briefly discuss nonparametric estimators of a conditional copula in a general setting. We then consider a special setting in which the covariate only influences the marginal distributions, and not the dependence structure itself. This simplifying assumption is often encountered, among others in pair-copula constructions. We study estimation of the copula under this simplifying assumption. Several estimation methods are discussed. We establish asymptotic properties, and investigate the finite-sample performances via a simulation study. Illustrations with real data examples are provided.

E540: **Some remarks on singular components of copulas**
Presenter: **Wolfgang Trutschnig**, University of Salzburg, Austria

The aim is to focus on some questions concerning singularity aspects of two-dimensional copulas. Motivated by the problem of maximizing the probability of joint default of two assets it will, firstly, be shown how much mass a copula can concentrate on the graph of a given (non-singular) transformation $T : [0,1] \to [0,1]$. Secondly, by constructing copulas $A_T^\star$ with full support $[0,1]^2$ for which all conditional distribution functions $y \mapsto F_x^{A_T^\star}(y)$ are continuous, strictly increasing and have derivative zero λ-almost everywhere, we will illustrate how surprisingly singular (from the purely analytical point of view) copulas and their conditional distribution functions may be. Finally, some singularity properties of Archimedean copulas will be discussed.

E1090: **On tail dependence coeffcients of transformed multivariate Archimedean copulas**
Presenter: **Elena Di Bernardino**, CNAM, France
Co-authors: Didier Rulliere

The impact of a class of transformations of copulas in their upper and lower multivariate tail dependence coefficients is presented. In particular we focus on multivariate Archimedean copulas. In the first part, we calculate multivariate tail dependence coefficients when the generator of the considered copula exhibits some regular variation properties, and we investigate the behaviour of these coefficients in cases that are close to tail independence. In the second part, we analyse the impact in the upper and lower multivariate tail dependence coefficients of a large class of transformations of dependence structures. In the third part, we show the utility of using transformed Archimedean copulas, as they permit to build Archimedean generators exhibiting any chosen couple of lower and upper tail dependence coefficients. The interest of such a study is also illustrated through applications in bivariate settings.

E246: **Two nonparametric tests of copula constancy in multivariate time series**
Presenter: **Ivan Kojadinovic**, University of Pau, France
Co-authors: Axel Buecher, Jean-Francois Quessy, Tom Rohmer, Johan Segers

Two classes of tests for change-point detection designed to be particularly sensitive to changes in the cross-sectional dependence of multivariate time series will be presented. The first one is based on the comparison of empirical copulas, while the second one focuses on differences of (multivariate versions) of sample Spearman's rhos. The asymptotic validity under the null of both approaches will be briefly discussed, practical recommendations will be made, and illustrations on financial data will be given. Both classes of procedures are implemented in the R package npcp.

E243: **Copula regression spline models for binary outcomes**
Presenter: **Giampiero Marra**, University College London, United Kingdom
Co-authors: Rosalba Radice, Malgorzata Wojtys

A framework is introduced for estimating the effect that a binary treatment has on a binary outcome in the presence of unobserved confounding. The method is applied to a case study whose aim is to estimate the effect of private health insurance on health care utilization. Unobserved confounding arises when variables which are associated with both treatment and outcome are not available. Also, treatment and outcome may exhibit a dependence that cannot be modeled using a linear measure of association, and observed confounders may have a non-linear impact on the responses. The problem of unobserved confounding is addressed using a two-equation structural latent variable framework, non-linear dependence between treatment and outcome is dealt with by using copula functions, and covariate-response relationships are flexibly modeled using a spline approach. Related model fitting and inferential procedures are developed and asymptotic arguments presented. The findings of our empirical analysis suggest that the issues discussed in this paper are present when estimating the impact of private health insurance on health care utilization, and that useful insights can be gained by using the proposed framework.

ES27 Room A1 OUTLIERS AND EXTREMES IN TIME SERIES **Chair: Roland Fried**

E1133: **Recent developments of the COGARCH model**
Presenter: **Claudia Klueppelberg**, Technische Universitaet Muenchen, Germany

In order to model high-frequency data in a realistic way a continuous-time GARCH (COGARCH) model with jumps was suggested in 2004. Since then various publications have investigated probabilistic and statistical properties of the model. For instance, the stationary COGARCH exhibits heavy tails and clusters in its extremes, thus retaining the attractive features of the GARCH model. Recently we have extended this model to allow for asymmetry and long range dependence in the data. We will present these models and discuss some of their properties.

E104: **A robust estimator of the tail index under short range dependence**
Presenter: **Svenja Fischer**, Ruhr-University Bochum, Germany
Co-authors: Roland Fried, Martin Wendler

Motivated by applications in hydrology the robust Generalized Median (GM) estimator is investigated under short range dependence. Needing results concerning the asymptotic distribution a central limit theorem is developed, which is shown for the broad class of Generalized L-Statistics, the GM estimator being a special case of it. In a concrete application in the hydrological context the improvement of the estimation using an AR(1)-process compared to the classical methods under independence is shown. Therefore monthly maximum discharges of gauges in Germany with an outlier caused by a large flood are investigated using the Peak over Threshold approach. A central question in this case is the estimation of the tail index. Finally some simulations are presented where the GM-estimator is compared to the Maximum-Likelihood estimator under the aspects robustness and efficiency.

E818: **Nonparametric two-sample tests based on divergences with special attention to extreme values**
Presenter: **Max Wornowizki**, TU Dortmund University, Germany

The concept of f-divergences provides a rich set of distance-like measures between pairs of distributions. Divergences do not focus on certain moments of random variables, but rather consider discrepancies between the corresponding probability density functions. Therefore, two-sample

tests based on these measures can detect arbitrary alternatives when testing the equality of the distributions. We propose a new nonparametric estimator, which performs stable and quite well in comparison to several existing non- and semiparametric divergence estimators. In addition, we construct tests for the homogeneity of two samples. These tests are compared to an asymptotic divergence test as well as to several traditional parametric and nonparametric procedures under different distributional assumptions and alternatives in simulations. According to the results, divergence based methods detect discrepancies between distributions more often than traditional methods if the distributions do not differ in location only. We give special attention to situations with outliers as well as data from generalised extreme value distributions.

E1236: **Statistical models for dynamics in extreme value processes**
Presenter: **Bernhard Spangl**, University of Natural Resources and Life Sciences - Vienna, Austria
Co-authors: Sascha Desmettre, Daria Pupashenko, Peter Ruckdeschel

A challenge in dealing with extreme events in river discharge data is to capture well time dynamics of these extremes, in particular in the presence of seasonal effects and trends. We study three different approaches to model time-dependent extremal behavior: dynamics introduced (a) by a state-space model, (b) by a GLM with GPD marginals (and previous extremal events as regressors), and (c) a particular copula-based autoregressive model with GPD marginals where the dependence is data driven. Each of the models is fit against data, and from the fitted data, we simulate corresponding paths according to the respective fitted models. At this simulated data, the respective dependence structure is analyzed in copula plots and judged against its capacity to fit the corresponding interarrival distribution.

E716: **TCLUST: Robust clustering based on trimming and restrictions**
Presenter: **Agustin Mayo-Iscar**, Universidad de Valladolid, Spain
Co-authors: Luis Angel Garcia-Escudero, Alfonso Gordaliza, Carlos Matran-Bea

TCLUST is a robust model based clustering approach. It gets a robust behavior by using jointly trimming and restrictions. Restrictions allow us to obtain a well posed estimation problem avoiding not only singularities, but also spurious solutions. The basic approximation assumes a normal model in each population, but the methodology also works by using skew-symmetric models in each population with the benefit of an increase of flexibility. By including additional restrictions in the eigenvectors, and in general, by restricting the covariance matrices is possible to increase the efficiency when prior information is available or to adapt the modelling flexibility to the sample size. Trimming and restrictions also work for clustering around linear subspaces. The joint application of these tools in cluster weighted modelling gets solutions which avoid, in a natural way, artifacts identified when traditional robust approaches were applied. This idea also works when groups are distributed around curves from a known family. A typical drawback of this methodology is the election of input parameters: the number of groups, the level of trimming and the level for the restrictions. New approaches allow, even, in an automatic way, to obtain sensible values for these parameters.

ES35 Room B1 THRESHOLD SELECTION IN STATISTICS OF EXTREMES Chair: Ivette Gomes

E259: **Computer intensive procedures in threshold selection**
Presenter: **Manuela Neves**, ISA and CEAUL, Portugal
Co-authors: Ivette Gomes, Fernanda Figueiredo, Dora Prata Gomes

In Extreme Value Analysis and whenever dealing with large values, there are a few primordial parameters, among which we mention the extreme value index, ξ, measuring the right tail-weight and the extremal index, θ, related to the degree of local dependence in the extremes of a stationary sequence, roughly defined as the reciprocal of the expectation of the extremes' duration. Most semi-parametric estimators of these parameters show the same type of behaviour: nice asymptotic properties, but a high variance for small values of k, the number of upper order statistics used in the estimation, a high bias for large values of k, and a very sharp plot for the Mean Squared Error, needing an adequate choice of k. Resampling computer intensive methodologies, like the jackknife and the bootstrap, are important tools for a reliable semi-parametric estimation of parameters of extreme or even rare events. After a brief introduction of some estimators of the aforementioned parameters and their asymptotic properties, we review and compare algorithms for the choice of k and the adaptive estimation of ξ and θ. Some results from a simulation study as well as applications to real data sets are shown.

E542: **Selection of regression quantile threshold**
Presenter: **Jan Picek**, Technical University of Liberec, Czech Republic

The contribution deals with estimators of extreme value index based on weighted averaged regression quantiles in the linear regression model. The weighted averaged quantiles can be seen as a possible generalization of the quantile idea. We consider a class of smooth functionals of the tail quantile function as a tool for the construction of estimators in the linear regression context. As in the location model, the estimation of the extreme value index usually performed on the basis of the largest regression quantiles exceeding a high threshold. The question that has been often addressed in many applications is the selection of the threshold. We focus on the resampling-based methods for a choice of the threshold in the considered linear regression model.The used methods will be demonstrated by numerical illustrations.

E600: **Extreme value mixture modelling**
Presenter: **Carl Scarrott**, University of Canterbury, New Zealand
Co-authors: Alfadino Akbar

Extreme value mixture models have received a lot of recent attention for automated estimation of the threshold beyond which the classical extreme value tail models provide a good approximation. The aim is to outline the key features of such models, including an R package called evmix which is available on CRAN. The package implements most of the mixture models defined in the literature, including profile likelihood estimation for the threshold and likelihood estimation for the other parameters. Advice for developers and users of such models based on recent simulation studies are also provided. Results are presented from a newly developed mixture model with a nonparametric density estimator below the threshold, which has the potential to be easily extended to incorporate continuity/derivative constraints as well as to nonstationary problems.

E741: **Threshold selection and tail least squares type estimators**
Presenter: **Margarida Brito**, Universidade Porto - ADFCUP-CMUP, Portugal
Co-authors: Ana Cristina Moreira Freitas

Several statistical analyses and methods require the selection and/or the estimation of a threshold, which is a delicate issue of the statistical modelling. When dealing with extreme events, we are faced with the study of properties of the tails of a distribution, defined via appropriate thresholds. We consider the Pareto-type distribution family and the problem of choosing the number of upper order statistics used in the estimation of the tail index and related parameters via least squares type estimators, with particular attention to the geometric-type estimator. The selection of the upper order statistics is made following different approaches, considering, in particular, an heuristic method specially adapted to this estimator. A simulation study is performed in order to analyse and compare the different procedures.

E778: **Extreme value analysis for truncated and non-truncated Pareto-type distributions**
Presenter: **Jan Beirlant**, KULeuven, Belgium
Co-authors: Isabel Fraga Alves, Ivette Gomes, Mark Meerschaert

A previous work derived the maximum likelihood estimator for the tail index of a truncated Pareto distribution with right truncation point T. Authors discuss and compare the use of the Hill and the maximum likelihood estimator under truncation in some practical settings. The Hill estimator is then treated as a limit case by letting T tend to infinity. The problem of extreme value estimation under right truncation was also

considered, proposing a similar estimator for the tail index and extending the estimator from Aban et al. introducing trimming of the top extreme order statistics. Other work discusses such estimation problems from a risk management perspective. In practice one does not always know if the distribution is truncated or not, which leads to the study of the properties of these estimators in case of a general class of Pareto-type distributions, both truncated and non-truncated. We also study the estimation of extreme quantiles and tail probabilities within this setting.

ES48 Room D1 HANDLING NUISANCE PARAMETERS: ADVANCES TOWARDS OPTIMAL INFERENCE Chair: Ioannis Kosmidis

E027: **Profile-score adjustments for incidental-parameter problems**
Presenter: **Koen Jochmans**, Sciences Po, France

Adjustments to the profile score are considered to deal with incidental-parameter bias in the estimation of common parameters from few observations on a large number of strata. The adjustments are based on a calculation of the profile-score bias evaluated at maximum-likelihood estimates of the nuisance parameters. If this bias does not depend on the incidental parameters, the incidental-parameter problem can be eliminated. More generally, inference based on adjusted profile scores leads to estimators that have general applicability, suffer from smaller bias than maximum likelihood, and exhibit desirable properties under rectangular-array asymptotics. The adjustment can be iterated, reducing the bias further.

E028: **Incidental parameter bias in panel quantile regressions**
Presenter: **Martin Weidner**, University College London, United Kingdom
Co-authors: Manuel Arellano

Linear quantile regression (QR) estimators are studied in panel data settings with fixed effects. The estimation error in the fixed effects causes an incidental parameter problem in the parameters of interest, and we work out the first order asymptotic bias under an asymptotic where both N and T grow to infinity. This leading incidental parameter bias is of order $1/T$, analogous to the situation in non-linear fixed effect panel models with smooth objective function. The key technical challenge in deriving our result is that the QR objective function is non-smooth, rendering the existing large T asymptotic bias results in the panel literature non-applicable. We provide analytic and Jackknife bias corrected estimators and study their performance in Monte Carlo simulations, and in an application to educational achievement of US high-school students.

E067: **Accurate likelihood inference with many nuisance parameters**
Presenter: **Nicola Sartori**, University of Padova, Italy

Frequentist inference on parameters of interest in models with many nuisance parameters, such as fixed effects panel data models, is reviewed. In particular, we describe two different methods based on the likelihood, which are the modified profile likelihood and the integrated likelihood. For the latter, the nuisance parameters are eliminated through integration using a suitable weight function. Strengths and weaknesses of each method are considered by means of theoretical results, when available, and simulation studies. For the special case of panel data models, the two methods are applied to static and dynamic models, both for continuous and discrete responses.

E074: **Bias corrected parametric bootstrap**
Presenter: **Ruggero Bellio**, University of Udine, Italy

Inference on a scalar parameter based on a parametric bootstrap of the signed root likelihood ratio statistic is considered. The approach falls within the prepivoting perspective, where parametric bootstrapping is performed for fixed value of the parameter of interest. We investigate the properties of the method with a high-dimensional nuisance parameter, in a two-index asymptotic setting. We also propose a bias corrected version, where simulation is performed after correcting for the bias of the constrained estimator of the nuisance parameter. Several alternative bias correction methods can be used for the task. Simulation results show the improved performance of the methodology.

E604: **Effects of bias on inference in the presence of nuisance parameters: case studies and questions**
Presenter: **Ioannis Kosmidis**, University College London, United Kingdom

The aim is to provide a unifying framework of the still-expanding repository of methods that have been developed to reduce bias in the estimation of parametric models. Case-studies on Beta regression and meta-regression are presented where the bias on the regression parameters is negligible or zero but where the reduction of bias of the nuisance dispersion or precision parameters can lead to a significant performance gain in first-order inference. Furthermore, we present a bias correction scheme that seems to recover the consistency of the MLE for the common log-odds ratio in a binary matched pair setting. The work concludes with discussion and open questions.

ES63 Room L1 MIXTURE MODELS FOR MODERN APPLICATIONS I Chair: Geoff McLachlan

E763: **A mixture model for the measurement of resonance production in heavy-ion collisions**
Presenter: **Angela Montanari**, Alma mater studiorum-Universita di Bologna, Italy
Co-authors: Daniela Giovanna Calo, Francesca Bellini, Rosario Nania

The measurement of resonance production in heavy-ion collisions provides information on the properties of the hadronic medium. The K^{*0} resonance is particularly interesting as it is a good candidate to study rescattering and regeneration effects in the expanding hadronic medium. After background subtraction, the K^{*0} signal peak appears sitting on the top of a residual background. The overall shape is usually fitted with a function that is the sum of a rescaled relativistic Breit-Wigner density (for the signal) and a polynomial function (for the noise). The K^{*0} raw yield is obtained as the integral of the difference between the two functions in a proper interval. The data are reported in binned form and truncation is required in order to prevent nearby particles to affect the signal. The problem could be alternatively rephrased as a mixture model problem. The data can be thought of as coming from two subpopulations: the real resonances which form the signal, coming with a certain probability from a relativistic Breit-Wigner, and the random combinations which form the noise, arising from a second distribution. The mixture will be fitted resorting to an EM algorithm which makes allowance for binned and truncated data and considers different component shapes.

E193: **Extending mixtures of factor models using the restricted multivariate skew-normal distribution**
Presenter: **Tsung-I Lin**, National Chung Hsing University, Taiwan
Co-authors: Geoffrey J. McLachlan, Sharon X. Lee

The mixture of factor analyzers (MFA) model provides a powerful tool for analyzing high-dimensional data as it can reduce the number of free parameters through its factor-analytic representation of the component covariance matrices. The MFA model is extended to incorporate a restricted version of the multivariate skew-normal distribution for the latent component factors, called mixtures of skew-normal factor analyzers (MSNFA). The proposed MSNFA model allows us to relax the need of the normality assumption for the latent factors in order to accommodate skewness in the observed data. The MSNFA model thus provides an approach to model-based density estimation and clustering of high-dimensional data exhibiting asymmetric characteristics. A computationally feasible ECM algorithm is developed for computing the maximum likelihood estimates of model parameters. The potential of the proposed methodology is exemplified using both real and simulated data, and the results are compared with those obtained from fitting the MFA model.

E494: **Model-based cluster and discriminant analysis for functional data**
Presenter: **Faicel Chamroukhi**, University of Toulon, France

Finite mixture models are one of the most popular and successful approaches in model-based cluster and discriminant analysis. Generally used in multivariate analysis of fixed-dimensional vectors, their flexibility has lead to extensions to the analysis of functional data. We give a brief

background on finite mixtures for functional data. We present some recent advances in fitting regression mixtures, namely with unknown number of components. We also present recent extensions to hidden process regression mixtures for non-stationary functional data. The developed expectation-maximization (EM) algorithms are illustrated on simulated data and various real data.

E621: **Modeling multivariate longitudinal data in the presence of unobserved heterogeneity**
Presenter: **Laura Anderlucci**, University of Bologna, Italy
Co-authors: Cinzia Viroli

A novel approach for modeling multivariate longitudinal data in the presence of unobserved heterogeneity is proposed for the analysis of the Health and Retirement Study (HRS) data. Our proposal can be cast within the framework of linear mixed models with discrete individual random intercepts. This differs from the standard formulation in that the proposed Covariance Pattern Mixture Model (CPMM) does not require the usual local independence assumption; therefore, it is able to simultaneously model the heterogeneity, the association among the responses and the temporal dependence structure. We focus on the investigation of temporal patterns related to the cognitive functioning in retired American respondents, aiming to understand whether it can be affected by some individual socioeconomic characteristics and whether it is possible to identify some homogeneous groups of respondents that share a similar cognitive profile, so that opportune government policy interventions can be addressed. Results identify three homogeneous clusters of individuals with specific cognitive functioning, consistent with the class conditional distribution of the covariates. The flexibility of CPMM allows for a different contribution of each regressor on the responses according to group membership. In so doing, the identified groups receive a global and precise phenomenological characterization.

E1114: **Statistical quantification of genomic tumoral alterations with a mixture model**
Presenter: **Christine Keribin**, Universite Paris Sud, France
Co-authors: Yi Liu, Yves Rozenholc

The study of genomic DNA alterations (recurrent regions of alteration, patterns of instability) contributes to tumor classification, and becomes of great importance for the personalization of cancer treatments. The use of Single-Nucleotide Polymorphism (SNP) arrays or of New Generation Sequences (NGS) techniques allows the simultaneous estimation of segmented copy number (CN) and B-allele frequency (BAF) profiles along the whole genome. In this context, the GAP method was previously proposed, based on pattern recognition with (BAF, CN) maps to detect genotype status of each segment in complex tumoral genome profiles. It takes into account the fact that the observations on these maps are necessarily placed on centers that depend -up to a proper scaling of the CN- only on the unknown proportion of non-tumoral tissue in the sample. Being deterministic and manually tuned, this method appears sensitive to noise. To overcome this drawback, we set a mixture model, allowing the automatic estimation of the proportion of non-tumoral tissue and the test of genotype for each segment along the whole genome. We present the model, its estimation with an adapted EM algorithm and results on prostate carcinomas.

ES67 Room C1 SMALL AREA ESTIMATION Chair: Domingo Morales

E194: **Small area models for estimating temporal mortality trends by age and region**
Presenter: **Tomas Goicoa**, Universidad Publica de Navarra, Spain
Co-authors: M. Dolores Ugarte, Jaione Exteberria, Ana F. Militino

The availability of high quality mortality registers has brought about a development of statistical methods to study the spatio-temporal distribution of mortality or incidence risks or rates. Typically, maps displaying how spatial patterns evolve with time, and temporal trends for different regions are provided as final outputs. Classical approaches consider age-standardised mortality rates under the assumption that all age groups are equally affected by the disease. Consequently, a single measure is provided for the whole region and all age groups. However, this may not be realistic in practice and age-specific mortality rates by region should be considered. Several models are fitted to provide mortality rates for the different age groups within a region throughout the years. As a result, age-region specific mortality trends are obtained. Results will be illustrated with Spanish prostate cancer mortality data during the period 1986-2010.

E515: **A multivariate model for the estimation of poverty gap in small areas**
Presenter: **Enrico Fabrizi**, Catholic University of the Sacred Hearth, Italy
Co-authors: Maria Ferrante, Carlo Trivisano

The estimation of the relative median at-risk-of-poverty gap for the health districts of the administrative regions Emilia-Romagna and Tuscany, Italy, is considered. Direct estimates for this parameter, which is calculated only on the incomes of the poor have unacceptably large variances for small and moderate area-specific sample sizes. We make use of area level models, in which the design-based estimates are improved with the help of models that establish a connection among the underlying area parameters and auxiliary information accurately known for each district. A typical problem when using this type of models is that of finding adequate predictors for the poverty gap. To overcome this problem we introduce a multivariate model in which the poverty gap is considered jointly with the at-risk-of-poverty rate and the Gini index. The joint model relies on the log-normality assumption; we checked the robustness of estimators obtained as posterior summaries with respect to this distributional assumption. We use data from the EU-SILC survey, complemented with auxiliary information from administrative sources in the framework of small area estimation. We adopt a hierarchical Bayes approach to inference assisted by MCMC integration.

E652: **The empirical best predictor in the two-fold nested error regression model**
Presenter: **Domingo Morales**, University Miguel Hernandez of Elche, Spain
Co-authors: Yolanda Marhuenda, Isabel Molina, Jon Rao

Nonlinear small area population parameters are proposed to be estimated by using empirical best predictors (EBP) based on a two-fold nested error model. The main interest is the estimation of poverty indicators. It is given a parametric bootstrap method to estimate the mean squared error of the EBPs. Some simulation studies show that the EBPs based on two-fold nested error models are more flexible than EBP based on one-fold nested error models. They are also quite competitive with respect to alternative methods appearing in the literature. We also present an application to real data from the Spanish survey on income and living conditions.

E673: **The use of Big Data as covariates in area level small area models**
Presenter: **Monica Pratesi**, University of Pisa, Italy
Co-authors: Fosca Giannotti, Caterina Giusti, Stefano Marchetti, Dino Pedreschi

The timely, accurate monitoring of social indicators, such as poverty or inequality, at a fine grained spatial and temporal scale is a challenging task for official statistics, albeit a crucial tool for understanding social phenomena and policy making. Big Data sensed from the digital breadcrumbs that humans leave behind in their daily activities, mediated by the Information Communication Technologies, provide accurate proxies of social life. Social data mining from these data, coupled with advanced model-based techniques for fine-grained estimates, have the potential of providing us with a novel microscope for understanding social complexity. We propose a model based area level approach that uses Big Data as auxiliary variables to estimate poverty indicators for the Local Labour Systems of the Tuscany region. This model allows us to take into account the measurement error in the auxiliary variables.

E676: **Small area estimation of poverty proportions under Poisson mixed models**
Presenter: **Maria Jose Lombardia**, Universidade da Coruna, Spain
Co-authors: Miguel Boubeta, Domingo Morales

Poisson regression mixed model is used for studying poverty proportions in the Spanish provinces. Model fitting based on the method of moments is described and compared with the Laplace Maximum Likelihood method. The empirical best predictor (EBP) is derived under this model and its mean squared error (MSE). Two approximations to MSE of the EBP are proposed, based on the plug-in estimator and a parametric bootstrap. In addition, a complete simulation study shows the good behavior of our proposal. Finally, an application to Spanish data is carried out to obtain estimates of poverty indicators for Spanish provinces in 2008.

ES74 Room F1 STATISTICAL INFERENCE FOR LIFETIME DATA WITH APPLICATION TO RELIABILITY Chair: Jean-Yves Dauxois

E265: **A stochastic process for partial degradation data**
Presenter: **Emmanuel Remy**, EDF RD, France
Co-authors: Sophie Mercier, Laurent Bordes, Emilie Dautreme

EDF periodically performs in-service inspections of the components within its electric power plants in order to ensure that their degradation levels remain admissible. These examinations allow collecting successive flaw sizes from which one can estimate the residual operation time. Unfortunately the non-destructive testing processes do not give a perfect image of the degradation. Small flaws with sizes below a specific threshold may be detected without possible measurement. Moreover in case of several competing flaws on one component, even if the processes are able to count the number of existing flaws, they cannot measure all flaws sizes but only the largest one. In order to take into account this partial information, the University of Pau and EDF R&D developed a specific stochastic model. In this model, flaws initiate following a Poisson process and propagate according to gamma processes. Parametric statistical inference procedures are proposed, tested on simulated data and then applied to a real industrial case study carried out on a passive component from EDF power plants. The fitted model is next used to make some prediction. Finally a comparison is performed between the developed model and a simplified one which does not take into account the number of existing flaws.

E316: **Some probabilistic and statistical aspects of time-homogeneous load-sharing models**
Presenter: **Fabio Spizzichino**, University La Sapienza, Rome, Italy

The term "Load-sharing" designates a special class of absolutely continuous, multivariate probability distributions for non-negative random variables. Stochastic dependence among variables is expressed in terms of the so-called multivariate hazard rate functions. For this reason such models had emerged in a natural way within the field of Reliability. More recently, and under a different terminology, related interest had arisen in the field of financial risk and default-correlation. We concentrate attention on the time-homogeneous case. We first describe remarkable probabilistic features of models admitting the special property of time-homogeneity. In particular, from the specific point of view of reliability problems, we point out properties related to no-ageing. Then we consider some related aspects of statistical analysis of observed data. In this frame we compare the different forms of stochastic dependence generated by the two different conditions of conditional independence and time-homogeneous load-sharing. Some related implications in the analysis of financial risk will also be briefly considered.

E618: **Chi-square test based on random cells with recurrent events**
Presenter: **Akim Adekpedjou**, Missouri University of Science and Technology, United States
Co-authors: Withanage De Mel, Gideon Zamba

Recurrent event data is often observed in a wide variety of disciplines including the biomedical, public health, engineering, economic, actuarial science, and social science settings. Consider n independent units that are monitored for the occurrence of a recurrent event. Suppose the inter-failure times are iid with common distribution function F. The interest is the problem of testing the null hypothesis that F belongs to some parametric family of distributions where the parameter θ is unknown. We present results pertaining to a general chi-square goodness of fit test where cell boundaries are data dependent, that is are cut free without being predetermined in advance. The test is based on an estimator of θ known as the minimum chi-square estimator and a nonparametric maximum likelihood estimator of F. Large sample properties of the proposed test statistic under the null hypothesis and a Pitman-like alternative will be presented. A simulation study is conducted to assess the performance of the test under the null and parameter misspecification. Finally, the procedures are applied to a fleet of Boeing 720 jet planes' air conditioning system.

E667: **A multistate semi-Markov model with panel data for degradation analysis and water loss prediction of a water supply network**
Presenter: **Vincent Couallier**, University of Bordeaux, France
Co-authors: Karim Claudio, Yves Le Gat, Jerome Saracco

Multistate semi-Markov continuous-time models are well suited for modeling degradation process of water supply networks when the monitoring of such networks consists of "Active Leakage Control Campains". In that case, random samples of pipes in the network are observed at discrete fixed times with a classification of the degradation states as "no failure observed" or "invisible leak detected through technological devices". Additionally to this data, critical failures of pipes leading to apparent leakages emerging from the roadway are available without delay. Hence a three-state continous-time unidirectional graph for a Semi-Markov model has to be adjusted to such data, called panel data in biostatistics. The challenge is to fit the model without observing transition times but only current states at discrete observation times. The assumption of unidirectional graph with a parameterization of the transition intensities provides an easy solution to the parametric maximum likelihood estimation problem using interval censored sojourn times. The model is applied on a real data set in order to compare the results with an already implemented tool that uses the Gompitz Model, a GLM model for categorical ordinal responses. Covariates are taken into account using a proportional sub-hazard regression model.

E1263: **A test statistic for assessing a diagnostic marker**
Presenter: **Alba Franco-Pereira**, Universidad Complutense, Spain
Co-authors: M. Carmen Pardo

The area under the receiver operating characteristic (ROC) curve (AUC) is a widely accepted summary index of the overall performance of diagnostic procedures. We propose a nonparametric test based on spacings as an indicator of how far the curve is from that of the uninformative test as alternative to AUC. This approach is compared with AUC and other test statistics through a simulation study. Furthermore, we apply the proposed test statistic to assess a diagnostic marker in a real data set.

ES80 Room M1 ADVANCES IN SPATIAL FUNCTIONAL DATA ANALYSIS Chair: Elvira Romano

E637: **Simultaneous modeling of phase and amplitude variation in functional data on high-dimensional domains**
Presenter: **Lars Lau Raket**, University of Copenhagen, Denmark
Co-authors: Bo Markussen

Two types of random effects seem to be ever-present in functional data: individual functional samples contain serial or spatially correlated noise and samples generally suffer from some degree of misalignment. We consider data on high-dimensional Cartesian domains and propose a simultaneous model for these types of variation. We model spatially correlated noise as an additive random effect, and misalignment as a nonlinear random effect, and propose an algorithmic setup for doing inference in this model. We show how so-called operator approximations can be used to approximate the likelihood function, and demonstrate how this allows for extremely efficient inference on massively parallel hardware.

E654: **Smoothing based clustering for functional data**
Presenter: **Stefano Antonio Gattone**, University of Rome Tor Vergata, Italy

A procedure to cluster functional observations is proposed. The model is estimated by using penalized least squares to take into account the functional nature of the data. The standard "first smooth then cluster" approach performs smoothing and clustering separately without regard to each other. In this context, the level of smoothing is a parameter not related to the prediction of the cluster membership. In our proposal, the smoothing is carried out within the clustering and its amount is adaptively calibrated. In particular, we allow for different degrees of smoothing in each cluster. The intuition behind this approach is based on the hypothesis that curves in the same cluster tend to share more similar variability than curves in different clusters. A simulation study shows how the automatic data-driven smoothing improves the performance of clustering by reducing the irrelevant information present in the data.

E529: **Spatial functional data analysis for biodiversity**
Presenter: **Tonio Di Battista**, CH-PE, Italy
Co-authors: Fabrizio Maturo, Francesca Fortuna

Spatial interdependence of phenomena is a common feature of many environmental applications. At the same time, in an ecological context, the observations are often represented by functions rather than finite dimensional vectors. In particular, we deal with the problem of biodiversity measures focusing on the use of a parametric family of biodiversity indices called diversity profiles. The latter, in fact, can be analyzed in a functional framework because they represent biodiversity as a function of the relative abundance vector in a fixed domain. In this context, the aim is to model the variation in species composition among sites within a geographical area of interest allowing us to identify ecological spatial patterns. The framework has been applied to a real data set concerning lichen biodiversity in Liguria Region (NW Italy).

E955: **An approach for using monotonic regression in discrete spatial statistics**
Presenter: **Christian Rohrbeck**, Lancaster University, United Kingdom

Monotonic regression is considered in several research areas including functional data analysis and Bayesian inference. However, there is little research on its application to discrete spatial statistics. We introduce a new approach allowing the use of monotonic regression in discrete spatial statistical applications. The method defines a Bayesian framework which accounts for potential similarities in the monotonic functions of neighbouring regions. The functions can take any form and are estimated by a reversible jump MCMC algorithm. The approach is applied to simulated data.

E852: **Geographically weighted regression model for functional data spatially dependent**
Presenter: **Elvira Romano**, Second University of Napoles, Italy
Co-authors: Jorge Mateu, Carlos Diaz Avalos

A Geographically Weighted Regression Model (GWR) for functional data spatially dependent In this paper is proposed. The basic idea behind the classic GWR is to explore how the relationship between a dependent variable (Y) and one or more independent variables might vary geographically. We pursue a somewhat different emphasis to the classic model by considering functional variables, especially a new spatial-functional weighting scheme is defined. Several weighting functions can be defined in the classic spatial model. We extend the bi-square weighting functions in the spatial functional framework. Extensive simulation studies and an application on real data show the performances of the proposed method.

ES94 Room I1 STATISTICS WITH ACTUARIAL/ECONOMIC APPLICATIONS Chair: Tim Verdonck

E466: **Robust estimation of real exchange rate process half-life**
Presenter: **Michele Bergamelli**, City University London, United Kingdom

It is argued that the data generating process of the real exchange rate is likely to include outliers that, if not accounted for, produce unreliable half-lives estimates. In order to obtain robust estimates of the half-life, we propose to detect outlying observations by means of an extension of the Dummy Saturation approach to ARMA processes, considering additional and innovative outliers as well as level shifts in the real exchange rate process. An empirical application involving US dollar real exchange rates allows us to conclude that the estimated half-lives are consistently shorter after outlying observations are correctly modelled, thus shedding some light on the PPP puzzle.

E485: **Sparse PCA for high-dimensional data with outliers**
Presenter: **Tom Reynkens**, KU Leuven, Belgium
Co-authors: Mia Hubert, Tim Verdonck, Eric Schmitt

A new sparse PCA algorithm is presented which is robust against outliers. The approach is based on the ROBPCA algorithm which generates robust but non-sparse loadings. The construction of the new ROSPCA method is detailed, as well as a selection criterion for the sparsity parameter. An extensive simulation study and a real data example are performed, showing that it is capable of accurately finding the sparse structure of datasets, even when challenging outliers are present. In comparison with a projection pursuit based algorithm, ROSPCA demonstrates superior robustness properties and comparable sparsity estimation capability, as well as significantly faster computation time. We also show that ROSPCA can be extended to be used on data sets with skewed variables.

E692: **A new multivariate approach to Value-at-Risk measures**
Presenter: **Fatima Palacios Rodriguez**, University of Seville, Spain
Co-authors: Elena Di Bernardino, Jose Maria Fernandez-Ponce, Maria del Rosario Rodriguez-Grinolo

A risk-based approach for supervision and regulation of the financial sector is gaining ground in both emerging and industrialized countries. Value-at-Risk (VaR) is one measure being explored for this purpose. Risks may also be strongly heterogeneous in nature and difficult to diversify away. In the last decade, much research has been devoted to risk measures and many multidimensional extensions have been investigated. Particularly, two alternative extensions of the classic univariate Conditional Value-at-Risk are introduced in a multivariate setting. The two proposed multivariate CoVaRs are constructed from level sets of multivariate distribution functions (*resp.* of multivariate survival distribution functions). These vector-valued measures have the same dimension as the underlying risk portfolio. Several characterizations of these new risk measures are provided in terms of the copula structure and stochastic orderings of the marginal distributions. Interestingly, these results are consistent with existing properties on univariate risk measures. Furthermore, comparisons between existent risk measures and the proposed multivariate CoVaR are developed. Illustrations are given in the class of Archimedean copulas.

E619: **Basket option pricing and implied correlation in a Levy copula model**
Presenter: **Daniel Linders**, KU Leuven, Belgium
Co-authors: Wim Schoutens

The Levy copula model is introduced and the problem of finding accurate approximations for the price of a basket option is considered. The basket is a weighted sum of dependent stock prices and its distribution function is unknown or too complex to work with. Therefore, we replace the random variable describing the basket price at maturity by a random variable with a more simple structure. Moreover, the Carr-Madan formula can be used to determine approximate basket option prices. In a second part we show how implied volatility and implied correlation can be defined in our Levy copula model. In our Levy copula model, each stock is described by a volatility parameter and these marginal parameters can be calibrated separately from the correlation parameters. However, the available market prices for basket options together with our newly designed basket option pricing formula enables us to determine implied Levy correlation estimates. We observe that implied correlation depends on the

strike and the so-called implied Levy correlation smile is flatter than its Gaussian counterpart. The standard technique to price non-traded basket options (or other multi-asset derivatives), is by interpolating on the implied correlation curve. In the Gaussian copula model, this can sometimes lead to non-meaningful correlation values. We show that the Levy version of the implied correlation solves (at least to some extent) this problem.

E824: **Detecting the asymptotic dependence and independence via measures of association**
Presenter: **Vali Asimit**, City University London, United Kingdom

A crucial piece of information needed to estimate joint extreme events is to find whether or not asymptotic dependence or independence is present in multivariate data. Since this property represents the starting point of any statistical inference on multivariate extreme, several approaches have been explored in the literature. Our contribution is the use of well-known measures of association, such as Kendall's tau and Spearman's rho, in order to detect the asymptotic dependence/independence. Many characterisations are made in order to illustrate our proposed method, but its simplicity makes the proposed estimators even more attractive. Three data sets are chosen from environmental statistics, insurance and finance fields, which illustrate the power of our method.

ES99 Room E1 METHODS FOR SEMIPARAMETRIC ANALYSIS OF COMPLEX SURVIVAL DATA Chair: Liming Xiang

E761: **Rank-based inference for semiparametric accelerated failure time model under length-biased sampling**
Presenter: **Sy Han Chiou**, University of Minnesota Duluth, United States
Co-authors: Gongjun Xu

A class of rank-based estimation methods is proposed for semiparametric accelerated failure time model under informative right censoring as arise under length-biased sampling scheme. The new estimators are obtained from a generalized monotone estimating equation via an iterative induced smoothing procedure, and are shown to be consistent and asymptotically normal. Their variance matrices are estimated with an efficient resampling approach that avoids solving estimating equations repeatedly. In numerical studies, the proposed estimators were found to be more efficient in comparison to existing methods. The methodology was applied to a Spanish unemployment duration study.

E768: **Marginal inference of recurrent events under competing risks**
Presenter: **Bowen Li**, Academia Sinica, Taiwan
Co-authors: Yi-Hau Chen, Weijng Wang

The motivation is a dialysis study in northern Taiwan. Dialysis patients, after shunt implantation, may experience two types ("acute" or "non-acute") of shunt thrombosis, both of which may recur. We formulate the problem under the framework of recurrent events data in the presence of competing risks. In particular, we focus on marginal inference for the gap time of each specific type. Of interest are the cumulative incidence function (CIF) and cause-specific hazard function (CHF). Here the major challenge for statistical inference is the problem of induced dependent censoring. We apply the technique of inverse probability of censoring weighting to adjust for the selection bias. Point estimation as well as standard errors given by a bootstrap re-sampling approach are provided for both CIF and CHF. Large sample properties and simulation results are examined. We apply the procedure proposed to analyze the dialysis data.

E1084: **Frailty modelling approaches for semi-competing risks data via H-likelihood**
Presenter: **Il Do Ha**, Pukyong National University, Korea, South
Co-authors: Jong-Hyeon Jeong , Youngjo Lee

The semi-competing risks situation is considered where a terminal event (e.g. death) censors a non-terminal event (e.g. disease recurrence), but not vice versa. A subject may experience both events which can lead to correlation. Such correlation can be usually modelled via frailty (i.e. unobserved random effect). We show that semi-parametric frailty models are useful for modelling the semi-competing risks data. We allow gamma and lognormal distributions for the frailty. For inference we develop a hierarchical likelihood (or h-likelihood) procedure, which obviates the need for an intractable integration over the frailty terms. For comparison we include the marginal likelihood approach. The proposed method is illustrated using a time-to-event data set from the B14 randomized multicenter breast cancer trial conducted by the NSABP of USA and a simulation study.

E1109: **A dimension reduction approach for conditional Kaplan-Meier estimators**
Presenter: **Valentin Patilea**, CREST-Ensai, France
Co-authors: Weiyu Li

In many situations, the quantities of interest in survival analysis are smooth, closed-form expression functionals of the law of the observations. This is for instance the case for the law of a lifetime of interest under random right censoring. In the presence of multivariate covariates, the nonparametric estimation of the quantities of interest suffers from the curse of dimensionality. A new dimension reduction approach for conditional model in survival analysis is proposed. First, we consider a single-index hypothesis on the conditional law of the observations and we propose a suitable semiparametric estimator. Next, we apply the smooth functionals to the estimator of the conditional law of the observations. This results in semiparametric estimators of the quantites of interest that avoid the curse of dimensionality. Moreover, bootstrapping from the estimated conditional law of the observations we easily build confidence regions. The vector of covariates need not have a density and the observed lifetimes need not be continuous random variables. The new methodology allows us to test the dimension reduction constraint and extends to other dimension reduction methods. Moreover, it can be applied to more general censoring mechanisms that allow for closed-form expression of the functionals of interest. An empirical study using simulated and real data illustrates the effectiveness of the proposed approach.

E664: **Penalized empirical likelihood for sparse additive hazards regression with diverging number of covariates**
Presenter: **Liming Xiang**, Nanyang Technological University, Singapore
Co-authors: Shanshan Wang

High-dimensional sparse modeling with censored survival data is of great practical importance, as exemplified by applications in high-throughput genomic data analysis. We propose a class of regularization methods, integrating both the penalized empirical likelihood and pseudoscore approaches, for simultaneous variable selection and inference in sparse and high-dimensional additive hazards regression models. When the dimensionality grows with the sample size, we establish the oracle property for the proposed penalized empirical likelihood method, i.e., with probability tending to 1: it identifies the true additive hazards regression and estimates the nonzero coefficients efficiently as if the sparsity of the true model was known in advance. We also develop a nonparametric likelihood approach based on a penalized empirical likelihood ratio statistic for the linear hypothesis of regression coefficients. Furthermore, simulation studies are carried out to evaluate the performance of the proposed method and a real data set is analysed.

ES122 Room G1 ADVANCES IN LATENT VARIABLES Chair: Paola Zuccolotto

E323: **Scaling for structure with STOPS**
Presenter: **Thomas Rusch**, WU - Vienna University of Economics and Business, Austria
Co-authors: Patrick Mair, Kurt Hornik

Multidimensional scaling (MDS) is a versatile statistical method whose general idea is to reduce the multivariate complexity in a data set by employing suitable proximities between the data points and finding low-dimensional configurations where the distances optimally approximate the proximities. The ultimate goal, however, is often not only to find the optimal configuration but to infer statements about the similarity of objects in the high-dimensional space. Since these two goals are somewhat different, it can happen that the resulting optimal configuration makes

inferring similarities rather difficult. In that case the MDS solution lacks "structuredness". The authors present a modified version of MDS, called structure-optimized proximity scaling (STOPS), that solves this conundrum by introducing more structure into the configuration while adhering to the general idea of MDS. In STOPS, an arbitrary stress function is parametrized by nonlinear power transformations and combined with an index that captures the "structuredness" of a configuration. The combination is called "structure-penalized stress" and is minimized to push any initial MDS configuration towards a more structured appearance. The method will be illustrated with an example for assessing similarities of countries based on the history of banking crises in the last 200 years.

E605: **Measuring change on latent variables using linear logistic models**
Presenter: **Marco Johannes Maier**, Vienna University of Economics and Business, Austria
Co-authors: Thomas Rusch, Patrick Mair

Assessing change on latent variables over time, based on categorical manifest variables (indicators), plays a pivotal role in longitudinal designs in the social sciences. Usually, the indicators are items from a psychological or educational test, questionnaire, or survey. If certain assumptions hold, the family of linear logistic models (LLM) provides a flexible framework to assess change on the latent variable over time. The LLM family has well-known and popular, but also hitherto unexplored models. Focusing on the latter, conceptual gaps can be filled by extending models to allow for continuous covariates and treating certain items in a set interchangeably. With these additions, various specifications of LLMs can be used to measure change, for example, uni-and multidimensional conceptualization of each item, item set, and/or latent variable, measurement models that a set of items adheres to, inclusion of categorical or continuous covariates, as well as varying degrees of assumptions on the mutual dependencies of items. A conditional maximum likelihood (CML) approach, which does not have any distributional assumptions about latent traits, is employed for estimation and inference using the implementation in the R package eRm. The usage of new LLM types is demonstrated with data from an introductory finance course at WU.

E462: **A unified proposal for modelling ordinal data**
Presenter: **Maria Iannario**, University of Naples Federico II, Italy

A generalized mixture model is introduced for the analysis of ordinal data in evaluation contexts. Mixtures with feeling and uncertainty latent components are statistical approaches used for interpreting decisional processes also when shelter effect or overdispersion are present. In this context we extend this proposal by considering a family which also includes standard models for the analysis of ordinal data (cumulative logit and probit). A careful analysis is devoted to the ability to capture the indecision or the guessing process. The EM algorithm is described to find maximum likelihood estimates of the parameters for these models. Some aspects concerning the inference issues are also analysed. This novel family of models is applied to real data sets and a comparison of performance with other well-established mixture-based approaches is presented.

E679: **A multivariate CUB model**
Presenter: **Sabrina Giordano**, University of Calabria, Italy
Co-authors: Roberto Colombi

CUB (Combination of Shifted Binomial and discrete Uniform variables) models are mixture models for ordinal data. The founding idea supposes that when respondents are asked to express their preference towards items (attitudes, services, objects) by selecting a category among ordered alternatives, the final choice depends on two latent components, i.e. feeling and uncertainty, so that the respondent's choice is modelled as a combination of a shifted binomial variable (feeling) and a discrete uniform variable (uncertainty). Quite often, people are requested to express preferences on several aspects related to an item or a service, it is therefore natural to expect the answers from the same individual to be related. Although CUB models have been widely investigated in the univariate framework, modeling the joint distribution of CUB responses is still a challenging problem. We propose the use of the Sarmanov multivariate distribution with CUB marginals, taking into account also the role of subject's and object's covariates. This distribution accommodates both positively or negatively associated variables, is mathematically tractable, allows the association parameters to be readily interpretable and its inferential issues are easily dealt with. The practical usefulness of the methodological results is illustrated on real data.

E1010: **Accounting for attribute non-attendance in the analysis of discrete choice experiments**
Presenter: **Michel Meulders**, KU Leuven, Belgium
Co-authors: Jeroen K. Vermunt

Nowadays stated preference techniques such as discrete choice experiments have become a popular tool to model the product preference of consumers as a function of product characteristics. Data from discrete choice experiments are often modeled with latent class conditional logit models to account for the fact that consumers may weigh product characteristics differently when choosing among products. Furthermore, a standard assumption when analyzing data from choice experiments is that subjects consider all attributes as relevant when choosing the most preferred alternative from a choice set. However, research has indicated that consumers may only attend to specific subsets of attributes when choosing between alternatives. This so-called attribute nonattendance can emerge for several reasons and failure to take it into account may lead to biased marginal willingness-to-pay estimates for product characteristics. We discuss how standard latent class conditional logit models (i.e. with class-specific regression parameters) can be extended to account for attribute non-attendance. As an illustration, the models are applied to analyze student housing data.

ES39 Room Q1 BAYESIAN SEMI- AND NONPARAMETRIC MODELLING I Chair: Matteo Ruggiero

E636: **Bayesian principal curve clustering by NGG-mixture models**
Presenter: **Alessandra Guglielmi**, Politecnico di Milano, Italy
Co-authors: Raffaele Argiento

The interest is in clustering data whose support is "curved". We follow a Bayesian nonparametric approach by considering a mixture model where the mixing distribution is a random probability measure, namely the normalized generalized gamma (NGG) process. The goal is to define a general but flexible class of distributions to model data from clusters with non-standard shape. To this end, we extend the definition of principal curve given by Tibshirani into a Bayesian framework. We propose a new hierarchical model, where data in each cluster are parametrically distributed around the Bayesian principal curve, and the prior cluster assignment is given on the latent variables at the second level of hierarchy according to the NGG process. We also build an efficient Gibbs sampler scheme to simulate from the posterior, based on the truncation of the infinite sum representing the NGG process. The parameter epsilon controlling the degree of approximation of the truncation method can be given a prior; this is particularly convenient from a computational point of view, since "large" values of epsilon give shorter run-times of the algorithm. We show two applications: a simulated dataset, and the detection of seismic faults using data from Italian earthquake catalogues.

E311: **Bayesian nonparametric modeling of clustered survival data**
Presenter: **Michele Guindani**, MD Anderson Cancer Center, United States
Co-authors: Bernardo Nipoti, Alejandro Jara

Clustered failure time data are frequently encountered in biomedical research. They typically arise when the outcome of interest is observed on a collection of subjects that may be grouped into a common structure, or cluster. For example, observations may be collected on members of the same family, patients in the same hospital, animals in litters, or multiple organs in a single patient. In all such cases, the affiliiation of subjects to a common cluster is likely to induce correlation among the individual outcomes, which needs to be appropriately modeled. We discuss proportional hazard models where the baseline hazard is modeled non-parametrically as a random hazard mixture. In particular, we focus on the dependence

structure that arises, in the multivariate setting, from different choices of mixing measures. We show 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. We illustrate our methodological framework by means of a simulation study as well as an application to a real dataset

E425: **Moment-constrained Ferguson-Klass algorithm**

Presenter: **Julyan Arbel**, CREST - Universite Paris Dauphine, France

Co-authors: Igor Prunster

Completely random measures have been used as mixing distributions for building a priori distributions in Bayesian nonparametric settings, and proved to result in flexible and accurate models. A commonly adopted strategy for posterior sampling is the Ferguson-Klass algorithm which consists in simulating trajectories of the completely random measures. Due to the discrete form as infinite sum of jumps of a completely random measure, one resorts to an approximate posterior scheme by sampling only a finite number of jumps. The goal is to quantify the approximation in the Ferguson-Klass algorithm by relying on a thorough moment-matching technique. We show that for large classes of completely random measures (Generalized Gamma Process, Beta Process), a limited number of jumps is sufficient to achieve a very good precision for the (a priori) sampling. Then we show that in the Normalized random measures with independent increments model (NRMI) and the Indian Buffet Process model, the precision of the posterior sampling is improved compared to the prior sampling precision. Simulation results indicate that the Ferguson-Klass algorithm represents a valid default choice for posterior sampling in Bayesian nonparametric problems based on completely random measures.

E434: **Comparison of mixture models through locally tied stick-breaking processes**

Presenter: **Jacopo Soriano**, Duke University, United States

Co-authors: Li Ma

Two- or multi-sample comparison of mixture distributions involving local differences in some mixture components is considered. We propose a nonparametric Bayesian method that achieves high statistical power through effectively combining information across the mixture distributions. Our method incorporates local dependencies between the models through spiked priors on both the weights and the locations in a stick-breaking representation. These priors allow one to test for variations in the mixture proportions and shifts in the mixture locations respectively. We focus in particular on a class of nonparametric distributions that induce a Dirichlet Process mixture marginally for each model, with posterior inference carried out efficiently using a blocked Gibbs sampler. Simulation studies show that our method compares favorably in detecting local cross-sample differences to other state-of-the-art methods. We apply the method to a flow cytometry study, and it successfully identifies rare cell subpopulations that differentiate multiple cell samples.

E711: **Functional Bayesian point process model for neuroimaging meta-analysis data**

Presenter: **Silvia Montagna**, University of Warwick, United Kingdom

Co-authors: Timothy D. Johnson, Thomas E. Nichols

Functional MRI (fMRI) is now 20 years old, and there is a massive and growing literature that requires meta-analytic tools to synthesise. While fMRI data comprise 3-dimensional images, a typical fMRI publication only reports the peak activation coordinates (foci), the location of local maxima in statistic images. Neuroimaging meta-analysis is used to 1) identify areas of consistent activation and 2) build a predictive model of task type or cognitive process for new studies. To simultaneously address both these goals and fill a current gap in the literature, we propose a Bayesian hierarchical model for neuroimaging point pattern data from multiple types of studies. We model the foci from each study as a spatial Cox process, and the study-specific log intensity function is characterised as a linear combination of a high-dimensional basis set. The functional representation leads to computational speed-ups over traditional log-Gaussian Cox processes, and sparse representation of the intensities is guaranteed through sparse latent factor modelling of the basis coefficients. Further, the latent factors are used to link the random intensities to a study-type, as part of a scalar (or vector)-on-image regression. We illustrate our model with a meta-analysis of nine emotions from 510 studies.

Saturday 6.12.2014 14:35 - 16:15 Parallel Session D – CFE

CSI02 Room Sala Convegni MODELING AND FORECASTING HIGH DIMENSIONAL TIME SERIES Chair: Manfred Deistler

C286: **Dynamic factor models: I(1) variables and cointegration**
Presenter: **Marco Lippi**, Universita di Roma La Sapienza, Italy

The case of Dynamic Factor Models with I(1) variables is systematically explored. By combining the classic Granger Representation Theorem with recent results on singular stochastic vectors, it is shown that for generic values of the parameters the Factors have an Error Correction representation with two unusual features: (i) the autoregressive matrix polynomial is of finite length, (ii) the number of error-terms is equal to the number of transitory shocks plus the difference between the dimension and the rank of the Factors. This result is the basis for the correct specification of an autoregressive model for the Factors. Estimation of impulse-response functions is also discussed. Results of an empirical analysis on a US quarterly database support the use of the model proposed.

C1246: **Short and long run second-order causality: theory, measures and inference**
Presenter: **Jean-Marie Dufour**, McGill University, Canada
Co-authors: Hui Jun Zhang

As financial markets being more integrated, financial volatilities move together more or less closely over time across assets (markets). It suggests that causal relationships may exist in conditional variances. Understanding such causal relationships at different horizons has important implications for financial crises and market integration, as well as can provide useful guides for investments, risk management and policy making. However, most of the literature has focused on mean causality, only limited attempts have been made to characterize causality for conditional variances. We extend the concept of second-order and variance causality proposed for the one period horizon to multiple horizons (horizons can be infinite), with considering the indirect causality transmitted by auxiliary variables. We also provide two new definitions of casuality for conditional covariances: cross-second-order and covariance causality. Properties of (cross-)second-order and (co-)variance noncausality at different horizons are presented. In addition, we propose the measures of (cross-)second-order causality at any given horizon. To evaluate the measures, we suggest a simple estimation method and a simulation-based inference procedure 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 between stock indexes in U.S. market and across U.S. and U.K. markets.

C1256: **Multivariate AR systems and mixed frequency data: identifiability and estimation**
Presenter: **Manfred Deistler**, Vienna University of Technology, Austria
Co-authors: Brian D.O. Anderson, Elisabeth Felsenstein, Bernd Funovits, Lukas Koelbl, Mohsen Zamani

The research is concerned with the problem of identifiability of the parameters of a high frequency multivariate autoregressive model from mixed frequency time series data. We demonstrate identifiability for generic parameter values using the population second moments corresponding to the observations. In addition we display a constructive algorithm for the parameter values and establish the continuity of the mapping attaching the high frequency parameters to these population second moments. These structural results are obtained using two alternative tools viz. extended Yule Walker equations and blocking of the the output process. The cases of stock and how variables as well as of general linear transformations of high frequency data are treated. Finally, we shortly discuss how our constructive identifiability results can be used for parameter estimation based on the sample second moments.

CS01 Room M2 FINANCIAL ECONOMETRICS Chair: Niklas Ahlgren

C113: **Momentum, uncertainty, and exchange rate predictability**
Presenter: **Matthijs Lof**, Aalto University, Finland

Empirical estimates and out-of-sample forecasting results of boundedly-rational heterogeneous agent models are presented using monthly data on nine major currencies. The models feature Fundamentalists, who expect exchange rates to adjust towards the value implied by Purchasing Power Parity, and Trend-followers (momentum traders), who expect recent trends in exchange rates to persist for at least into the near future. In terms of forecasting performance, the heterogeneous-agent models outperform a simple random walk, as well as linear alternatives. In addition, it is found that trend-following behavior is less prevalent in periods of high market volatility, which is consistent with recent findings in the literature on momentum in stock returns.

C132: **Generalized forecast error variance decomposition for linear and nonlinear multivariate models**
Presenter: **Henri Nyberg**, University of Helsinki, Finland
Co-authors: Markku Lanne

A new generalized forecast error variance decomposition is proposed with the property that the proportions of the impact accounted for by innovations in each variable sum to unity. Our decomposition is based on the well-established concept of the generalized impulse response function. The use of the new decomposition is illustrated with an empirical application to U.S. output growth and interest rate spread data.

C176: **The power of wild bootstrap tests of cointegration rank with unconditional and conditional heteroskedasticity**
Presenter: **Paul Catani**, Hanken School of Economics, Finland
Co-authors: Niklas Ahlgren

Heteroskedasticity in its two forms, unconditional and conditional, is pervasive in financial time series. When heteroskedasticity is present, asymptotic and IID bootstrap tests of cointegration rank become unreliable. Wild bootstrap tests are valid under both forms of heteroskedasticity. We investigate the impact of unconditional heteroskedasticity and conditional heteroskedasticity in the form of strong persistence in volatility on the power of wild bootstrap tests of cointegration rank. Monte Carlo simulations show that wild bootstrap tests have low power in such circumstances. In an application to credit default swap prices and credit spreads, we show that more than 1000 observations are required to obtain high power of tests for cointegration.

C901: **Further evidence on long-run stock returns after corporate events**
Presenter: **Seppo Pynnonen**, University of Vaasa, Finland
Co-authors: James Kolari, Ahmet Tunez

Recent work is revisited on long-run abnormal stock returns after merger and acquisitions, IPOs, SEOs, and dividend initiations. We repeat the regression analyses and confirm the findings for the most part using the normalized factor procedure but not non-normalized (original) factor values. Extending the analyses, we apply Sharpe ratios to the computation of abnormal returns. Sharpe abnormal returns are risk-adjusted, weight returns by the statistical precision, and allow observation of intertemporal patterns after corporate events. Significant long-run abnormal returns are detected for all four corporate events. Contrary to the previous work, we conclude that risk adjustment alone cannot explain these significant abnormal returns.

CS03 Room C2 ADVANCES IN IDENTIFICATION OF STRUCTURAL VECTOR AUTOREGRESSIVE MODELS Chair: Christiane Baumeister

C168: **The analytics of SVARs: A unified framework to measure fiscal multipliers**
Presenter: **Dario Caldara**, Board of Governors, United States
Co-authors: Christophe Kamps

Whether fiscal policy stimulates output is analyzed. SVARs have been used to address this question but no stylized facts have emerged. We derive analytical relationships between the output elasticities of fiscal variables and fiscal multipliers. We show that standard identification schemes imply different priors on elasticities, generating a large dispersion in multiplier estimates. We then use extra-model information to narrow the set of empirically plausible elasticities, allowing for sharper inference on multipliers. Our results for the U.S. for the period 1947-2006 suggest that the probability of the tax multiplier being larger than the spending multiplier is below 0.5 at all horizons.

C021: **Sign restrictions, structural vector autoregressions, and useful prior information**
Presenter: **Christiane Baumeister**, Bank of Canada, Canada
Co-authors: James D. Hamilton

Many empirical studies have used numerical Bayesian methods for structural inference in vector autoregressions that are identified solely on the basis of sign restrictions. Because sign restrictions only provide set-identification of structural parameters, over certain regions of the parameter space the posterior inference could only be a restatement of prior beliefs. We characterize these regions, explicate the beliefs about parameters that are implicit in conventional priors, provide an analytical characterization of the full posterior distribution for arbitrary priors, and analyze the asymptotic properties of this posterior distribution. We show that in a bivariate supply and demand example, if the population correlation between the VAR residuals is negative, then even if one has available an infinite sample of data, any inference about the supply elasticity is coming solely from the prior distribution. More generally, the asymptotic posterior distribution of contemporaneous coefficients in an n-variable VAR is confined to the set of values that orthogonalize the population variance-covariance matrix, with the height of the posterior proportional to the prior at any point within that set. We suggest that researchers should defend their prior beliefs explicitly and report the difference between prior and posterior distributions for key magnitudes of interest.

C050: **Identification and estimation of the macroeconomic effects of monetary policy shocks in the U.S.**
Presenter: **Luca Fanelli**, University of Bologna, Italy
Co-authors: Emanuele Bacchiocchi, Efrem Castelnuovo

A novel identification scheme is employed to quantify the macroeconomic effects of monetary policy shocks in the United States. The identification of the shocks is achieved by exploiting the instabilities in the contemporaneous coefficients of the structural VAR (SVAR) and in the covariance matrix of the reduced-form residuals. Different volatility regimes can be associated with different transmission mechanisms of the identified structural shocks. We formally test and reject the stability of our impulse responses estimated with post-WWII U.S. data by working with a break in macroeconomic volatilities occurred in the mid-1980s. We show that the impulse responses obtained with our non-recursive identification scheme are quite similar to those conditional on a standard Cholesky-SVARs estimated with pre-1984 data. In contrast, recursive vs. non-recursive identification schemes return substantially different macroeconomic reactions conditional on Great Moderation data, in particular as for inflation and a long-term interest rate. Using our non-recursive SVARs as auxiliary models to estimate a small-scale new-Keynesian model of the business cycle with an impulse response function matching approach, we show that the instabilities in the estimated VAR impulse responses are informative as for the estimation of some key-structural parameters.

C030: **Mixed frequency structural VARs**
Presenter: **Claudia Foroni**, Norges Bank, Norway
Co-authors: Massimiliano Marcellino

A mismatch between the time scale of a structural VAR (SVAR) model and that of the time series data used for its estimation can have serious consequences for identification, estimation and interpretation of the impulse response functions. However, the use of mixed frequency data, combined with a proper estimation approach, can alleviate the temporal aggregation bias, mitigate the identification issues, and yield more reliable responses to shocks. The problems and possible remedy are illustrated analytically and with both simulated and actual data.

CS17 Room N2 LIQUIDITY AND CONTAGION Chair: Serge Darolles

C411: **Illiquidity transmission from spot to futures markets**
Presenter: **Erik Theissen**, University of Mannheim, Germany
Co-authors: Olaf Korn, Paolo Krischak

The aim is to develop a model of the illiquidity transmission from spot to futures markets that formalizes the "derivative hedge theory". The model shows that spot market illiquidity does not translate one-to-one to the futures market, but rather interacts with price risk, liquidity risk, and the risk aversion of the market maker. The predictions of the model are tested empirically with data from the stock market and the market for single-stock futures. The results support our model. In particular, they show that the derivative hedge theory is important for the explanation of the liquidity link between spot and futures markets. Our results provide no evidence in favor of the substitution hypothesis.

C424: **The determinants of ETF liquidity: theory and evidence from European markets**
Presenter: **Laurent Deville**, Universite de Nice Sophia Antipolis - CNRS, France
Co-authors: Anna Calamia, Fabrice Riva

Professionals conventionally argue that, as open-ended funds, ETFs should exhibit a degree of liquidity that is proportional to that of their underlying basket. We develop a model where a risk averse ETF market maker liquidates her ETF inventory against the underlying stock basket and bears illiquidity costs when closing out her position. Though our model confirms that ETFs replicating illiquid indices should trade with higher spreads, it also predicts that ETF spreads should be positively correlated with the underlying index volatility and that higher ETF trading volumes will result in lower ETF spreads since market makers will face lower expected inventory at the end of the day. We test these predictions on a large sample of European equity ETFs. Spreads appear to positively depend on variables related to inventory risk. The stock basket spread does not affect ETF spread in the whole sample, but still does so for less traded ETFs. These findings suggest that ETFs are dealt like stocks by market makers when the trading volume is sufficiently high to manage their inventory with low risk. Still, market makers have to pass on the underlying stock basket illiquidity to their quoted spreads when trading volumes are low.

C469: **Tracking illiquidities in daily and intradaily characteristics**
Presenter: **Gulten Mero**, Universite de Cergy-Pontoise and THEMA, France
Co-authors: Serge Darolles, Gaelle Le Fol

Two types of liquidity problems are distinguished, called respectively liquidity frictions and illiquidity events. The first one is related to order imbalances that are resorbed within the trading day. It can be assimilated to "immediacy cost" and impacts the traded volume at the intraday and daily frequencies while affecting the price increments only at the intraday periodicity. The second one is inherent to the long lasting liquidity problems and is responsible for the time-dependence of the daily returns and volume. We extend the MDHL framework of Darolles to account for the presence of the illiquidity events. We then propose a two-step signal extraction formulation of the MDHL model in order to separate the

two liquidity problem impacts on the daily returns and volume. We also provide, for a set of FTSE100 individual stocks, long lasting illiquidity indicators.

C873: **Non-synchronous market impact and hedge fund portfolio construction**
Presenter: **Serge Darolles**, Paris Dauphine, France
Co-authors: Mathieu Vaissie

Technical change, regulatory change, and institutional change have profoundly changed the face of financial markets. A greater variety of intermediaries can now engage in a greater diversity of transactions, allowing for a better utilization of the risk-bearing capacity of the economy. The flipside is that the linkages between the myriad of nodes composing the system are tighter than they have ever been. A wealth of indirect effects, in addition to the initial shock, are now driving the dynamics of the market. Against this backdrop, the use of traditional factor models, where variables are assumed to be independent and identically distributed appears to be ill-suited. This is particularly true when the return series of the underlying assets have been shown to exhibit a significant level of autocorrelation. We propose to build on the literature on contagion to disentangle the direct and indirect impacts of an initial shock on the risk structure, and in turn, to choose the appropriate variance/covariance matrix to design a robust hedge fund portfolio.

CS32 Room P2 CO-MOVEMENTS IN MACROECONOMICS AND FINANCE Chair: Alain Hecq

C632: **Factor-augmented autoregressiove models: representation, estimation, and forecasting**
Presenter: **Gianluca Cubadda**, University of Rome Tor Vergata, Italy
Co-authors: Elisa Scambelloni

The purpose is to examine the condition under which each individual series that is generated by an n-dimensional Vector Autoregressive Model can be represented as an autoregressive model that is augmented with the lags of q linear combinations of all the variables in the system. We call this modelling Factor-Augmented Autoregressions (FAAR). We show that the parameters of the FAAR can be estimated by a switching algorithm that increases the Gaussian likelihood at each iteration. Provided that the number of factors q times the VAR order p is small compared to the sample size T, our approach can be applied even when the number of series n is large. We illustrate the usefulness of the FAAR modelling both by empirical applications and simulations.

C646: **Cross-border effects of coordinated fiscal policy in the Eurozone**
Presenter: **Lenard Lieb**, Maastricht University, Netherlands
Co-authors: Andreea Bicu

Spillovers from fiscal policy shocks in the euro area are empirically quantified. We use a multi-country factor-augmented Panel-VAR identified with sign restrictions and analyse the domestic and international effects of coordinated fiscal policy. By extracting the information from an extended set of country specific and cross-border variables, we look at the different channels through which either revenues or expenditure shocks are transmitted within as well as across borders. We find significant negative effects of fiscal consolidations on domestic output and consumption. More importantly, coordinated spending cuts in the periphery induce delayed but significant output spillovers in Germany and France.

C680: **A vector heterogeneous autoregressive index model for bi-power variation**
Presenter: **Barbara Guardabascio**, ISTAT - Italian Statistic Institute, Italy
Co-authors: Gianluca Cubadda, Alain Hecq

A new modelling is introduced for detecting the presence of commonalities in the volatility of a set of time series. Our Vector Heterogeneous Autoregressive Index model (or VHARI) is simple to estimate when applied to realized volatility or bi-power variation measures of the latent volatility feature. Moreover, the VHARI specification has nice properties in terms of parsimony as well as for the interpretation of the connections between univariate and multivariate representations. We illustrate its usefulness when forecasting the volatility of 20 equity indexes.

C823: **Exuberance: an empirical investigation of sentiment driven buoyancy**
Presenter: **Guillaume Chevillon**, ESSEC Business School, France

Sentiments, as measures of optimism or pessimism have recently been proposed as explanations for some of the observed large fluctuations in financial markets. We study their interactions and spillovers through a large class of asset and commodity price returns and markets. We exploit for this the recent proposal of Near Explosive Random Coefficient autoregressive models which we extend to allow for a multivariate framework where several sentiments drive the observed fluctuations, large deviations and bubbles.

CS42 Room B2 TIME-SERIES ECONOMETRICS Chair: Robert Kunst

C057: **Forecasting tourism demand with Google trends: The case of Vienna**
Presenter: **Ulrich Gunter**, MODUL University Vienna, Austria
Co-authors: Irem Onder

The purpose is to investigate whether Google Trends data have predictive power in terms of improving forecasting accuracy of tourism demand relative to a baseline model. Vienna is chosen as a case example and four simplified ADL models that allow for seasonality are derived to test this hypothesis: (1) baseline autoregressive model with own lags of tourism demand only, (2) web search index model, (3) image search index model, and (4) web and image search index model with Google Trends web and/or image search indices as explanatory variables. The ex-post forecasting accuracy is assessed in terms of MAPE and RMPSE for horizons 1, 2, 3, 6, and 12 months ahead. The accuracy is improved for horizons 1, 2, 3, and 6 when Google Trends data are included, thereby underlining the usefulness of Google Trends data as predictor for short-term forecasting. In addition, naive-1 and seasonal naive benchmarks are significantly outperformed across horizons according to the Hansen test. Current research is extending the topic by the following dimensions: (a) origin of Google queries and tourists to Vienna, (b) original language of Google queries per source market, and (c) impact of the use of deseasonalized data.

C361: **Estimation of generalized long-memory stochastic volatility: Whittle and wavelets**
Presenter: **Michael Hauser**, Vienna University of Economics and Business, Austria
Co-authors: Alex Gonzaga

The aim is to compare the Whittle and a wavelet based Whittle estimator, WWE, for k-GARMA and generalized stochastic long-memory volatility models, GLMSV. As the decorrelation properties of wavelets for FI are shown to hold also for k-GARMA and GLMSV the consistency of the WWE can be derived. The small sample properties of Whitcher's, WWE and Whittle's estimator are compared. WWE clearly dominates Whichter's, and is essentially indistinguishable to Whittle's. The estimators are illustrated by fitting a GLMSV to Microsoft realized volatilities.

C610: **Multiple breaks in long memory time series**
Presenter: **Heiko Rachinger**, University of Vienna, Austria

Least squares (LS) estimation of breaks in long memory time series is analyzed. We show that the estimator of the break fraction is consistent and converges at rate T when there is a break in the level, in the memory or in both parameters. We also derive the asymptotic distribution of the break fraction under shrinking break magnitudes. Further, we analyze tests for the number of breaks. When testing for breaks in the memory, the asymptotic results correspond to standard ones in the literature. When testing for breaks in the level and when testing for breaks in both parameters, the results differ in terms of the asymptotic distribution of the test statistic. In this case, the LS-procedure loses its asymptotic pivotality. We further

propose a method in order to distinguish between long memory, breaks in the memory and breaks in the level. Such a distinction is difficult but is important for reasons such as shock identification, forecasting and detection of spurious fractional cointegration. In a simulation exercise, we find that the tests based on asymptotic critical values are oversized in finite samples. Therefore, we suggest using the bootstrap, for which we derive validity and consistency, and we confirm its better size properties. Finally, we use the method to test for breaks in the U.S. inflation rate.

C271: **Forecasting seasonal data and nonparametric unit-root tests**
Presenter: **Robert Kunst**, Institute for Advanced Studies, Austria

Nonparametric unit-root tests tend to trade off power for enhanced robustness features. We consider a variant of the RURS (seasonal range unit roots) test statistic, a variant of the level-crossings count adapted to classes of seasonal patterns, and a new combined test. These tests exploit two main characteristics of seasonal unit-root models, the range expansion typical of integrated processes and the frequency of changes among seasonal shapes. For standard designs, parametric seasonal unit-root tests of the HEGY (Hylleberg Engle Granger Yoo) type dominate non-parametric rivals in large samples. In small samples, surprising local power gains by range tests have been reported. It is of interest whether such power advantages transfer into enhanced predictive accuracy out of sample. We explore the implications of test-based decisions on predictions of seasonal time series. Apart from generating processes with seasonal unit roots and with deterministic seasonality, we also study seasonal time deformation. For predictions from test-based selection, non-parametric tests are found to be competitive with traditional HEGY tests in standard data generating models. In some non-standard designs, however, the non-parametric tests can perform poorly. Some applications to real-world data confirm this impression.

CS53 Room O2 ADVANCES IN DSGE MODELLING Chair: Alexander Meyer-Gohde

C493: **Efficient simulation of DSGE models with occasionally binding constraints**
Presenter: **Tom Holden**, University of Surrey, United Kingdom
Co-authors: Michael Paetz

A computationally efficient algorithm is presented for the solution and simulation of dynamic stochastic general equilibrium models with occasionally binding constraints. The algorithm can deal with arbitrarily many bounded variables, and thanks to a hybrid local/global approximation, the algorithm is able to capture the precautionary effects associated with the risk of hitting a bound, even when the rest of the model is approximated to first order. Our procedure is much faster than comparable methods for a given level of accuracy, and can readily handle large models. To illustrate the usefulness and efficiency of this algorithm we provide a variety of applications to models incorporating a zero lower bound (ZLB) on nominal interest rates, and to multi-country irreversible-investment models.

C547: **Generalized exogenous processes in DSGE: a Bayesian approach**
Presenter: **Alexander Meyer-Gohde**, Humboldt University Berlin, Germany
Co-authors: Daniel Neuhoff

A Reversible Jump Markov Chain Monte Carlo (RJMCMC) method to Bayesian DSGE estimation is introduced. The method enables us to sample from a posterior distribution spanning nonnested models with parameter spaces of different dimensionality. We use the method to jointly sample from an ARMA process of unknown order along with the associated parameters. We apply the method to a canonical neoclassical growth model using post war US GDP data and find that the posterior decisively rejects the standard AR(1) assumption for the mean reverting technology process in favor of higher order processes. While the posterior contains significant uncertainty regarding the exact order, it provides tight posterior credible sets over the impulse responses. At the same time, the responses are hump-shaped. A negative response of hours to a positive technology shock is within the posterior credible set when noninvertible MA representations are considered.

C1163: **Estimation of the DSGE models with multivariate detrending**
Presenter: **Artem Duplinskiy**, Maastricht University, Netherlands
Co-authors: Franz Palm, Jean-Pierre Urbain

The focus is on the role of the detrending in the estimation of a simple Dynamic Stochastic General Equilibrium (DSGE) model and try to engage the information contained in the non-cyclical component in the estimation. We apply the same filter to the data and the model variables. The filtered variables are stationary, for both permanent or transitory shocks in the model. We use restricted estimators of the deterministic trend parameters to improve the efficiency of the existing quasi-differencing estimation method. Also, we propose the multivariate Beveridge-Nelson decomposition as a possible candidate for a robust estimator. We match the moments of the transitory components of the data and the model variables. We compute the moments analytically but additionally consider an indirect inference approach to estimate the model. Simulations show that proposed estimators perform well especially when the shocks are highly persistent yet stationary. In such cases, we improve upon existing methods in terms of root mean square error and bias.

C691: **Solving nonlinear rational expectations models by approximating the stochastic equilibrium system**
Presenter: **Michael Evers**, Bonn University, Germany

Dynamic stochastic rational expectations models are commonly solved by computing the linear solution from the deterministic equilibrium system. Therefore, the linear solution must be independent of the distribution of the exogenous shocks. It is proposed to compute the linear solution from a higher-order approximation to the true stochastic equilibrium system. The approximated equilibrium system (denoted by AES) is the k-th order Taylor polynomial in the exogenous disturbances at the deterministic equilibrium system when disturbances are nil. The AES is non-stochastic and preserves the nonlinearity in the endogenous variables, but it is linear in the first k moments of the exogenous disturbances. Hence, the linear solution to the AES - the steady state and the linear coefficients - fully captures the equilibrium implications of the stochastic environment up to the first k moments on existence and uniqueness, equilibrium dynamics, propagation of shocks, and the equilibrium distribution of the endogenous variables.

CS58 Room H2 STATISTICAL MODELLING IN BANKING AND INSURANCE REGULATIONS Chair: Gareth Peters

C496: **Bayesian daily tail-risk forecasting employing intra-day data**
Presenter: **Richard Gerlach**, University of Sydney, Australia
Co-authors: Cathy Chen

Bayesian methods have proved effective for forecasting distributional tails, including for financial Value-at-Risk and Expected Shortfall (ES). Intra-day measures, such as realized volatility, realized range and intra-day range, have shown promise to add efficiency and accuracy to return volatility forecasts and potentially to tail risk measures as well. The class of conditional autoregressive expectile (CARE) models is extended to directly incorporate intra-day measures as an input to forecast related tail risk measures. Adaptive Markov chain Monte Carlo sampling schemes are employed for estimation and forecasting. Models employing intra-day measures are favoured in an empirical study forecasting multiple financial return series, in particular during the recent global financial crises, relative to a raft of popular competing time series models and methods.

C539: **Onset of systemic fragility due to counterparty risk in a stylized banking system**
Presenter: **Tomaso Aste**, University College London, United Kingdom
Co-authors: Annika Birch

Interbank lending is essential to an efficient banking system; however lending exposes banks to counterparty risk. We show that, in a stylized

banking system, there exists a critical level of interbank lending above which the whole system becomes fragile. In such a state the banks operate normally with a small probability of default of each single bank but with a sizable probability of default of the entire system. This fragile state is irreversible: once defaulted there is a large cost to pay to stabilize the system back to an operating state. We compute this quantitatively for a cascade model using liabilities and assets to define banks' balance sheet. Banks are connected through an interbank lending network and, whenever a bank is distressed, their creditors can potentially become distressed themselves. We solve the problem analytically for a homogeneous system and numerically with simulations of more complex -realistic- systems calibrated on UK regulatory data. We quantify how financial stability of a banking system can be influenced by regulatory decisions, we discuss the effect of central bank actions -such as quantitative easing- and we determine the cost of rescuing the system using re-capitalisation.

C599: **Systemic crisis timeline using tails dependences**
Presenter: **Guillaume Bagnarosa**, University Paris Pantheon Sorbonne, United Kingdom
Co-authors: Matthew Ames, Gareth Peters

A sequential analysis of the recent financial crisis is proposed using the upper and the lower tails dependences associated to specific speculative portfolios and comparing it to leverage estimators available in the market. This statistical approach provides an indicator which could help regulators and central banks in detecting and measuring the risk originated in the financial sector through highly levered positions in specific financial assets. When bubbles start swelling, investors indeed use the money they borrowed from banks or directly from the market to invest it in portfolio of risky assets leading to a higher upper tail dependence among risky assets and a higher leverage in the market. However, during the bursting stage they simultaneously liquidate all their positions at the same time increasing accordingly the lower tail dependence between these very assets and reducing obviously the leverage in the market. We empirically show these phenomena using specific self-financing strategies in currencies and equities markets, which thus demonstrate the interest of tail dependences for measuring and forecasting systemic risk.

C315: **Sequential Monte Carlo for capital allocation**
Presenter: **Gareth Peters**, University College London, United Kingdom
Co-authors: Rodrigo Targino, Pavel Shevchenko

The task of capital allocation is an important quantitative challenge faced by practitioners in both banking and insurance environments. Recently, this issue has received increasing attention specifically with regard to the development of theoretical properties in the form of axioms that will result in a coherent capital allocation framework. This coherency in the allocation principle must be understood relative to a given risk measure, axioms have been developed that cover a variety of different risk measures that may be utilised to quantify the capital required under either the Basel II/III or the Solvency II capital adequacy standards for banks or insurers. In practical capital allocation settings, one should have the flexibility to model the individual risk processes in different divisions or business units of the bank or insurance institution with flexible dependence structures between the loss processes. In such cases, these axioms of coherent capital allocation result in allocation rules for capital risk measures such as Value at Risk (VaR) and Expected Shortfall (ES) which are challenging to evaluate in practice as they involve intractable integrations with respect to multivariate constrained distributions. It is the sole purpose of the work to start from the assumption that a risk model has been developed for an institution and capital derived. Given this starting point, we develop a new class of efficient rare event sampling techniques which we demonstrate have several advantageous features relative to asymptotic approximations and importance sampling based approaches recently developed.

CS62 Room D2 FILTERS WAVELETS AND SIGNALS II Chair: Stephen Pollock

C279: **Bayesian inference of autoregressive models**
Presenter: **Kostas Triantafyllopoulos**, University of Sheffield, United Kingdom
Co-authors: Dler Kadir

Autoregressive (AR) and more general ARMA models form a standard tool of statistical analysis of stationary time series data, with many applications to finance and commerce. We discuss simulation-based Bayesian approaches to the estimation and forecasting of AR models. In particular, we discuss Metrolpolis-Hastings MCMC schemes by considering a new class of prior distributions. In the centre of our study is our desire to place a prior distribution directly on the AR coefficients, hence avoiding indirect priors, placed on the partial autocorrelations or the roots of the characteristic polynomial, to ensure stationarity. For such a prior to commence we revisit the classic problem of stationarity conditions, dated back in the late 30s with the pioneering work of H. Wold. We are able to provide simple stationarity conditions involving the AR coefficients, which enable us to propose suitable prior distributions. The proposed methodology is illustrated by considering simulated and real data, utilising the AR(2) and AR(3) models. Some extensions and lines of future work will be discussed.

C555: **Convergence rates of sieve estimation in a univariate nonlinear cointegration model**
Presenter: **Maarten van Kampen**, Technical University Dortmund, Germany
Co-authors: Martin Wagner

Sieve estimation of a univariate nonlinear cointegration model is considered, with the sieve space consisting of polynomials. As is common in the cointegration literature, we allow for error serial correlation and regressor endogeneity. We derive the convergence rates for both the OLS series estimator as well as for a fully modified OLS-type series estimator.

C419: **Current account core periphery dualism in the EMU**
Presenter: **Tatiana Cesaroni**, Bank of Italy, Italy
Co-authors: Roberta De Santis

The role of financial integration in determining the so called Eurozone current account "core-periphery dualism" is investigated. To this end, the determinants of current account imbalances are analyzed for 22 OECD countries and 15 European Union (EU) members, comparing the behavior of core and peripheral countries using panel econometric models. Empirical evidence shows that within OECD and EU groups, financial integration indicators significantly contributed to explain current account positions dispersion. Whereas business cycle and budget balance seem to assume more importance over time, competitiveness seems to have played a minor role in Explaining CA dynamics after EMU.

C781: **A testing procedure for parameter constancy in stochastic volatility models**
Presenter: **Carlos Rivero**, Universidad Autonoma de Madrid, Spain
Co-authors: Juan del Hoyo, Guillermo Llorente

Rapid increases in conditional volatility followed by persistent behavior, autocorrelations which resemble long memory behavior, and premia associated to volatility jumps are widely acknowledged in financial market returns. To accommodate these empirical stylized facts it is necessary to introduce changing parameters into Stochastic Volatility (SV) models. A general method is developed, easy to implement, to test for constant parameters in SV models. The test has a well-known null asymptotic distribution free of nuisance parameters. Monte Carlo simulations demonstrate that it has good size and power properties. An empirical application is analyzed.

CS64 Room G2 REGIME CHANGE MODELING IN ECONOMICS AND FINANCE I Chair: Willi Semmler

C1009: **Melting down: systemic financial instability and the macroeconomy**
Presenter: **Kirstin Hubrich**, European Central Bank, Germany
Co-authors: Philipp Hartmann, Manfred Kremer, Robert J. Tetlow

The role of systemic financial instability in an empirical macro-financial model for the euro area is investigated, employing a richly specified Markov-Switching Vector Autoregression model to capture the dynamic relationships between a set of core macroeconomic variables and a novel indicator of systemic financial stress. We find that at times of widespread financial instability the macroeconomy functions fundamentally differently from tranquil times. Not only the variances of the shocks, but also the parameters that capture the transmission of shocks change regime, especially around times of high systemic stress in the financial system. In particular, financial shocks are larger and their effects on real activity propagate much more strongly during high-financial-stress regimes than during tranquil times. We find an economically important role of loan growth in the propagation of financial stress to the macroeconomy. We also show that prospects for detecting episodes of what we call financial fragility appear promising, although we argue that more research is required. We conclude that macroprudential and monetary policy makers are well advised to take these non-linearities into account.

C1101: **Overleveraging in the banking sector: evidence from Europe**
Presenter: **Willi Semmler**, New School for Social Research, United States
Co-authors: Frauke Schleer

Overleveraging of the banking sector has been viewed as one ot the main causes of the 2007-9 financial crisis and the subsequent great recession. It was also of considerable concern of the subsequent BIS regulatory policies resulting in Basil III and its request for higher capital requirements. It has become now of great importance for the planned European banking union. Overleveraging of the banking sector exposes the financial sector and the macroeconomy to vulnerabilitites but also, as critiques state, seem to constrain credit flows to private activitites. We present a model of overleveraging, as defined by the difference of actual and sustainable debt, conduct an empirical study on overleveraging for 40 banks in Europe, and study the vulnerabilities and credit contractions that can arise from this. We are using a nonlinear macroeconometric VSTAR model to evaluate the hypothesis that periods of high leveraging are accompanied by more severe credit constraints than periods of decreasing leveraging. We demonstrate this for individual banks as well as for country groups in Europe

C1045: **Money, banking and interest rates: monetary policy regimes with Markov-switching VECM evidence**
Presenter: **Giulia Ghiani**, Politecnico di Milano, Italy
Co-authors: Max Gillman, Michal Kejak

The Euler equation is used within a cash-in-advance economy, with endogenous velocity via a banking sector, to set out an equilibrium equilibrium determination of the nominal interest rate through the money supply growth rate along with other Taylor rule related variables.We then test that equilibrium condition using US postwar data and find cointegration of inflation, unemployment and the money supply with the Federal Funds rate. Our VECM approach includes a Markov-switching identification of regime shifts with three key regimes resulting: 1) one similar to Nber contractions, 2) one similar to Nber expansions, and 3) one similar to negative real interest rate periods including most of the post 2000 "Unconventional" period. Results indicate that the money supply growth rate strongly explains postwar interest rates along with inflation and unemployment. The key contribution is a money supply based regime classification alternative to chronological regimes tied to Fed Chairmen. Results show the sense in which claims can be lodged of model misspecification for certain classes of "Taylor rule estimations" for most of the postwar US if the money supply growth is left out.

C953: **Uncertainty and monetary policy in good and bad times**
Presenter: **Giovanni Caggiano**, University of Padova, Italy
Co-authors: Efrem Castelnuovo, Gabriela Nodari

A nonlinear VAR is employed to document the asymmetric reaction of real economic activity to uncertainty shocks. An uncertainty shock occurring in recessions triggers an abrupt and deep drop followed by a quick rebound and a temporary overshoot. Differently, a volatility shock hitting in expansions induces a milder slowdown, a longer-lasting recovery, and no overshoot. The employment of linear models is shown to offer a distorted picture of the timing and the severity of heightened uncertainty. Volatility shocks are found to be deflationary in recessions only. Monetary policy responds quite aggressively during bad times, and more mildly during booms. Counterfactual simulations point to monetary policy ineffectiveness during the first months after the shock, especially in recessions, and to policy effectiveness in the medium-term, especially during expansions. This holds true considering as policy tools both the federal funds rate and a long-term interest rate. Our results call for microfounded models admitting nonlinear effects of uncertainty shocks.

CS80 Room E2 FINANCIAL MODELLING Chair: Genaro Sucarrat

C141: **Maximum likelihood estimation of partially observed diffusion models**
Presenter: **Tore Selland Kleppe**, University of Stavanger, Norway
Co-authors: Jun Yu, Hans Julius Skaug

A maximum likelihood (ML) method to estimate partially observed diffusion models based on data sampled at discrete times is developed. The method combines two techniques recently proposed in the literature in two separate steps. In the first step, the closed form approach of Ait-Sahalia is used to obtain a highly accurate approximation to the joint transition probability density of the latent and the observed states. In the second step, the efficient importance sampling technique of Richard and Zhang is used to integrate out the latent states, thereby yielding the likelihood function. Using both simulated and real data, we show that the proposed ML method works better than alternative methods. The new method does not require the underlying diffusion to have an affine structure and does not involve infill simulations. Therefore, the method has a wide range of applicability and its computational cost is moderate.

C527: **Unbiased QML estimation of log-GARCH models in the presence of zero returns**
Presenter: **Genaro Sucarrat**, BI Norwegian Business School, Norway
Co-authors: Alvaro Escribano

A critique that has been directed towards the log-GARCH model is that its log-volatility specification does not exist in the presence of zero returns. A common "remedy" is to replace the zeros with a small (in the absolute sense) non-zero value. However, this renders Quasi Maximum Likelihood (QML) estimation asymptotically biased. We propose a solution to the case where the true return is equal to zero with probability zero. In this case zero returns may be observed because of non-trading, missing values and measurement error (e.g. discreteness approximisation error). The solution we propose treats zeros as missing values and handles these by combining QML estimation via the ARMA representation with an Expectation-Maximisation (EM) type algorithm. Monte Carlo simulations confirm that the solution corrects the bias. Several empirical applications illustrate that the bias-correcting estimator can make a substantial difference, and a case-study suggests past zeros and zero-probabilities do not have an effect on volatility dynamics.

C548: **Consistency and asymptotic normality in log-GARCH-X models**
Presenter: **Steffen Groenneberg**, BI Norwegian Business School, Norway
Co-authors: Benjamin Holcblat, Genaro Sucarrat

A generally applicable asymptotic framework is introduced for inference in log-GARCH-X models, i.e. a log-garch model which includes exogenous covariates. log-GARCH-X models always have an ARMA-X representation which we use for estimation. Under weak conditions we use this representation and some new results for ARMA-X residuals to provide the inference framework for the log-GARCh-X parameters.

C172: **Nonlinear shrinkage for portfolio selection: Markowitz meets goldilocks**

Presenter: **Michael Wolf**, University of Zurich, Switzerland

Co-authors: Olivier Ledoit

Markowitz portfolio selection requires estimates of (i) the vector of expected returns, and (ii) the covariance matrix of returns. Many successful proposals to address the first question exist by now. We address the second question. We promote a new nonlinear shrinkage estimator of the covariance matrix that is more flexible than previous linear shrinkage estimators and has 'just the right number' of free parameters to estimate (that is, the Goldilocks principle). It turns out that this number is the same as the number of assets in the investment universe. Under certain high-level assumptions, we show that the nonlinear shrinkage estimator is asymptotically optimal for portfolio selection in the setting where the number of assets is of the same magnitude as the sample size. For example, this is the relevant setting for mutual fund managers who invest in a large number of stocks. In addition to theoretical analysis, we study the real-life performance of our new estimator using backtest exercises on historical stock return data. We find that it performs better than previous proposals for portfolio selection from the literature and, in particular, that it dominates linear shrinkage.

Saturday 6.12.2014 14:35 - 16:15 Parallel Session E – ERCIM

ESI01 Room A1 ROBUSTNESS AGAINST ELEMENTWISE CONTAMINATION Chair: Peter Rousseeuw

E682: **A first step towards robust regression under elementwise contamination**
Presenter: **Andreas Alfons**, Erasmus University Rotterdam, Netherlands
Co-authors: Viktoria Ollerer, Christophe Croux

Regression methods that are robust against outlying data points are now widely used in applied data analysis. Such estimators simultaneously identify observations with a large residual and downweight them for estimating the regression parameters. Even if only one component of an observation causes the large residual, the whole observation is downweighted, which results in a loss of information. We propose the shooting S-estimator as a first step towards regression in situations where a large number of observations suffer from contamination in a small number of (not necessarily the same) variables. This new estimator combines the ideas of the coordinate descent algorithm (also known as shooting algorithm) and simple S-regression, making it robust against elementwise contamination.

E832: **Three-step robust regression for handling cell-wise and case-wise contamination**
Presenter: **Andy Leung**, University of British Columbia, Canada
Co-authors: Ruben Zamar

Traditional robust regression methods may fail when data contains cell-wise outliers. Cell-wise outliers are likely to occur together with case-wise outliers in modern datasets. The proposed method, called 3S-regression, proceeds as follows: first it uses a univariate filter to detect and eliminate extreme cell-wise outliers; second it applies a robust estimator of multivariate location and scatter to the filtered data to down-weight case-wise outliers; third it computes robust regression coefficients from the estimates obtained in the second step. The estimator is Fisher consistent and asymptotically normal at the central model under mild assumptions on the tail distributions of the predictors. Extensive simulation results show that 3S-regression is resilient to cell-wise outliers. It also performs well under case-wise contaminations when comparing with traditional high breakdown point estimators. The method implementation can handle both continuous and dummy predictors using an iterative algorithm.

E856: **Robust least angle regression with categorical variables**
Presenter: **Stefan Van Aelst**, Ghent University, Belgium
Co-authors: Ruben Zamar, Fred Zhang

Least angle regression (LARS) is a very efficient method for variable selection in linear regression based on least squares. LARS lists the candidate predictors in order of importance. To address the non-robustness of LARS it has been shown that LARS only depends on Pearson correlations. These correlations have been replaced by a fast and robust correlation measure to obtain robust LARS. By using only pairwise correlation, robust LARS can easily handle elementwise contamination in the data. Another problem of (robust) LARS is that it cannot handle well categorical variables. To overcome this problem we introduce a new method to measure correlation between a categorical and numerical variable. We show that this correlation measure can be used successfully in LARS and robust LARS which largely extends the applicability of this method. The results show good behavior of the extended LARS method compared to forward selection and group LASSO.

ES07 Room H1 STATISTICAL METHODS FOR DEPENDENT SEQUENCES Chair: Henryk Zaehle

E209: **Asymptotic behavior of the Whittle estimator for the increments of a Rosenblatt process**
Presenter: **Jean Marc Bardet**, University Paris Pantheon-Sorbonne, France
Co-authors: Ciprian Tudor

The purpose is the estimation of the self-similarity index of the Rosenblatt process by using the Whittle estimator. Via chaos expansion into multiple stochastic integrals, we establish a non-central limit theorem satisfied by this estimator. We illustrate our results by numerical simulations.

E742: **New methods for statistical functionals - with applications to weakly dependent sequences and long-memory processes**
Presenter: **Eric Beutner**, Maastricht University, Netherlands
Co-authors: Henryk Zahle

Using the notion of quasi-Hadamard differentiability that has recently been introduced and the theory of weighted empirical processes two general results are presented that can be used to obtain asymptotic properties for statistical functionals based on weakly dependent sequences and long-memory sequences. As examples we consider L- and U-statistics, in particular tail-dependent L-statistics as well as U-statistics with unbounded kernels. The results based on quasi-Hadamard differentiability provide new central and non-central limit theorems for L- and U-statistics. However, similar to previous approaches, the method of quasi-Hadamard differentiability does not allow for a general treatment of U-statistics based on long-memory sequences. To overcome the difficulties encountered by previous approaches, we derive a new representation for U-statistics. Using this representation and the theory of weighted empirical processes the asymptotic distribution of U-statistics can be derived by a direct application of the Continuous Mapping Theorem. This approach yields a new and powerful tool to derive the asymptotic distribution of very general U-statistics based on long-memory sequences.

E428: **Asymptotics of plug-in estimators under long-range dependence**
Presenter: **Jannis Buchsteiner**, Ruhr-University Bochum, Germany

A useful tool to establish the asymptotic behaviour of plug-in estimators under dependence is the functional delta method. Usually Hadamard differentiability is needed to apply this method but under some additional conditions it is still applicable if the functional fails to be differentiable. To these conditions belongs weak convergence of the underlying process with respect to a non-uniform sup-norm. We formulate such a limit theorem for the two parameter empirical process, where the observations are given by a subordinated Gaussian process, which exhibits long-range dependence. The limiting process is the product of a deterministic function and a Hermite-process and the convergence takes place in a subspace of $D(\mathbb{R} \times [0,1])$ equipped with a weighted sup-norm.

E846: **Inference of weighted V-statistics for non-stationary time series and its applications**
Presenter: **Zhou Zhou**, University of Toronto, Canada

The behavior of Fourier transforms is investigated for a wide class of non-stationary nonlinear processes. Asymptotic central and noncentral limit theorems are established for a class of non-degenerate and degenerate weighted V-statistics through the angle of Fourier analysis. The established theory for V-statistics provides a unified treatment for many time and spectral domain problems in the analysis of non-stationary time series, ranging from nonparametric estimation and specification to the inference of periodograms and spectral densities.

ES14 Room E1 SUFFICIENCY AND DATA REDUCTIONS IN REGRESSION Chair: Efstathia Bura

E258: **Semiparametric approach to dimension reduction**
Presenter: **Yanyuan Ma**, University of South Carolina, United States
Co-authors: Liping Zhu

A novel and completely different approach to dimension-reduction problems from the existing literature is provided. We cast the dimension-reduction problem in a semiparametric estimation framework and derive estimating equations. Viewing this problem from the new angle allows

us to derive a rich class of estimators, and obtain the classical dimension reduction techniques as special cases in this class. The semiparametric approach also reveals that in the inverse regression context while keeping the estimation structure intact, the common assumption of linearity and/or constant variance on the covariates can be removed at the cost of performing additional nonparametric regression. The semiparametric estimators without these common assumptions are illustrated through simulation studies and a real data example. The work has online supplementary material.

E264: **A general theory for nonlinear sufficient dimension reduction: formulation and estimation**
Presenter: **Kuang-Yao Lee**, Yale University, United States
Co-authors: Bing Li, Francesca Chiaromonte

A general framework for nonlinear sufficient dimension reduction is proposed, and its ramifications and scope are explored. It subsumes recent work employing reproducing kernel Hilbert spaces, and reveals many parallels between linear and nonlinear sufficient dimension reduction. Using these parallels we analyze the properties of existing methods and develop new ones. We begin by characterizing dimension reduction at the general level of σ-fields, and proceed to that of classes of functions, leading to the notions of sufficient, complete, and central dimension reduction classes. We show that, when it exists, the complete and sufficient class coincides with the central class, and can be unbiasedly and exhaustively estimated by a generalized slice inverse regression estimator (GSIR). When completeness does not hold, this estimator captures only part of the central class. However, in these cases we show that a generalized sliced average variance estimator (GSAVE) can capture a larger portion of the class. Both estimators require no numerical optimization, because they can be computed by spectral decomposition of linear operators. Finally, we compare our estimators with existing methods by simulation and actual data sets.

E319: **Exploiting structure to reduce and integrate high dimensional, under sampled "omics" data**
Presenter: **Francesca Chiaromonte**, The Pennsylvania State University, United States
Co-authors: Yang Liu

Dimension reduction techniques especially formulated for regression problems can be very useful in the analysis of high dimensional "omics" data. These techniques produce a small number of composite predictors that can then be used to construct effective models. Compared to other application settings however, high dimensional "omics" data are often characterized by (1) marked heterogeneity and structure, with both samples and features differing in terms of origin and/or information available on their nature or function, and (2) under-sampling, with the number of available samples rather small relative to the number of measured features. This data thus requires a dimension reduction approach that can simultaneously exploit known structure in samples and features and withstand under-sampling. Once formulated, such an approach will allows us to integrate data collected through multiple studies and diverse high-throughput platforms into unified and parsimonious regression models. We are currently in the process of defining theory and methods for this approach, building upon a number of recent developments in the field of Sufficient Dimension Reduction. Some of our preliminary results are illustrated through analyses of simulated and actual data.

E844: **Sufficient dimension reduction in functional regression**
Presenter: **Anne Francoise Yao**, University Blaise Pascal, France

The generalization of the sufficient dimension reduction method to functional framework is useful to find out the relevant functional subspace in regression problem (based on functional data). In the finite dimensional setting, it is used to improve the rate of convergence of the non-parametric regression estimators (by reducing the dimension of the regressor). But, in the functional setting, the interest of this approach mainly (but not exclusively) lies in the fact that if the observations belong intrinsically to an infinite dimensional space, in practice, they often belong to a small dimensional subspace. Then, the goal is to find the best (in some meaning to be specified) subspace for the semi-parametric model. Some estimators of the subspace of interest have been proposed in the non-parametric statistics literature. We will present the asymptotic behavior of some of them. Their rate of convergence as well as their use in some applications will be discussed.

ES20 Room P1 CHANGE-POINT ANALYSIS Chair: Serguei Dachian

E252: **Change detection for time series following generalized linear models**
Presenter: **Edit Gombay**, University of Alberta, Canada
Co-authors: Abdulkadir Hussein, Fuxiao Li

The considered models are of great practical importance as they are used in measuring health care performance, evaluating financial markets, analysing industrial processes, and in climate studies. We survey recent theoretical developments concerning logistic and other regression models that allow AR(p)-type dependence structure together with the presence of covariates. Conditions are set for the Maximum Partial Likelihood Estimator's existence and its convergence to the true value. We can prove that this convergence is at the optimal rate. The performance of the score vector of the partial likelihood function is analysed. We can use it for change detection and in sequential monitoring. Its usefulness will be demonstrated on data from clinical studies.

E267: **Fast change point analysis with applications to physiological series**
Presenter: **Pierre Bertrand**, University Clermont-Ferrand, France

New devices allow for the measurement of big series of physiological parameters, e.g. heart rate, skin conductance, motion, ... Outside laboratory, the environmental conditions can vary with time. Thus the series should be modeled as a stationary series with a piecewise constant mean. We are concerned with change point analysis to detect the change times. On the other hand, we now have access to series with big data, e.g. datasets of size $N = 100,000$, or larger. Thus, we have to focus on change point analysis for big series and the complexity of change point methods. The introduction of the Filtered Derivative method with p-value (FDpV) allows for change detection with both time and memory complexities of order $O(N)$. We firstly recall the FDpV method for change detection on the mean, and some generalizations to change detection on other structural parameters. Secondly, we recall cardiologists point of view on heart rate series. Thirdly, we apply FDpV method for change point analysis to heart rate series for shift workers, sportmen or emergency physicians. Eventually, we apply classification methods to disentangle the different levels of stress in the previous heart rate series.

E404: **Resampling methods in change point analysis**
Presenter: **Zuzana Praskova**, Charles University in Prague, Czech Republic

Critical values of change point tests in location and regression models are usually based on limit distribution of the respective test statistics under the null hypothesis. However, the limit distribution is very often a functional of some Gaussian processes depending on unknown quantities that cannot be easily estimated. In many situations, convergence to the asymptotic distribution is rather slow and the asymptotic critical values are not well applicable in small and moderate size samples. It has appeared that resampling methods provide reasonable approximations for critical values of test statistics for detection changes in location and regression models. Some recent results will be reviewed and proper bootstrap methods for dependent data discussed. Dependent wild bootstrap procedure for testing changes in linear model with weakly dependent regressors and errors will be proposed and its validity justified. Using the procedure in sequential change point testing will be also considered.

E397: **On limiting likelihood ratio processes encountered in statistical inference on the change-point location parameter**
Presenter: **Serguei Dachian**, Universite Blaise Pascal, France

Different change-point type models encountered in statistical inference for stochastic processes give rise to different limiting likelihood ratio processes. We review several such likelihood ratio processes and discuss recently discovered relations between them. We also pinpoint that, besides their theoretical interest, these relations equally have some practical implications. Finally, we illustrate the results by numerical simulations.

ES41 Room B1 ROBUST ESTIMATION IN EXTREME VALUE THEORY Chair: Armelle Guillou

E024: **Robust conditional variance and value-at-risk estimation**
Presenter: **Debbie Dupuis**, HEC Montreal, Canada

Losses due to idiosyncratic events can have a disproportionate impact on traditional value-at-risk estimates, upwardly biasing these estimates, increasing capital requirements and unnecessarily reducing the available capital and profitability of financial institutions. We propose new bias-robust conditional variance estimators based on weighted likelihood at heavy-tailed models, as well as value-at-risk estimators based on the latter and on volatility updated historical simulation. The new value-at-risk estimators use optimally chosen rolling window length and smoothing parameter value.

E049: **Bounded-influence robust estimation of copulas**
Presenter: **Roberto Molinari**, University of Geneva, Switzerland
Co-authors: Stephane Guerrier, Maria-Pia Victoria-Feser, Samuel Orso

Copula functions are very convenient for modelling multivariate observations. Popular estimation methods are the (two-stage) maximum likelihood and an alternative semi-parametric with empirical cumulative distribution functions of the margins. Unfortunately, they can often be biased whenever relatively small model deviations occur at the marginal and/or copula levels. We propose two robust estimators that do not share this undesirable feature. Since skewed and heavy tailed parametric marginals are considered in many applications, we also propose a bounded-bias robust estimator for such distributions that is corrected for consistency by means of indirect inference. In a simulation study we show that our robust estimators outperform the conventional approaches.

E185: **Detecting influential data points for the Hill estimator in Pareto-type distributions**
Presenter: **Dina Vanpaemel**, KU Leuven, Belgium
Co-authors: Mia Hubert, Goedele Dierckx

Pareto-type distributions are extreme value distributions for which the extreme value index $\gamma > 0$. Classical estimators for $\gamma > 0$, like the Hill estimator, tend to overestimate this parameter in the presence of outliers. The empirical influence function plot, which displays the influence that each data point has on the Hill estimator, is introduced. To avoid a masking effect, the empirical influence function is based on a new robust GLM estimator for γ. This robust GLM estimator is used to determine high quantiles of the data generating distribution, allowing to flag data points as unusually large if they exceed this high quantile.

E920: **On the harmonic moment tail index estimator**
Presenter: **Dieter Schell**, University of Konstanz, Germany
Co-authors: Jan Beran, Milan Stehlik

The aim is to derive asymptotic properties of the harmonic moment tail index estimator. The results depend on the first and second order tail properties of the underlying distribution and a tuning parameter which also determines the behaviour of the influence function. In comparison to classical tail index estimators a reduced sensitivity to large outliers is demonstrated by a small simulation study.

ES50 Room G1 DIRECTIONAL STATISTICS Chair: Thomas Verdebout

E335: **Improved kernel density estimation for directional data under rotational symmetry**
Presenter: **Christophe Ley**, Universite libre de Bruxelles, Belgium
Co-authors: Eduardo Garcia-Portugues, Davy Paindaveine, Thomas Verdebout

Rotational symmetry is an often encountered hypothesis when dealing with directional data on unit hyperspheres $\mathcal{S}^{k-1} := \{x \in \mathbb{R}^k : x'x = 1\}$. It is, for instance, naturally related to earth rotation, and most of the classical distributions on $\mathcal{S}^{k-1}$ are rotationally symmetric (e.g., the Fisher-von Mises-Langevin distribution). It is therefore natural, for problems where such a symmetry is known to be present, to take advantage of that information when using a Kernel Density Estimator (KDE). Such a shape-constrained KDE is the object of the work. The new KDE is obtained by means of a new operator, which we term "rotsymmetrizer": applied to a given directional KDE, it ensures the resulting estimator, denoted as RSKDE, is rotationally symmetric. The RSKDE improves on the classical KDE (under rotational symmetry), both by theoretical arguments (MSE) and computational ones (simulation study). The success of the RSKDE may well open up new research directions for shape-constrained KDEs in directional settings.

E461: **Bandwidth selection in nonparametric directional regression**
Presenter: **Eduardo Garcia-Portugues**, Santiago de Compostela, Spain
Co-authors: Rosa M. Crujeiras, Ingrid Van Keilegom, Wenceslao Gonzalez-Manteiga

New bandwidth selection rules for nonparametric kernel regression with directional predictor and linear response are explored. From the expressions for the conditional bias and variance for a recent proposal of a local linear estimator with directional predictor, new plug-in rules are obtained. In addition, different cross validation bandwidth selectors are adapted to this setting. The new selectors are compared in an extensive simulation study for different combinations of models, estimators, dimensions and sample sizes. Finally, the bandwidth selectors are applied to analyze real datasets.

E641: **Two proposals for circular order aggregation**
Presenter: **Cristina Rueda**, University of Valladolid, Spain
Co-authors: Miguel Fernandez, Sandra Barragan

The problem of obtaining a circular order on a set of n items using angular values from p heterogeneous data sets is addressed. The problem resembles the classical problem of determining the true order or rank among n objects using the ranks assigned by p independent judges. Although there exists a huge literature in ranking aggregation for Euclidean data, the problem is unexplored in the circular setting, where the Euclidean methods cannot be applied directly. We consider two original proposals. The first one is based on solving a Traveling Salesman Problem (TSP). The TSP approach for Circular Ordering Aggregation is formulated as a problem of searching the shortest tour in a graph where the vertices are the items to order and the lengths of the edges measure the aggregated pairwise relationships. The second is based on Hodge theory, which has been developed within the fields of graph theory, linear algebra and topology and that has been proposed recently to derive algorithms for rank aggregation using pairwise information. The originality behind this proposal is that of using triplet-wise input information. Besides some theoretical results for Hodge theory, we present an extensive simulation study to compare the different techniques.

E892: **CircSiZer for the assessment of significant features in nonparametric circular curve estimates**
Presenter: **Maria Oliveira**, Durham University, United Kingdom
Co-authors: Rosa M. Crujeiras, Alberto Rodriguez-Casal

A critical issue in nonparametric circular density and circular-linear regression estimation is the selection of the smoothing parameter. From a practical perspective, the concerns about the requirement of an appropriate smoothing value for constructing a density or regression estimator may discourage the use of nonparametric techniques. Exploring the estimators at different smoothing levels, by trying a reasonable range between oversmoothing and undersmoothing scenarios, will provide a thorough perception of the data structure. However, the significance of the observed features in the family of smoothed curves must be statistically assessed. This is the goal of CircSizer method. CircSiZer is a graphical tool based

on nonparametric curve estimates which considers a wide range of smoothing parameters and, for each of them, addresses the question of which features, like peaks and valleys, observed in the density or regression curves are really there and which ones are simply artifacts of the sampling noise. This method assesses the significance of such features by constructing confidence intervals for the derivative of the smoothed curve. Different ways of constructing the confidence intervals for the derivative will be discussed. The proposed technique will be illustrated with simulated datasets and real datasets from different scientific fields.

ES60 Room D1 SIMULTANEOUS EQUATION MODELS CONTROLLING FOR UNOBSERVED CONFOUNDING Chair: Giampierro Marra

E344: **On the assumption of joint normality in selection models: A flexible copula based approach for estimating HIV prevalence**
Presenter: **Mark McGovern**, Harvard University, United States
Co-authors: Till Barnighausen, Giampiero Marra, Rosalba Radice

Heckman-type selection models can be used to adjust HIV prevalence estimates for correlation between participation in testing and knowledge of HIV status. However, existing methods for selection on omitted confounders typically rely on strong parametric assumptions. We demonstrate a novel approach for relaxing the assumption of bivariate normality in selection models to correct for bias and inefficiency associated with model misspecification. In addition, we apply a ridge penalty method to correct for collinearity problems, and calculate confidence intervals using a Taylor-series expansion which simultaneously acknowledges uncertainty due to cluster effects and the presence of sampling weights. We apply this method to estimating HIV prevalence in the 2007 Zambian Demographic and Health Survey where 21% of men and 20% of women refuse to test. Using interviewer identity as the selection variable which predicts test consent but not HIV status, we show how to allow for non-linear association between the participation and outcome equations using flexible copula functions. For men, we estimate a population HIV prevalence of 21%, compared to 12% among those who consent to test. For women, the corresponding figures are 19% and 16%. This approach has many other potential applications where the assumption of missing at random is unrealistic.

E083: **Semi-parametric copula sample selection models for count response**
Presenter: **Karol Wyszynski**, University College London, United Kingdom
Co-authors: Giampiero Marra

Non-random sample selection arises when observations do not come from a random sample. Instead, individuals select themselves into (or out of) the sample on the basis of observed and unobserved characteristics. This problem can be addressed using sample selection models. In the methodological literature a lot of attention has been given to sample selection models with continuous response. At the same time, not much work has been attributed to sample selection models with count response. We will consider a copula-based sample selection model for count data with flexible covariate effects. Beforehand, some examples of sample selection models will be provided. Hence, the proposed sample selection model for count data will be described. The model will be illustrated on data originating from the United States Veterans' Administration. Finally, the blueprint for future work will be presented.

E451: **Asymptotics of penalized spline estimators in generalized sample selection model**
Presenter: **Malgorzata Wojtys**, Plymouth University, United Kingdom
Co-authors: Giampiero Marra

A sample selection model where the outcome equation and the selection equation take the form of a generalized additive model and the bivariate distribution of the outcomes of the two equations is expressed with the use of a copula is considered. Asymptotic theory for penalized spline estimators in such models is discussed. In particular, asymptotic bias and variance of such estimators is presented and the main aspects of their derivation are outlined.

E181: **Semiparametric bivariate regression with non-gaussian dependent errors**
Presenter: **Giuliano Galimberti**, University of Bologna, Italy
Co-authors: Giampiero Marra, Gabriele Soffritti

Bivariate gaussian linear regression models with dependent errors represent useful tools for dealing with endogeneity. An extension of this class of models is proposed, by allowing departures from both the Gaussian and the linear assumptions. On one hand, Gaussian mixture models for approximating the distribution of the error term are considered. On the other hand, smooth nonlinear effects for (a subset of) the exogenous regressors are introduced, by resorting to a regression spline approach. Parameter estimation is performed using the penalized maximum likelihood method. In particular, a penalized expectation-maximization algorithm is developed. Solutions for automatic selection of the number of mixture components and the smoothing parameters are suggested, as well as inferential procedures on model parameters. The usefulness of the proposed methodology is illustrated through the analysis of a real dataset.

ES65 Room L1 GOODNESS-OF-FIT TESTS Chair: Simos Meintanis

E321: **GoF tests for semi- and parametric hypotheses based on the probability weighted empirical characteristic function**
Presenter: **James Allison**, Northwest University, South Africa
Co-authors: Simos Meintanis, Leonard Santana

The finite-sample properties of certain procedures which employ the novel notion of the probability weighted empirical characteristic function are investigated. The procedures considered include a test for symmetry of the error distribution in regression models and a test for multivariate normality with independent observations. Along with the new tests, procedures based on the ordinary empirical characteristic function as well as other more well-known procedures are implemented.

E455: **Goodness-of-fit tests for finite mixture distributions**
Presenter: **Olivier Thas**, University of Ghent, Belgium
Co-authors: Thomas Suesse, John Rayner

Mixture distributions have become a very flexible and common class of distributions, used in a wide variety of applications. We present two tests of goodness-of-fit. The first test is classical in the sense that it does not make use of the mixture construction, whereas the second test does. By doing so, the latter test allows for assessing the quality of the fit of each component distribution separately. Asymptotic null distributions are provided and the usefulness of the tests is demonstrated in a simulation study and on some example data sets.

E516: **Goodness-of-fit tests for multivariate stable distributions based on the empirical characteristic function**
Presenter: **Emanuele Taufer**, University of Trento, Italy
Co-authors: Simos Meintanis, Joseph Ngatchou-Wandji

Goodness-of-fit testing for multivariate stable distributions is considered. The proposed test statistics exploit a characterizing property of the characteristic function of these distributions and are consistent under some conditions. The asymptotic distribution is derived under the null hypothesis as well as under local alternatives. Conditions for an asymptotic null distribution free of parameters and for affine invariance are provided. Computational issues are discussed in detail and simulations show that with proper choice of the user parameters involved, the new tests lead to powerful omnibus procedures for the problem at hand.

E301: **Inference procedures for discrete-valued time series**
Presenter: **Simos Meintanis**, University of Athens, Greece

Co-authors: Jan Swanepoel, Nikolai Uskakov

The aim is to review certain properties of probability weighted quantities and suggest a nonparametric estimator of the probability weighted characteristic function. Some properties of this estimator are studied and corresponding inference procedures for symmetry and for the two-sample problem are proposed.

ES92 Room F1 CURE MODELS AND COMPETING RISKS IN SURVIVAL ANALYSIS Chair: Anouar El Ghouch

E417: **Mixture cure models: a completely nonparametric approach**

Presenter: **Maria Amalia Jacome Pumar**, Universidade da Coruna, Spain

Co-authors: Ana Lopez Cheda, Ricardo Cao

A nonparametric approach is applied to the estimation in the mixture cure model, as an alternative to the current parametric and semiparametric methods in the literature. The derived nonparametric estimators, deeply based on the Beran estimator of the conditional survival function, are proved to be the local maximum likelihood estimators of the incidence and the latency. The choice of the optimal smoothing bandwidth is addressed by a cross validation procedure. The MSE and MISE of the proposed estimators are compared to those of existing semiparametric methods in a simulation study for small and moderate sample sizes. Finally, as a result of the cooperation of the research group MODES with the University Hospital of A Coruna (CHUAC), both the proposed and semiparametric methods are applied to a database of colorectal cancer in CHUAC patients to determine the prognosis based on age, among other variables.

E950: **Cox proportional hazard cure models with time-varying covariates**

Presenter: **Cedric Heuchenne**, University of Liege, Belgium

Co-authors: Alessandro Beretta, Paul Wilson

Cox proportional hazard models are extended where a portion of the population is unsusceptible to time-varying covariates. In this model, the incidence is a logistic function depending on time (through covariates). Such a relation can occur for example when studying time to bankruptcy of banks as a function of managerial practices: these are varying over time and influencing the incidence. We obtain an estimation that we can then interpret as the estimated probability to be fully immune (to bankruptcy) at each time. This information can be of particular interest in the context of banks rating, complementing the conditional hazard rate function in the susceptible population.

E1022: **Variable selection in a flexible parametric mixture cure model with interval-censored data**

Presenter: **Sylvie Scolas**, Universite catholique de Louvain, Belgium

Co-authors: Anouar El Ghouch, Catherine Legrand

Mild cognitive impairment (MCI) may be a precursor of Alzheimer disease or other dementia. Generally, within this field, it is assumed that if the follow-up time is long enough, then the event of interest will be observed for each individual. In our case, not everybody will show signs of impairment. We then say that a proportion of the population is cured, or long-term survivor. Also, patients come to scheduled interviews and thus we can only detect MCI to have appeared between two visits. That is, the database contains interval censored data. Thus, we propose to extend the existing survival models to the case where interval censored data and cure may be present. We present the method we want to use: to model event times (i.e. the latency part), we utilize an accelerated failure time (AFT) regression model, adapted to interval censored data, together with an extended generalized gamma (EGG) distribution for the error term of the AFT. In addition, modeling the cure proportion (i.e. the incidence part) is made by a logistic regression. Furthermore we show the good behavior of the method thanks to results of simulations. Then, we address some issues concerning variable selection in such a model and apply this method to our Alzheimer disease database, which consist in 241 at-risk patients followed-up between 1988 and 2008 with regular checks for the appearance of MCI.

E1259: **Testing an "exponential delay time model" against a "random sign censoring model" in reliability**

Presenter: **Jean-Yves Dauxois**, INSA-IMT Toulouse, France

Co-authors: Sarah Jomhoori, Fatemeh Yousefzadeh

An industrial system is considered subject to different causes of failure and different types of maintenance: a corrective maintenance is performed after a critical failure and a preventive maintenance can be performed in order to decrease the risk of critical failure. The recurrence of these types of maintenance has been often modeled in a competing risks framework. However, rather light statistical inference has been carried out in these models. In particular, there is a need to introduce statistical tests in order to help the engineers to select the model which better fits their data. Thus, we introduce a nonparametric test with the aim to decide between a Delay Time model with exponential distribution and a Random Sign model. We prove the asymptotic normality of our test statistic and we carry out a Monte Carlo simulation to learn how our test works on finite sample sizes. An application on a real dataset is also given.

ES101 Room M1 FUNCTIONAL REGRESSION MODELS AND APPLICATIONS Chair: Ana M. Aguilera

E205: **Generalized multilevel function-on-scalar regression and principal component analysis**

Presenter: **Jeff Goldsmith**, Columbia University, United States

Co-authors: Vadim Zipunnikov, Jennifer Schrack

Regression models for generalized, multilevel functional responses are considered: functions are *generalized* in that they follow an exponential family distribution and *multilevel* in that they are clustered within groups or subjects. This data structure is increasingly common across scientific domains and is exemplified by our motivating example, in which binary curves indicating physical activity or inactivity are observed for nearly six hundred subjects over five days. We use a generalized linear model to incorporate scalar covariates into the mean structure, and decompose subject-specific and subject-day-specific deviations using multilevel functional principal components analysis. Thus, functional fixed effects are estimated while accounting for within-function and within-subject correlations, and major directions of variability within and between subjects are identified. Fixed effect coefficient functions and principal component basis functions are estimated using penalized splines; model parameters are estimated in a Bayesian framework using `Stan`, a programming language that implements a Hamiltonian Monte Carlo sampler. Simulations designed to mimic the application indicate good estimation accuracy and inference with reasonable computation times for moderate datasets, in both cross-sectional and multilevel scenarios; code is publicly available. In the application we identify effects of age and BMI on the time-specific change in probability of being active over a twenty-four hour period; in addition, the principal components analysis identifies the patterns of activity that distinguish subjects and days within subjects.

E383: **P-spline smoothing for functional data with spatial dependence**

Presenter: **M Carmen Aguilera-Morillo**, Universidad Carlos III de Madrid, Spain

Co-authors: Maria Durban, Ana M. Aguilera

In functional data analysis sometimes is usual to find data (curves) which have some spatial dependence. For example, Canadian Weather is a very well known functional data set, which in some cases it was considered as a set of independent curves related to daily temperature and precipitation at 35 different locations in Canada averaged over 1960 to 1994. But this is a clear example of functional data presenting spatial dependence and in that sense it was studied by different authors making use of geostatistical techniques such as functional kriging. The aim is to study noisy functional data with spatial dependence. So, we focus on spatial functional variables whose observations are realizations of a spatial functional process. In that context, and in order to smooth the possible noise in the curves, a new smoothing method for functional data presenting spatial dependence

is proposed. This method consists of a P-spline approach for spatial functional data. As alternative to other smoothing methods (kriging, kernel smoothing, among others), this P-spline approach can be used to estimate the functional form of a set of sample paths, and also to predict the corresponding function at a new location within the plane of study. Finally, in order to test the good performance of the proposed method, it will be compared with functional kriging on an application to real data.

E535: **Goodness-of-fit tests for the functional linear model based on random projections**
Presenter: **Wenceslao Gonzalez-Manteiga**, University of Santiago de Compostela, Spain
Co-authors: Juan Antonio Cuesta-Albertos, Eduardo Garcia-Portugues, Manuel Febrero-Bande

Different goodness-of-fit tests for the null hypothesis of a functional linear model are given. The tests are based on the residual marked empirical process indexed by random projections of the functional data. The asymptotic distribution of the projected empirical process is obtained under general conditions on the functional estimator. This yields in particular the asymptotic distribution of the Kolmogorv-Smirnov and Cramer-von Mises tests for the functional linear model, considering estimation by functional principal components. The exact distribution of the tests is calibrated by bootstrap and the finite simple properties of the tests are illustrated by simulations as well as real data applications.

E541: **Bayesian hierarchical additive models for sparse functional data**
Presenter: **Mathew McLean**, Texas A and M University, United States
Co-authors: Fabian Scheipl, Giles Hooker, Sonja Greven, David Ruppert

The purpose is to discuss additive models for scalar-on-function regression involving sparse, noisy longitudinal data. The presented models allow for more complex response-predictor relationships than the commonly used functional linear model. In the first model, the conditional mean response given a functional predictor $X(t)$ is modelled as the integral with respect to t of $F(X(t),t)$ where F is an unknown bivariate function. In the second model, the conditional mean response additively includes unknown univariate functions of the functional principal component scores for $X(t)$. A typical approach in practice is to first estimate complete predictor trajectories from the irregular, noisy measurements and then fit a functional regression model as if the functional predictors had been completely observed. Using semiparametric Bayesian methods, hierarchical models will be introduced which allow for the complete functional predictor trajectories to be simultaneously estimated along with other model parameters, so that variability in the estimated predictors is automatically taken into account during model fitting. Monte Carlo and variational Bayes algorithms for fitting the models will be presented. For the second model, we will discuss the use of spike-and-slab priors for performing variable selection.

ES129 Room I1 STATISTICAL METHODS AND APPLICATIONS Chair: Luisa Cutillo

E523: **Analysis of positively correlated counts by using a hierarchical Gamma Poisson model**
Presenter: **Luisa Cutillo**, Fondazione Telethon - TIGEM, Italy
Co-authors: Annamaria Carissimo, Vincenzo Belcastro, Claudia Angelini

Many biological experiments, such as High-throughput sequencing based techniques, produces measurements on a discrete scale. A key step in exploring such data is the identification of dependencies between them, which is commonly achieved by correlation analysis. However, it is known that such techniques can produce unreliable results since the observed data take the form of relative fractions of genes or species, rather than their absolute abundances. We infer the correlation network of such data based on a hierarchical Poisson model with a gamma prior. In particular, we show that the proposed model accounts for over-dispersion and is able to catch positive correlations from the data. Despite the limit imposed on the structure of the correlation, positive correlations often arise in several applications. We will illustrate the results achieved on simulated and biological datasets.

E549: **Pathways identification in cancer survival analysis by network-based Cox models**
Presenter: **Antonella Luliano**, Consiglio Nazionale delle Ricerche, Italy
Co-authors: Annalisa Occhipinti, Claudia Angelini, Italia De Feis, Pietro Lio

Gene expression data from high-throughput assays, such as microarray, are often used to predict cancer survival. However, available datasets consist of a small number of samples (n patients) and a large number of gene expression data (p predictors). Therefore, the main challenge is to cope with the high-dimensionality. Moreover, genes are co-regulated and their expression levels are expected to be highly correlated. In order to face these two issues, network based approaches have been proposed. In our analysis, we compare four network penalized Cox models for high-dimensional survival data aimed to determine pathway structures and biomarkers involved in cancer progression. Using these network-based models, it is possible to obtain a deeper understanding of the gene-regulatory networks and investigate the gene signatures related to the cancer survival time. We evaluate cancer survival prediction to illustrate the benefits and drawbacks of the network techniques and to understand how patient features (i.e. age, gender and coexisting diseases-comorbidity) can influence cancer treatment, detection and outcome. In particular, we show results obtained in simulation and real cancer datasets using the Functional Linkage network, as network prior information.

E427: **Validation of community robustness**
Presenter: **Annamaria Carissimo**, Fondazione Telethon - TIGEM, Italy
Co-authors: Luisa Cutillo, Italia De Feis

The large amount of work on community detection and its applications leaves unaddressed one important question: the statistical validation of the results. We present a methodology able to clearly detect the truly significance of the communities identified by some technique, permitting us to discard those that could be merely the consequence of edge positions in the network. Given a community detection method and a network of interest, our procedure examines the stability of the partition recovered against random perturbations of the original graph structure. To address this issue, we specify a perturbation strategy and a null model to build a stringent statistical test on a special measure of clustering distance, namely Variation of Information. The test determines if the obtained clustering departs significantly from the null model, hence strongly supporting the robustness against perturbation of the algorithm that identified the community structure. We show the results obtained with the proposed technique on simulated and real dataset.

E704: **Analysis of two-way functional data: an application to the comparison of mortality ratio models**
Presenter: **Maria Francesca Carfora**, Istituto per le Applicazioni del Calcolo CNR, Italy
Co-authors: Luisa Cutillo, Albina Orlando

The literature on mortality forecasting is widely based on the Lee-Carter parametric model and its different extensions. In the original model, the authors seek to summarize the mortality rate m_{xt} as a function of both age x and time period t and, after estimating the model coefficients, they use statistical time series methods to forecast the mortality changes over time and finally to produce a forecast of age-specific mortality. We analyse some of these models on the Italian mortality rates from a smoothing perspective. Moreover, we compare their projections by using methods of functional data analysis.

ES77 Room O1 CONTRIBUTIONS TO CLASSIFICATION AND CLUSTERING Chair: Agustin Mayo-Iscar

E061: **Multiclass classification of Gaussian spatial data based on pairwise discriminant functions**
Presenter: **Kestutis Ducinskas**, Klaipeda University, Lithuania
Co-authors: Lina Dreiziene

Given a training sample, the problem of classifying Gaussian spatial data into one of multiple classes specified by Gaussian random field with different parametric means is considered. The classifier based on the plug-in pairwise discriminant functions with inserted ML estimators is investigated. This is the extension of the previous one from the two-class case to the multiclass case. The novel close form expressions for the actual error rate and approximation of the expected error rate incurred by proposed classifier are derived. These error rates are suggested as performance measures for the proposed classifier. The three-class case with feature modeled by scalar stationary Gaussian random field on regular lattice with exponential covariance function is used for the numerical analysis of the proposed classifier performance. The accuracy of the obtained approximation is checked through a simulation study for various parametric structure cases.

E1210: **Kernel multilogit algorithm for multiclass classification**
Presenter: **Oscar Dalmau**, Centre for Mathematical Research CIMAT, Mexico
Co-authors: Teresa Alarcon, Graciela Gonzales

An algorithm for multi-class classification is proposed. The soft classification problem is considered, where the target variable is a multivariate random variable. The proposed algorithm transforms the original target variable into a new space using the multilogit function. Assuming Gaussian noise on this transformation and using a standard Bayesian approach the model yields a quadratic functional whose global minimum can easily be obtained by solving a set of linear system of equations. In order to obtain the classification, the inverse multilogit-based transformation should be applied and the obtained result can be interpreted as a 'soft' or probabilistic classification. Then, the final classification is obtained by using the 'Winner takes all' strategy. A Kernel-based formulation is presented in order to consider the non-linearities associated with the feature space of the data. The proposed algorithm is applied on real data, using databases available online. The experimental study shows that the algorithm is competitive with respect to other classical algorithms for multiclass classification.

E536: **A fuzzy clustering for time series based on quantile autocovariances**
Presenter: **Borja Lafuente**, Universidade da Coruna, Spain
Co-authors: Jose A. Vilar

Unlike conventional clustering, fuzzy cluster analysis allows data elements to belong to more than one cluster by assigning membership degrees of each data to clusters. We propose a fuzzy K-medoids algorithm to cluster time series data by using an innovative dissimilarity measure. Specifically, dissimilarity between two time series is measured as the Euclidean distance between the corresponding estimated quantile autocovariance functions (QAF). The quantile autocovariances examine the serial dependence structure, so accounting for sophisticated dynamic features that conventional autocovariances are unable to detect. The algorithm works recursively as follows. Starting by selecting an initial set of medoids representing each cluster, distances between elements and medoids are computed using the QAF dissimilarity weigthed by membership degrees. The medoids are updated and the process is repeated until convergence is reached. The methodology is applied in several simulated scenarios considering different generating models, and compared with other existing clustering methods. In all cases, the proposed method reports very good results, outperforming the considered alternatives. Robustness against the type of underlying models is a salient point of the method. An illustrative application is also discussed.

E1135: **Fuzzy double k-means clustering for simultaneous classification of objects and variables**
Presenter: **M. Brigida Ferraro**, University La Sapienza, Rome, Italy
Co-authors: Maurizio Vichi

Two-mode clustering consists in simultaneously clustering modes (e.g., objects, variables) of an observed two-mode data matrix. This idea arises to confront situations in which objects are homogeneous only within subsets of variables, while variables may be strongly associated only on subsets of objects. There are many practical applications presenting the above situations, for example, DNA microarrays analysis and market basket analysis. Other applications include biology, psychology, sociology and so on. We focus on an extension of standard k-means, the double k-means, to simultaneously cluster objects and variables. We propose this model in a fuzzy framework and discuss the advantages of this approach. Finally, we check its adequacy by means of simulation and real case studies.

ES29 Room N1 DEPENDENCE MODELS AND COPULAS: THEORY II Chair: Wolfgang Trutschnig

E232: **Assignment of risk in a cost cooperative game induced by a modified Expected Shortfall**
Presenter: **Arsen Palestini**, Sapienza University of Rome, Italy
Co-authors: Mauro Bernardi

Coherent risk measures have been widely studied in recent years as a crucial instrument to assess individual institutions' risk. Those measures fail to consider individual institutions as part of a system which might itself experience instability and spreads new sources of risk to the market participants. We take into account a multiple institutions framework where some of them jointly experience distress events in order to evaluate their individual and collective impact on the remaining institutions in the market. To carry out this analysis, we define a new risk measure (SCoES), generalising the Expected Shortfall and we characterise the riskiness profile as the outcome of a cost cooperative game played by institutions in distress. Each institution's marginal contribution to the spread of riskiness towards the safe institutions is then evaluated by calculating suitable solution concepts of the game such as the Banzhaf–Coleman and the Shapley–Shubik values. In the empirical part of the work, we apply the proposed risk measure to a large panel of US companies belonging to different sectors of the Standard and Poor's Composite Index and compare our results with other systemic risk measures (CoVaR, SRISK, MES).

E906: **Characterization of copulas with given diagonal and opposite diagonal sections**
Presenter: **Enrique de Amo**, University of Almeria, Spain

In recent years special attention has been devoted to the problem of finding a copula C when its diagonal section δ and opposite diagonal section ω are known. For a given diagonal function δ and an opposite ω, we provide necessary and sufficient conditions on $\%\delta$ and ω for the existence of a copula with δ and $\%\omega$ as diagonal and opposite diagonal sections. We make use of techniques related to linear interpolation between the diagonals, the construction of checkerboard copulas, and linear programming. This result allows us to solve two open problems: to characterize the class of copulas where the knowledge of δ and ω determines the copula in a unique way, and to formulate necessary and sufficient conditions for each pair (δ, ω) to be the diagonal and opposite diagonal sections of a unique copula.

E972: **New measure of dependence from conditional variance**
Presenter: **Noppadon Kamnitui**, Chulalongkorn University, Thailand
Co-authors: Tippawan Santiwipanon, Songkiat Sumetkijakan

Conditional variance of two random variables motivates us to define, in terms of copulas, a measure of association which detects mutual complete dependence. Due to its inability to identify independence, we modify the measure into a measure of dependence using a subclass of shuffles of Min.

E1185: **A class of B-spline copulas: dependence structure and estimation**
Presenter: **Xiaoling Dou**, Waseda University, Japan
Co-authors: Satoshi Kuriki, Gwo Dong Lin, Donald Richards

A class of copulas based on B-spline basis functions is proposed. The proposed B-spline copulas include the Bernstein copula (Baker's distribution)

as a special case. The range of correlation of the B-spline copulas is examined, and the Frechet-Hoeffding bound is proved to be attained when the number of B-spline basis functions goes to infinity. The B-spline is well-known as a weak Tchebycheff system, from which the property of total positivity of order r (TP_r) follows for the maximum correlation case. For the estimation of parameters, the EM algorithm designed for the Bernstein copula is applicable with minor changes. A data set is analyzed as an illustrative numerical example.

ES45 Room Q1 BAYESIAN SEMI- AND NONPARAMETRIC MODELLING II **Chair: Antonio Lijoi**

E079: **A Bayesian nonparametric approach for time series clustering**
Presenter: **Luis E Nieto-Barajas**, ITAM, Mexico
Co-authors: Alberto Contreras-Cristan

A model-based clustering method for time series is presented. The model uses an almost surely discrete Bayesian nonparametric prior to induce clustering of the series. Specifically we propose a general Poisson-Dirichlet process mixture model, which includes the Dirichlet process mixture model as a particular case. The model accounts for typical features present in a time series like trends, seasonal and temporal components. All or only part of these features can be used for clustering according to the user. Posterior inference is obtained via an easy to implement Markov chain Monte Carlo (MCMC) scheme. The best cluster is chosen according to a heterogeneity measure as well as the model selection criterion LPML (logarithm of the pseudo marginal likelihood). We illustrate our approach with a dataset of time series of share prices in the Mexican stock exchange.

E087: **Bayesian feature allocation estimation**
Presenter: **Sara Wade**, University of Cambridge, United Kingdom
Co-authors: Zoubin Ghahramani

Feature allocation is an increasingly popular extension of clustering, which allows objects to belong to multiple clusters. A Bayesian treatment involves assigning a prior over the space of all feature allocations and computing the posterior. An important problem is how to summarize this posterior; indeed, often the first question one asks is what is an appropriate point estimate of the feature allocation based on the posterior. In practice, the last sampled feature allocation in a MCMC algorithm or the posterior mode is used as a point estimate. The former is undesirable, as it is highly susceptible to Monte Carlo error. The latter is the optimal estimate under the 0-1 loss function, which does not take into account similarity between two feature allocations and often suffers from computational drawbacks. Motivated by these issues, our aim is to develop more general and more appropriate loss functions and propose alternative estimates derived from minimization of the posterior expected loss. In particular, following the construction of Binder's loss for the clustering setting, we define basic principles for loss function and explore the class of functions satisfying these principles, starting from the simplest cases and pointing out key differences with the clustering case. Simulated examples are provided.

E221: **A nonparametric latent distance model for dynamic relational networks**
Presenter: **Isadora Antoniano-Villalobos**, Bocconi University, Italy
Co-authors: Maxim Nazarov, Sonia Petrone

Network and relational data arise in a variety of fields. They constitute the source of extensive literature and, specially in recent years, an active area of research among statisticians. One of the most popular approaches for the study of this type of data is based on a latent variable representation, justified under the assumption of row-column exchangeability for the error terms in a model which, in the general case, incorporates additional information through covariates. We present a new model for Bayesian inference on dynamic network data which extends the static latent variable representation by introducing a finite hidden Markov model driving the temporal evolution within the latent space. The model can be interpreted in terms of global time-varying phenomena affecting the dynamics of the network's edge density and other structural quantities of interest, while maintaining, at each point in time, the desirable properties provided by latent distance representations for static networks. The resulting model has a complex infinite-dimensional structure which makes it intractable. We therefore propose a computational algorithm that enables MCMC posterior inference.

E202: **Priors for random count matrices derived from a family of negative binomial processes**
Presenter: **Mingyuan Zhou**, University of Texas at Austin, United States
Co-authors: Oscar-Hernan Madrid-Padilla, James Scott

A family of probability distributions for random count matrices with a potentially unbounded number of rows and columns is defined. The three distributions are derived from the gamma-Poisson, gamma-negative binomial (GNB), and beta-negative binomial (BNB) processes. A key aspect of our analysis is the recognition that, although the proposed random count matrices are defined by a row-wise construction, their columns can be shown to be i.i.d. By analyzing these matrices' combinatorial structure, we describe how to construct a column-i.i.d. random count matrix one row at a time, and derive the predictive distribution of a new row count vector with previously unseen features. We describe the similarities and differences between the three priors, and argue that the greater flexibility of the GNB and BNB processes - especially their ability to model over-dispersed, heavy-tailed count data - makes these well suited to a wide variety of real-world applications. As an example of our framework, we construct a naive-Bayes text classifier to categorize a count vector to one of several existing random count matrices of different categories. The classifier supports an unbounded number of features, and unlike most existing methods, it does not require a predefined finite vocabulary to be shared by all the categories.

ES06 Room C1 BIG DATA ANALYSIS: PENALTY, PRETEST AND SHRINKAGE ESTIMATION I **Chair: S. Ejaz Ahmed**

E367: **Model selection in high-dimensional misspecified models**
Presenter: **Yang Feng**, Columbia University, United States
Co-authors: Pallavi Basu, Jinchi Lv

Model selection is vital to high-dimensional modeling in selecting the best set of covariates among a sequence of candidate models. Most existing work assumes implicitly that the model under study is correctly specified or of fixed dimensions. Both model misspecification and high dimensionality are, however, common in real applications. We investigate two classical Bayesian and Kullback-Leibler divergence principles of model selection in the setting of high-dimensional misspecified models. Asymptotic expansions of these model selection principles in high dimensions reveal that the effect of model misspecification is crucial and should be taken into account, leading to the generalized BIC and generalized AIC. With a natural choice of prior probabilities, we suggest the generalized BIC with prior probability ($GBIC_p$) which involves a logarithmic factor of the dimensionality in penalizing model complexity. We further establish the consistency of the covariance contrast matrix estimator in the general setting. Our results and new method are also supported by numerical studies.

E342: **A local ANOVA-type nonparametric test for trend synchronism in multiple time series**
Presenter: **Vyacheslav Lyubchich**, University of Waterloo, Canada
Co-authors: Yulia Gel

The problem of identifying joint trend dynamics in multiple time series, i.e. testing whether two or more observed processes follow the same common trend, is essential in a wide spectrum of applications, from economics and finance to climate and environmental studies. However, most of the available tests for comparing multiple mean functions either deal with independent errors or are applicable only to a case of two time series, which constitutes a substantial limitation in many high-dimensional studies. We propose a new nonparametric test for synchronism of trends exhibited by multiple linear time series where the number of time series N can be large but N is fixed. The core idea of our new approach is based on

employing the local regression test statistic, which allows to detect possibly non-monotonic non-linear trends. The finite sample performance of the new synchronism test statistic is enhanced by a nonparametric hybrid bootstrap approach. The proposed methodology is illustrated by simulations and a case study on climate dynamics.

E625: **Shrinkage estimation in an additive survival model**

Presenter: **Abdulkadir Hussein**, University of Windsor, Canada

Co-authors: Katrina Tomanelli, Severien Nkurunziza

Additive regression models are good alternatives to the Cox's PH model when the proportionality assumption is violated. We propose shrinkage estimators for the coefficients of Aalen's additive survival regression model. We compare such estimators to the LASSO and Ridge estimators via Monte Carlo simulations. We apply the methods to data on survival times of patients with primary billiary cirrhosis.

E895: **Fence methods for gene set selection in microarray studies**

Presenter: **JSunil Rao**, University of Miami, United States

Co-authors: Jiming Jiang, Thuan Ngyuen

Fence methods are procedures that isolate subgroups of correct models from incorrect ones, with the optimal model being a member of the isolated set. We extend fence methods to situations where a true model may not exist resulting in a new version called the invisible fence (IF). Focusing on gene set analysis of expression data, we develop a fast algorithm for the IF, and demonstrate a novel type of signal consistency of the method. We also develop a variation called the relative invisible fence (RIF) which adjusts for imbalances in the number of competing models across dimensions. We compare these approaches to other competing methods via simulations and apply them to the problem of tracking differential expression of biological pathways in colon cancer metastasis.

Saturday 6.12.2014 16:45 - 18:50 Parallel Session E – CFE

CS02 Room N2 VOLATILITY AND CORRELATION MODELLING FOR FINANCIAL MARKETS Chair: Cristina Amado

C059: **Assessing risk premium over time: Inference on GARCH-in-mean models with time-varying coefficients**
Presenter: **Gustavo Fruet Dias**, Aarhus University, Denmark

How the risk premium parameter varies over time is examined, shedding light on the behaviour of the risk aversion parameter during periods of financial distress. To accommodate a time-varying coefficient on the mean equation of a GARCH-in-mean model, we introduce the time-varying GARCH-in-mean (TVGARCH-in-mean) model, where the risk premium parameter is allowed to be a time-varying stochastic process. We propose an estimation strategy based on the Nonlinear Iterative Least Squares (NL-ILS) estimator that successfully estimates the time-varying risk premium parameter. A Monte Carlo study shows that the proposed algorithm has good finite sample properties. We investigate the time-varying risk premium using excess returns computed using the CRSP index. We document that the risk premium parameter is indeed time-varying and shows high degree of persistence. We find that the time-varying risk premium parameter is statistically different from zero on 46.5% of the observations when considering the monthly frequency. Considering point-wise analyses, we find that weekly estimates of the time-varying risk premium parameter anticipate bear market phases and business cycles, being therefore countercyclical. Furthermore, we show that the TVGARCH-in-mean specification generates risk premium functions that can virtually assume any shape.

C095: **Handling conditional correlation GARCH models in R: The ccgarch2 package**
Presenter: **Tomoaki Nakatani**, Hokkaido University, Japan

The ccgarch2 package is an add-on package for the open source statistical environment R and is designed to provide functions for estimation and simulation of variants of conditional correlation (CC-) GARCH models. A couple of R packages are available for handling the major variants of the CC-GARCH models. An advantage of ccgarch2 over the other existing packages is that it allows for modeling volatility spillovers in the GARCH part of the model. In particular, in the bivariate model, the estimating functions are constructed in such a way that it can capture negative volatility spillovers. In addition to inheriting many of the functionalities from its predecessor, ccgarch2 improves user-interface by defining classes and associated methods. Numerical optimization of the likelihood function is now carried out by the solnp() function in the Rsonlp package, which makes it possible to impose non-linear restrictions on the parameters. These restrictions are necessary to keep the time-varying conditional covariance matrices positive definite as well as to keep the sequence of conditional variances stationary.

C349: **A latent factor model for panels of realized volatilities**
Presenter: **Roxana Halbleib**, University of Konstanz, Germany
Co-authors: Aygul Zagidullina

A parsimonious way of modeling the long-memory in the dynamics of large panels of realized volatilities is provided. The model is based on the Kalman-filter approach and models the dynamics of panels of realized volatilities within a latent factor structure with dynamic factors. The aggregation of short memory latent factors is able to capture in a parsimonious framework both the commonality of the panel dynamics, but also the long memory of the underlying series. Both on simulated and real data, the model outperforms the standard approaches, such as ARFIMA and HAR.

C190: **Realizing commodity correlations and the market Beta**
Presenter: **Harry Vander Elst**, Universite libre de Bruxelles, Belgium
Co-authors: Niels Hansen, Asger Lunde, Kasper Olesen

The purpose is to document evidence of large economic gains in the commodities space from including high-frequency based measures of realized volatility and correlation in the investor's information set. Using both model-free and model-based methods for the intraday transaction data of 6 selected commodity futures, we present three main findings: (1) correlations and market betas were highly volatile over the last 7 years, but are in general back at the same level as in 2007; (2) dynamic models based on intraday information are superior to more restricted alternatives; (3) economic gains can be achieved by more sophisticated methods.

C755: **Testing for tail breaks in bank equity index returns: international evidence**
Presenter: **Paulo Rodrigues**, Bank of Portugal, Portugal
Co-authors: Joao Nicolau

Based on recently proposed methods, the aim is to analyse the tail properties of international and sectoral bank equity index returns and whether changes (breaks) in the tail indices have occurred. After the collapse of Lehman Brothers in 2008 and the subsequent shocks that featured the Great Recession and the European sovereign debt crisis, financial institutions worldwide have faced an extremely uncertain operating environment. Repeated rating downgrades, sharp regulatory changes, and widening funding-spreads have led to unprecedented levels of market volatility in the sector, particularly, in peripheral Europe. While it is perfectly clear that the mean of the volatility process shifted as a consequence of the financial crisis, it is possible that other characteristics of the data generating process, may have changed as well. The severity of the consequences on the banking industry provides us with the perfect ground to detect potential changes in the tail index of these series, bringing new evidence to the field. Given that unusually large movements in economic and financial time series seem to occur more often than would be implied by normality and given that in practice the exact distribution of the innovations is not known, it has become increasingly important to obtain insights on the tail properties of data.

CS103 Room F2 CONTRIBUTIONS TO APPLIED ECONOMETRICS I Chair: Hilde C. Bjornland

C544: **Composite indices based on partial least squares**
Presenter: **Jisu Yoon**, University of Goettingen CRC PEG, Germany
Co-authors: Tatyana Krivobokova, Stephan Klasen, Axel Dreher

Composite indices are constructed using partial least squares (PLS) and principal component analysis (PCA) focusing on the treatment of non-metric variables. A composite index is built typically as a linear combination of variables and the quality depends on the weighting. PCA calculates weights from the relationships between variables, whereas PLS calculates weights from the relationships between outcome variables and variables building composite index. One often has non-metric variables in applications, which need appropriate treatments to apply PCA or PLS. We review the treatments of non-metric variables in PCA and PLS in the literature and compare their performances by means of a simulation study. PLS with binary coding performs better than others in terms of prediction and additionally it is easier to interpret. We build two wealth indices and a globalization index as applications.

C921: **How do business cycles become global? Common shocks or spillovers?**
Presenter: **Christopher Otrok**, University of Missouri, United States
Co-authors: Thomas Helbling, Raju Huidrom, Ayhan Kose

A model is developed that allows us to decompose cross-country comovement into a portion due to common shocks and a portion due to spillovers from one country to another. Our definition of a spillover is a shock that originates in one country and subsequently impacts other countries. We argue that to identify a spillover you must first isolate country specific shocks from global shocks. A spillover in our model is a country specific

shock that transmits to other countries. Our framework for measuring global comovement , isolating country specific shocks, and then quantifying the impact of spillovers is a parametric dynamic factor model. Our model is flexible enough that we can also identify the channels of transmission. Spillovers from one country can either affect all countries symmetrically or each country have a unique impact on a country. Empirically we find that spillovers are generally from the US to other countries. The channel of transmission has historically been that US shocks affect other countries symmetrically through the global factor, not through country specific channels. When we apply the model to subsamples of the data we find that spillovers from the US increased dramatically during the great recession.

C1117: **Testing for optimal monetary policy via moment inequalities**
Presenter: **Laura Coroneo**, University of York, United Kingdom
Co-authors: Valentina Corradi, Paulo Santos Monteiro

The specification of an optimizing model of the monetary transmission mechanism requires selecting a policy regime, commonly commitment or discretion. We propose a new procedure for testing optimal monetary policy, relying on moment inequalities that nest commitment and discretion as two special cases. The approach is based on the derivation of bounds for inflation that are consistent with optimal policy under either policy regime. We derive testable implications that allow for specification tests and discrimination between the two alternative regimes. The proposed procedure is implemented to examine the conduct of monetary policy in the United States economy.

C1171: **Identification of network effect in the buyer-seller network**
Presenter: **Ryohei Hisano**, University of Tokyo, Japan
Co-authors: Takayuki Mizuno, Takaaki Ohnishi, Tsutomu Watanabe

Origins of aggregate fluctuation have attracted researchers for decades. However the source of such behavior has not been fully explained provoking recent debates. We analyze a unique data set which gives detailed description of the buyer-seller network of over one million Japanese firms. By analyzing this huge data set we provide evidence that a certain proportion of the cyclic behavior in average log growth rate of firms could indeed be explained by the buyer-seller trade network. We would estimate the proportion of this network effect, explain issues surrounding identification and also analyze to what extent firms mitigate shocks by renewing their trade relations. And last but not least, we would also present a model with endogenous link renewal.

C1250: **Monetary policy, leaning and concern for financial stability**
Presenter: **Hilde Bjornland**, BI Norwegian Business School, Norway
Co-authors: Leif Brubakk, Junior Maih

The recent financial crisis and subsequent recession has initiated new interest in the question to what extent an inflation targeting central bank should take financial stability concerns into account when conducting monetary policy. Higher house prices and debt accumulation over the business cycle increases the risk of abrupt and severe disturbances to activity and inflation somewhat further ahead. This justifies increasing interest rates today more than would otherwise have been warranted. We ask to what extent a Central Bank should take account on financial stability when conducting monetary policy, by leaning against the wind. To explore this issue, we estimate a simple New Keynesian model for a small open economy that allows for two different regimes; normal times and a period with financial stress. We allow for Markov switching between the two regimes. Our proposed mechanism assumes the costs of entering the periods of financial stress is endogenously linked to some measure of financial imbalances. We find that some degree of leaning will reduce the costs of recession further ahead.

CS06 Room B2 NON-STATIONARY TIME SERIES AND THE BOOTSTRAP Chair: Peter Boswijk

C042: **Inference in VARs with conditional heteroskedasticity of unknown form**
Presenter: **Carsten Trenkler**, University of Mannheim, Germany
Co-authors: Ralf Brueggemann, Carsten Jentsch

A framework is proposed for asymptotically valid inference in stable vector autoregressive (VAR) models with conditional heteroskedasticity of unknown form. In particular, we study the joint asymptotic behaviour of estimators of both the VAR mean parameters as well as of the parameters in the covariance matrix of the error term vector. Thereby, we provide the necessary asymptotic results for important VAR inference methods, like e.g. structural impulse responses, that depend both on the level and variance parameters. Conventional asymptotic as well as bootstrap inference is covered. We show that the moving block bootstrap (MBB) results in asymptotically valid inference in the presence of conditional heteroskedasticity. This is not true for the commonly applied wild and pairwise bootstrap schemes if inference on (functions) of the error terms' variance parameters is of interest. These bootstrap approaches fail because they do not correctly replicate the relevant fourth moments' structure of the error terms in case of conditional heteroskedasticity. Our simulation results indicate that finite-sample (impulse response) inference can be strongly distorted in the presence of conditional heteroskedasticity.

C066: **Determining the co-integration rank in heteroskedastic VAR models of unknown order**
Presenter: **Luca De Angelis**, University of Bologna, Italy
Co-authors: Giuseppe Cavaliere, Anders Rahbek, Robert Taylor

The aim is to investigate the asymptotic and finite sample properties of different methods for estimating the co-integration rank in integrated vector autoregressive systems of unknown order driven by heteroskedastic shocks. We allow for both conditional and unconditional heteroskedasticity of a very general form. We establish the conditions required such that standard information criterion-based methods, such as the BIC, when employed either sequentially or jointly, can be used to consistently estimate both the co-integration rank and the lag order. We also extend the available large sample theory for the conventional sequential approach of Johansen and the associated wild bootstrap implementation thereof to the case where the lag order is unknown. We show that these methods remain valid under heteroskedasticity and an unknown lag length provided this is first chosen by a consistent method. The finite sample properties of the different methods are investigated in a Monte Carlo simulation study. For the simulations DGPs considered, we find that the best performing methods are a wild bootstrap implementation of the Johansen procedure implemented with BIC selection of the lag length and an approach which uses the BIC to jointly select the lag order and the co-integration rank.

C187: **Bootstrap inference on deterministic trends in the presence of heteroskedastic and possibly integrated errors**
Presenter: **Stephan Smeekes**, Maastricht University, Netherlands
Co-authors: Robert Taylor

Inference on deterministic trends is investigated in the presence of unconditional heteroskedasticity and uncertainty about the order of integration. We propose a novel bootstrap method that combines the wild bootstrap for robustifying against heteroskedasticity with a switching mechanism that prevents the usual invalidity of the bootstrap for local-to-unity processes and delivers asymptotically valid inference whether a (near) unit root is present or not. Though the method is conservative for autoregressive parameters close to unity, our simulations show that with a careful choice of the tuning parameters this effect can be minimized and good power properties can be retained.

C266: **Bootstrap-based inference on cointegration parameters in heteroscedastic vector autoregressions**
Presenter: **Anders Rahbek**, University of Copenhagen, Denmak
Co-authors: Giuseppe Cavaliere, Peter Boswijk, Robert Taylor

It is well established that the shocks driving many key macro-economic and financial variables display time-varying volatility. We consider estimation and hypothesis testing on rank, coefficients of the co-integrating relations and the adjustment coefficients in vector autoregressions

driven by both conditional and unconditional heteroskedasticity of a quite general and unknown form in the shocks. We show that the conventional results in a previous work for the maximum likelihood estimators and associated likelihood ratio tests derived under homoskedasticity do not in general hold in the presence of heteroskedasticity. As a consequence, standard confidence intervals and tests of hypothesis on these coefficients are potentially unreliable. Solutions to this inference problem based on the use of the wild bootstrap are discussed. These do not require the practitioner to specify a parametric model for volatility, or to assume that the pattern of volatility is common to, or independent across, the vector of series under analysis. We formally establish the conditions under which these methods are asymptotically valid. A Monte Carlo simulation study demonstrates that significant improvements in finite sample size can be obtained by the bootstrap over the corresponding asymptotic tests in both heteroskedastic and homoskedastic environments. An application to the term structure of interest rates in the US illustrates the difference between standard and bootstrap inferences regarding hypotheses on the co-integrating vectors and adjustment coefficients.

C280: **Adaptive testing for a unit root with nonstationary volatility**
Presenter: **Peter Boswijk**, University of Amsterdam, Netherlands
Co-authors: Yang Zu

Recent research has emphasised that permanent changes in the innovation variance (caused by structural shifts or an integrated volatility process) lead to size distortions in conventional unit root tests. It has been shown that these size distortions may be resolved using the wild bootstrap. We first derive the asymptotic power envelope for the unit root testing problem when the nonstationary volatility process is known. Next, we show that under suitable conditions, adaptation with respect to the volatility process is possible, in the sense that nonparametric estimation of the volatility process leads to the same asymptotic power envelope. Implementation of the resulting test involves cross-validation and the wild bootstrap. A Monte Carlo experiment and an empirical application show that the asymptotic results are reflected in finite sample properties.

CS16 Room E2 MODELLING FINANCIAL CONTAGION Chair: Raffaella Calabrese

C1065: **Using agent-based network models to assess financial contagion**
Presenter: **Christoffer Kok**, European Central Bank, Germany

The purpose is to present how network formation techniques can be used to study contagion among financial intermediaries and between the financial sector and real economic agents. Agent-based modelling approaches will be exploited in order to form networks between different economic agents (banks in particular) and to run counter-factual simulations to assess how shocks (exogenous or policy-induced) can propagate through the financial system and beyond while taking into account agents' heterogeneous and dynamic behaviour. This approach allows for a more realistic description of the propagation mechanism than more traditional network-based contagion models. A number of practical applications illustrating how agent-based network models can be used to study financial contagion in a policy context is presented.

C1258: **Systemic risk in financial networks**
Presenter: **Stefano Battiston**, UZH, Switzerland

Financial inter-linkages play an important role in the emergence of financial instabilities and the formulation of systemic risk can greatly benefit from a network approach. We focus on the role of linkages along the two dimensions of contagion and liquidity, and we discuss some insights that have recently emerged from network models. With respect to the issue of the determination of the optimal architecture of the financial system, models suggest that regulators have to look at the interplay of network topology, capital requirements, and market liquidity. With respect to the issue of the determination of systemically important financial institutions the findings indicate that both from the point of view of contagion and from the point of view of liquidity provision, there is more to systemic importance than just size. In particular for contagion, the position of institutions in the network matters and their impact can be computed through stress tests even when there are no defaults in the system.

C1238: **Global macro-nets: systemic risk from within country sectoral imbalances and cross border exposures of national banks**
Presenter: **Sheri Markose**, University of Essex, United Kingdom

The lack of integration of the financial sector with the non-financial macro-economic sectors within national economies and in a cross-border setting has been identified as a major drawback of traditional macro-economic modelling which led to a lack of early warning for the 2007 crisis and the onset of the Great Recession. The innovative cross border macro-net framework integrates the exposures of national banking systems to economic sectors within countries such as financial sector, public sector and the non-bank private sector. Sectoral imbalances in the flow of funds within countries, especially with the substantial growth of the financial and housing sectors relative to other sectors in the US, UK and the Euro zone, have been noted by many as a major cause for the Great Recession. The network approach can show the interconnections between cross border financial flows and the macro-economic sectors within countries. In principle, sub-networks at any degree of granularity can be incorporated in the new macro-net model. Our systemic risk analytics can identify not only systemically important debtor countries, but also which sectors within them are the cause of this. We illustrate the efficacy of the macro-net model for the analysis of the recent and ongoing Euro zone crisis.

C1044: **Early warning of global financial instability: a spectral systemic risk index**
Presenter: **Simone Giansante**, University of Bath, United Kingdom

The purpose is to address the issue of providing systemic risk measures with early warning capability for global financial flows by employing network-based spectral analysis. We investigate the empirical topological structure of global banking in all its aspects by focusing on the foreign claims of 22 national banking systems during 2005-2010 provided by BIS consolidated banking statistics. The global banking networks show small-world features, with low average shortest path length, fat tail connectivity distribution that enables faster and easy propagation of contagion and high degree of disassortativity. Following the Eigen-Pair approach introduced previously, we construct a Spectral Systemic Risk Index for global flows by specifying the matrix of foreign liabilities of countries to major country banking systems, relative to their bank equity capital. The global network becomes markedly unstable in the first quarter of 2007, well before the global financial crisis, providing early warning signals and remaining generally unstable in the post-crisis, though the instability decreased as connectivity falls. The liabilities of certain systemically important countries relative to aggregate banking equity of counterparty countries become far too elevated, contribute to the global instability. The corresponding right and left eigenvectors are used to rank Systemically Important Debtor Countries (SIDCs), those who pose threats to other countries banking systems, and vulnerable banking systems respectively.

C1003: **Early-warning signals of topological collapse in interbank networks**
Presenter: **Tiziano Squartini**, Sapienza University of Rome, Italy
Co-authors: Iman van Lelyveld, Diego Garlaschelli

The financial crisis marked a paradigm shift, from traditional studies of individual risk to recent research on the systemic risk generated by whole networks of institutions. However, the reverse effects of realized defaults on network topology are poorly understood. We analyze the Dutch interbank network over the period 1998-2008, ending with the global crisis. We find that many topological properties, after controlling for overall density effects, display an abrupt change in 2008, thus providing a clear but unpredictable signature of the crisis. By contrast, if the intrinsic heterogeneity of banks is controlled for, the same properties undergo a slow and continuous transition, gradually connecting the crisis period to a much earlier stationary phase and highlighting a gradual built-up phase starting three years in advance of the crisis. The pre-crisis phase is thus an early-warning signal of the upcoming topological collapse. This early-warning signal begins in 2005 and is preceded by an even earlier period of "risk auto-catalysis" characterized by an anomalous number of debt loops. In order to further detect these remarkable "earlier precursors" an even more complex model is required, controlling for reciprocated interactions.

CS18 Room O2 STATISTICAL SIGNAL PROCESSING IN ASSET MANAGEMENT Chair: Serge Darolles

C743: **Mixture of experts for binary classification: application to the S&P500 index prediction**
Presenter: **Marc Mitri**, UBP, France
Co-authors: Emmanuelle Jay, Stephan Clemencon

Financial time series often exhibit multimodal probability distributions, which makes difficult for a single global model to learn input-target relationships existing in each regime. A dynamic mixture of experts model is proposed, combining a neural network able to quantize and cluster the input space into disjoint sets (Self-Organizing Maps), with an ensemble machine learning algorithm for binary classification called AdaBoost. We explore this divide-and-conquer approach consisting in local fitting (one AdaBoost expert for each input region) in order to forecast the S&P500 weekly movements using well known technical and fundamental indicators. Our findings provide empirical evidence that (i) this hybrid model is able to identify different market states and automatically select relevant variables in each regime, and (ii) the mixture of experts approach outperforms single global expert model in term of prediction accuracy.

C851: **Practical uses of signal processing in asset management**
Presenter: **Rafael Molinero**, Molinero Capital Management, United Kingdom

The purpose is to approach the use of digital signal processing from the practical side and explain how some these algorithms are used for trading. We will cover the strengths and weaknesses of these algorithms, as well as the increased difficulty driven by the high noise levels in financial markets. We will also try to illustrate how hedge funds may use digital signal processing to identify trading opportunities.

C888: **Trend-following meets Risk-Parity**
Presenter: **Nick Baltas**, UBS, United Kingdom

Trend-following strategies take long positions in assets with positive past returns and short positions in assets with negative past returns and are usually constructed using futures contracts across all asset classes. These strategies have historically exhibited great diversification benefits during dramatic market downturns. The conventional weighting scheme that is used in multi-asset class allocations instructs equating the ex-ante volatility of all assets in the portfolio (known as inverse-volatility weighting scheme or volatility-parity scheme). However, this approach ignores pairwise correlations and can turn out to be suboptimal in an environment of increasing correlations. Following a jaw-dropping performance in 2008, a trend-following strategy has failed to generate strong returns over the most recent period that has been characterised by large degree of co-movement even across asset classes. Extending the conventionally long-only risk-parity portfolio construction technique, we construct a long/short trend-following strategy that makes use of risk-parity principles. Not only do we enhance the performance of a simple volatility-parity trend-following strategy, but we also show that this enhancement in mostly driven by the more sophisticated weighting scheme in periods of increased correlation.

C1196: **Simulating and analyzing order book data: the queue-reactive model**
Presenter: **Mathieu Rosenbaum**, University Pierre and Marie Curie, France
Co-authors: Weibing Huang, Charles-Albert Lehalle

Through the analysis of a dataset of ultra high frequency order book updates, a model is introduced which accommodates the empirical properties of the full order book together with the stylized facts of lower frequency financial data. To do so, we split the time interval of interest into periods in which a well chosen reference price, typically the mid price, remains constant. Within these periods, we view the limit order book as a Markov queuing system. Indeed, we assume that the intensities of the order flows only depend on the current state of the order book. We establish the limiting behavior of this model and estimate its parameters from market data. Then, in order to design a relevant model for the whole period of interest, we use a stochastic mechanism that allows for switches from one period of constant reference price to another. Beyond enabling to reproduce accurately the behavior of market data, we show that our framework can be very useful for practitioners, notably as a market simulator or as a tool for the transaction cost analysis of complex trading algorithms.

C1294: **On the joint dynamics of equity and bond - a no arbitrage dynamic asset pricing approach**
Presenter: **Linlin Liu**, University of Notre Dame, United States

The aim is to study the dynamics of the cross section of bond yields and stock return in a joint framework by developing a common stochastic discounting factor with which the investor discount expected future cash flows. In order to identify the interconnection between the stock and bond market, vector autoregressive (VAR) analysis is conducted on the joint dynamics of stock returns, the term structure of interests and the macroeconomic variables. Risk factors for the asset returns are extracted using principal component analysis (PCA). The pricing relations are derived using the Gaussian Term Structure framework and the model implied returns and yields are determined by a set of no arbitrage restrictions. It is found that the set of risk factors for the stock market and bond market are un-spanned by each other and the macro risk factors are also un-spanned by the asset returns. Therefore only incorporate risk factors of one market when constructing the stochastic discount factor misses information of the market expectation about future returns conveyed by the asset dynamics in another market and in the macro economy. To the extent that stock (bond) returns co-vary with the bond (stock) risk factors, the estimates of stochastic discount factor will need to take into account of this correlation.

CS21 Room A2 BEHAVIOURAL AND EMOTIONAL FINANCE: THEORY AND EVIDENCE Chair: Richard John Fairchild

C224: **Determining the customer costs of using personal current accounts**
Presenter: **John Ashton**, Bangor University, United Kingdom
Co-authors: Andros Gregoriou

The aim is to examine the influence of offering an overdraft facility on the customer costs of using a personal current account (also termed checking accounts). This assessment informs the wider debate as to whether overdraft use is a significant factor in paying for current account use within 'free banking' systems. A UK data set of 222 current accounts, recorded monthly between 1995 and 2011 is used in combination with interest rates from 1,200 instant access deposit accounts offered simultaneously by the same firms. We use a panel framework to undertake the econometric analysis encapsulating contemporaneous correlation amongst UK current accounts. Our results do not support predictions that cross-subsidies flow from overdraft users to other current account customers.

C578: **Between fear and hope: optimal portfolio choice in a model combining expected utility and safety first preferences**
Presenter: **XiaoHua Chen**, University of Bath, United Kingdom
Co-authors: Richard Fairchild, Gulnur Muradoglu

A weighted average expected utility and safety first utility function is constructed, where fear underlies the concern for security and hope underlies the concern for potential. We demonstrate how the weights allocated to the expected utility preference and safety first preference affect the optimal portfolio choice in the mean-variance framework at low or high level of weighting on expected utility - an investor allocates all of his wealth in the near risk-free asset, or allocates a relatively stable percentage in the risky market portfolio. However, there is a critical point of the weighting on expected utility, where the investor experiences a large 'jump' out of the near risk-free asset towards the risky market portfolio.

C762: **Information processing and non-Bayesian learning in financial markets**
Presenter: **Stefanie Schraeder**, University of Lausanne, Switzerland

Ample empirical and experimental evidence documents that individuals place greater weight on information gained through personal experience –

a phenomenon called "availability bias". I embed this bias in an overlapping generations equilibrium model in which the period that investors first enter the market establishes the starting point of their experience history. The difference in the individuals' experience leads to heterogeneity among agents and perceived noise trading. The model captures several empirical findings. It explains why returns on high-volume trading days tend to revert. Furthermore, it provides explanations for a high trading volume, a connection between trading volume and volatility, excess volatility, and overreaction and reversal patterns. Consistent with empirical evidence, young investors buy high and sell low, trade frequently, and obtain lower returns. For intraday trading, it predicts a high trading volume around the opening hours, especially for cross-listed stocks.

C885: **Entrepreneurial under-diversification: over optimism and overconfidence**
Presenter: **Enrico Maria Cervellati**, Universita di Bologna, Italy
Co-authors: Pierpaolo Pattitoni, Marco Savioli

Previous studies claim that overconfidence and over optimism may help explain the decision to become entrepreneur or to overinvest in their already existing company. While these behavioral biases have been extensively used in the theoretical behavioral corporate finance (e.g., managerial overconfidence effects on M&As activity), there are still few theoretical contributions in the field of entrepreneurial finance. We intend to fill this gap, theorizing how these biases lead entrepreneurs to underestimate the risk and to overestimate the expected returns associated with their private company, and thus to overinvest in it. We consider an entrepreneur that has to choose her portfolio allocation, i.e., which part of her wealth to invest in her private company and which in the stock market. We include overconfidence and over optimism as parameters in our model, and show how they affect a risk-return analysis. Using this parameterization, we measure the bias in the entrepreneur portfolio allocation induced by the biases, explaining entrepreneurial under-diversification. A simulation analysis allows us to calculate overconfidence and over optimism levels that, assuming certain model parameters, are implicit in the entrepreneurial portfolio choices. Our approach allows an immediate and explicit measure of entrepreneurial under-diversification that could also be used in empirical analyses.

C1308: **Victory desease, earnings and hindsight bias: An experimental study**
Presenter: **Brian Kluger**, University of Cincinnati, United States
Co-authors: Patricia Chelley-Steeley, Jim Steeley

The hypothesis to test is that hindsight bias, the propensity for subjects to recall probabilities estimates as being more accurate than they actually were, is connected to the phenomenon of overconfidence following superior investment performance. We conduct experiments where subjects estimate the likelihoods for a series of future events, and are then allowed to buy state contingent claims whose values are linked to the same events. We find that hindsight bias is greater for events where subjects earned more through the purchases of the state-contingent claims.

CS47 Room P2 CYCLICAL COMPOSIT INDICATORS Chair: Gianluigi Mazzi

C090: **Econometric filters**
Presenter: **Stephen Pollock**, University of Leicester, United Kingdom

A variety of filters that are commonly employed by econometricians are analysed with a view to determining their effectiveness in extracting well-defined components of economic data sequences. These components can be defined in terms of their spectral structures -i.e. their frequency content- and it is argued that the process of econometric signal extraction should be guided by a careful appraisal of the periodogram of the detrended data sequence. Whereas it is true that many annual and quarterly economic data sequences are amenable to relatively unsophisticated filtering techniques, it is often the case that monthly data that exhibit strong seasonal fluctuations require a far more delicate approach. In such cases, it may be appropriate to use filters that work directly in the frequency domain by selecting or modifying the spectral ordinates of a Fourier decomposition of data that have been subject to a preliminary detrending.

C334: **Advanced spectral methods for macroeconomic indicators**
Presenter: **Michael Ghil**, Ecole Normale Superieure, France
Co-authors: Andreas Groth, Lisa Sella, Gianna Vivaldo

Macroeconomic growth dynamics is studied via non-equilibrium modeling and multivariate time series analysis of business cycles. The work is motivated by the trend towards "integrated models" that couple natural and socio-economic phenomena; it includes the analysis of a set of economic indicators for the United States and for several European countries. Recent model findings show the macroeconomic response to natural disasters in a non-equilibrium setting to be more severe during expansions than during recessions. Cyclic behavior in the U.S. and European economies is analyzed using multivariate singular spectrum analysis (M-SSA); M-SSA enables one to identify and describe nonlinear trends and dominant cycles, including seasonal and interannual components; it provides frequency information on multivariate data sets and complements pure analysis in the time domain. Analyzing nine aggregate indicators for the years 1954-2005 confirms that the U.S. economy changes significantly between intervals of growth and recession, with higher volatility during expansions. Next, M-SSA is applied to three countries of the European Union (Italy, The Netherlands, and the United Kingdom) and it identifies similar spectral components among the indicators being analyzed. Finally, the analysis is extended to include several U.S. economic indicators and the way they affect the European market.

C354: **Anticipating business-cycle turning points in real time using density forecasts from a VAR**
Presenter: **Sven Schreiber**, Macroeconomic Policy Institute IMK and Free U Berlin, Germany

For the timely detection of business-cycle turning points we suggest to use medium-sized linear systems (subset VARs with automated zero restrictions) to forecast the relevant underlying variables, and to derive the probability of the turning point from the forecast density as the probability mass below (or above) a given threshold value. We show how this approach can be used in real time in the presence of data publication lags and how it can capture the part of the data revision process that is systematic. Then we apply the method to US and German monthly data. In an out-of-sample exercise (for 2007-2012/13) the turning points can be signalled before the official data publication confirms them (but not before they happened in reality).

C435: **Using state-level data as predictors of National recessions: a model-averaging approach**
Presenter: **Danilo Leiva-Leon**, Bank of Canada, Canada
Co-authors: Pierre Guerin

The aim is to estimate and forecast U.S. business cycle turning points with state-level data. The probabilities of recession are obtained from univariate and multivariate regime-switching models based on a pairwise combination of national and state-level data. We use three different combination schemes to summarize the information from these models: (i) Bayesian model averaging, (ii) Dynamic model averaging, and (iii) weights based on the optimal prediction pools. Our best specification provides timely updates of the U.S. business cycles. In particular, the estimated turning points from this specification largely precede the announcements of business cycle turning points from the NBER business cycle dating committee, and compare favorably with existing models.

C512: **A unified framework for euro area and member countries real-time business cycle analysis**
Presenter: **Gian Luigi Mazzi**, Eurostat, Luxembourg
Co-authors: Monica Billio, Jacques Anas, Laurent Ferrara

An integrated framework for the euro area and member countries' business cycle analysis is presented. It relays on an analytical framework called ABCD allowing for a unified treatment of business and growth cycle with the possibility of incorporating also the acceleration cycle. According to this framework we developed a set of historical dating for euro area and member countries based on a simple non-parametric dating rule applied

to GDP, industrial production and unemployment rate. The series of historical dating are updated quarterly and they are considered final after 4 years. They are complemented by a set of coincident turning points indicators based on MS-VAR models to timely detect the occurrence of turning points. Both historical dating and the coincident indicators provide a variety of descriptive and parametric measures offering a detailed picture of cyclical movements. Such measures include the severity of recessions, the degree of convergence and synchronisation as well as the diffusion of turning points. This framework is used to study the post-recession behaviour of the euro area and its largest economies. The presentation ends with some ideas for a graphical and easily understandable way to carry out business cycle analysis.

CS55 Room G2 VOLATILITY MODELS AND THEIR APPLICATIONS Chair: Yasuhiro Omori

C251: **The SIML estimation of integrated covariances and hedging coefficients under micro-market noise and random sampling**
Presenter: **Naoto Kunitomo**, University of Tokyo, Japan
Co-authors: Hiroumi Misaki

For estimating the integrated volatility and covariance by using high frequency data, the Separating Information Maximum Likelihood (SIML) method has been proposed when there are micro-market noises. The SIML estimator has reasonable finite sample properties and asymptotic properties when the sample size is large under general conditions with non-Gaussian processes or volatility models. We show that the SIML estimation is useful for estimating the integrated covariance and hedging coefficient when we have micro-market noise and financial high frequency data are randomly sampled. The SIML estimation is consistent and has the stable convergence (i.e. the asymptotic normality in the deterministic case) and it has reasonable finite sample properties with these effects.

C405: **Network clustering of Japanese stock returns with multivariate GARCH model**
Presenter: **Takashi Isogai**, Bank of Japan, Japan

The high dimensional correlation structure of over 1,400 Japanese stock returns is analyzed to find a more data-oriented grouping than the Japanese standard sector classification. Every stock return is filtered by a univariate ARMA-GARCH model to ensure that volatilities are separated from residuals which are i.i.d. distributed, and a correlation matrix of the standardized residuals is calculated assuming a CCC (constant conditional correlation)-type multivariate model. Then, an undirected network of the stock returns is built from the correlation matrix, which is divided into several homogeneous and balanced groups by recursive spectral clustering of the network. A network based group detection technique is applied to the hierarchical divisive clustering. The clustering result is compared with the standard sector classification to explore statistically how these two categories are linked. We also perform random portfolio simulations to confirm if the grouping can actually improve portfolio risk control. The dynamic change of the correlation matrix is examined under a reduced size of the network, where a small number of stocks are selected as representatives for every group identified. We mainly focus on the stock returns; however, our method can be applicable for clustering other fat-tailed financial asset returns.

C406: **Predicting executions in high-frequency trading**
Presenter: **Ko Sugiura**, Keio University, Japan
Co-authors: Teruo Nakatsuma, Kenichiro McAlinn

High-frequency trading is widely practised in the financial markets and becomes one of the most important factors in price formulation of financial assets. At the same time, a huge amount of data on high-frequency transactions, so-called tick data, became accessible to both market participants as well as academic researchers, which paved the way for studies on the efficacy of the high-frequency trading and the micro-structure of the financial markets. One of the well-documented stylized facts on high-frequency transaction data is bid-ask bounce. Bid-ask bounce is a phenomenon that execution prices tend to move back and forth between the best-ask and the best-bid. We try to develop a new time series model to capture the characteristics of the bid-ask bounce and use it to predict executions in high-frequency trading.

C623: **Moment-based estimation of stochastic volatility models in the presence of intraday seasonality**
Presenter: **Isao Ishida**, Konan University, Japan
Co-authors: Virmantas Kvedaras

GMM estimators based on moments of realized variance and other realized measures constructed from high-frequency intraday price observations have been used to estimate the spot variance models for financial assets. We show by simulation the extent to which intraday volatility seasonality, neglected in previous studies, may bias GMM estimators and how various seasonality filters perform in reducing the bias.

CS57 Room C2 RISK MEASURES Chair: Katerina Panopoulou

C178: **Decomposition of the asset return distribution by joint autoregressive quantile models**
Presenter: **Evangelia Mitrodima**, University of Kent, United Kingdom
Co-authors: Jim Griffin, Jaideep Oberoi

A family of dynamic joint quantile models is proposed that accounts for both the scale and the shape of the conditional asset return distribution. Inter-quantile range is used as a convenient basis for the estimation of the scale of quantiles. By using a small number of quantile estimates we approximate the distribution of asset returns. Using daily returns on an index and two stocks we produce rolling one-day-ahead joint quantile forecasts out-of-sample for several benchmark models for the left and the right tail of the distribution. We conduct comparisons based on in-sample and out-of-sample criteria and we find that the proposed models exhibit a superior performance. In particular, we find that the time series of quantile forecasts better describes the evolution of the right and the left tail in-sample. We also find that the proposed models produce accurate expected number of exceedances out-of-sample. These findings indicate a better trade-off in the modelling of the scale and the shape of the distribution.

C393: **A role for GlueVaR risk measures under the Solvency II framework**
Presenter: **Jaume Belles-Sampera**, University of Barcelona, Spain
Co-authors: Montserrat Guillen, Miguel Santolino

An applied perspective on risk measures is taken. Concretely, it is intended to provide European insurance supervisory authorities with tools to track, year after year, changes on the risk profile of those insurance undertakings which use internal models. These tools are based on comparisons between risk aversion characteristics linked to Value-at-Risk (VaR), the regulatory risk measure of reference, and to GlueVaR risk measures. Particular GlueVaR risk measures are chosen to conduct the comparisons, being the selection procedure itself a contribution. Simplifying, candidates are required to satisfy two conditions: 1) GlueVaR must be tail-subadditive and 2) capital requirements based on GlueVaR should not extremely differ from those based on VaR. Both VaR and GlueVaR may be understood as Choquet integrals. In case of multiple candidates arising from the previous procedure, the final selection of the GlueVaR is based on indicators linked to the Choquet integral. Additional information about the underlying risk is provided by these indicators. Several concerns on considering the confidence level of VaR as the main risk aversion feature when measuring risk are highlighted.

C402: **Do realized measures improve VaR and ES forecasts**
Presenter: **Christos Argyropoulos**, University of Piraeus, Greece
Co-authors: Ekaterini Panopoulou

Recent financial turbulence has initiated the reevaluation of the Value at Risk measure and its suitability. Moving one step further, regulatory authorities raised the issue of a future switch to the coherent Expected Shortfall measure in order to alleviate these empirical discrepancies. One of

the main doubts especially in the Market Risk scheme is the inadequacy of conditional variance modeling to depict the current level of volatility and more importantly the poor reaction to new information. To this end two major approaches stand out in recent literature. The first approach consists of the traditional time series analysis implemented on a variety of realized measures while the second seeks to incorporate the additional "realized" information to the conditional variance modeling. Both approaches produce gains in volatility forecasting despite the inherent limitations of realized measures. Our findings suggest that the use of Realized Measures improves the performance of both quantile (Value at Risk) and Expected Shortfall forecasts. These improvements are particularly important from a risk management perspective.

C592: **Exchange rate volatility forecasting: a multivariate realized-GARCH approach**
Presenter: **Elena Dumitrescu**, Paris West University, France
Co-authors: Janine Balter, Peter Hansen

A new simple and parsimonious way is proposed to jointly model and forecast the returns, realized and conditional measures of volatility and correlation of different foreign exchange markets while accounting for the no-arbitrage condition. Our multivariate realized GARCH model combines the simplitude and flexibility of the GARCH approach and the advantages of high-frequency (intraday) data in terms of statistical precision in the estimation of volatility. We hence introduce a multivariate specification that jointly models the dynamics of the conditional and realized measures of covariance while accounting for spillover effects.It makes it possible to investigate the dynamics of exchange rates volatility and the propagation of shocks in foreign exchange markets. Most importantly, the multivariate realized GARCH model allows us to forecast the volatility and correlations one-step and multi-step ahead. The empirical analysis looks at the EUR/USD and USD/JPY exchange rates from January 2005 to December 2012. The superior out-of-sample performance of our method with respect to that of existing models indicates that it is a useful tool for forecasting risk in forex markets.

C614: **Volatility during the financial crisis through the lens of high frequency data: a realized GARCH approach**
Presenter: **Denisa Georgiana Banulescu**, University of Orleans, France
Co-authors: Peter Reinhard Hansen, Zhuo Huang, Marius Matei

The aim is to study the financial volatility during the global financial crisis and use the largest volatility shocks to identify major events during the crisis. Our analysis makes extensive use of high-frequency (HF) financial data for the modelling of volatility and, importantly, for determining the timing within the day when the largest volatility shocks occurred. The latter helps us to identify the event(s) that can be associated with each of these shocks, and serves to illustrate the benefits of using high-frequency data. Some of the largest volatility shocks coincide, not surprisingly, with the bankruptcy of Lehman Brothers on September 15, 2008 and Congress's failure to pass the Emergency Economic Stabilization Act on September 29, 2008. The day with the largest volatility shock was February 27, 2007 - a date where Freddie Mac announced a stricter policy for underwriting subprime loans and a date that was marked by a crash on the Chinese stock market. However, the intraday HF data shows that the main culprit for this shock was a computer glitch in the trading system. On the other hand, the days with the largest drops in volatility can in most cases be related to interventions by governments and central banks.

CS60 Room D2 CONTRIBUTIONS ON TIME SERIES ECONOMETRICS I Chair: Manabu Asai

C1189: **Are you sure that you took the right model? Estimating impulse responses under model uncertainty**
Presenter: **Jan Lohmeyer**, Maastricht University, Netherlands
Co-authors: Jean-Pierre Urbain, Franz Palm

The focus is on how to take model uncertainty into account when estimating univariate impulse responses. We consider the issue of selecting an appropriate model, or a combination of models, when all the models considered are at best approximations of the true underlying data generating process. We study the case of univariate impulse reponse functions based on finite order AR models when the underlying DGP is an ARMA model, which is not part of the set of models considered and estimated. We then compare the model selection methods with several estimator averaging methods by simulations. Both, the model selection as well as the selection of averaging weights, are based on a number of different criteria and methods like equal weighting, AIC, BIC, FIC, and C_p. Simulation evidence indicates that no model selection method based on different model selection criteria is found to perform notably better (in terms of mean squared error of the estimated impulse responses) than the rest over a large range of the parameter space. Preliminary results show that averaging methods (among other least squares averaging) stabilize the estimators and should be preferred to selecting a single model from the set of models entertained.

C1048: **Can a stochastic cusp catastrophe model explain housing market crashes?**
Presenter: **Juanxi Wang**, University of Amsterdam, Netherlands
Co-authors: Cees Diks

Like the patterns of stock market prices, house prices also exhibit temporary bubbles and crashes in the history. One possible explanation for such abrupt changes in stock market prices is the catastrophe model. However, due to the deterministic nature of catastrophe theory, applications to real market data are rare. It remains a question whether the catastrophe model can be used to explain the changes in the housing market. The aim is to fit a stochastic cusp catastrophe model to housing market data. Using the deviations of house prices from their fundamental value and the quarterly data on long term government interest rates, we fit the stochastic cusp catastrophe model to different countries. Moreover, we support our methods of estimation by fitting the model to simulation results from a housing market model with heterogeneous beliefs. Our results show that the long term interest rate serves as a bifurcation parameter in the housing market of different countries. It highlights the key role the interest rate policy plays on the housing market.

C1066: **Integrated modified OLS estimation andfixed-b inference for one-nonlinear-variable cointegrating polynomial regressions**
Presenter: **Peter Grabarczyk**, Technical University Dortmund, Germany
Co-authors: Martin Wagner

Integrated modified OLS (IM-OLS) parameter estimation and inference in cointegrating polynomial regressions (CPR) is considered, where only one of the integrated regressors enters with powers larger than one. We thereby extend the IM-OLS estimation principle introduced for linear cointegrating relationships. We investigate the properties of the estimator and parameter tests based upon it by means of a simulation study in which we also compare its performance with FM-OLS and D-OLS. Finally we apply the method to estimate the enviromental Kuznets curve for SO_2 and CO_2 emissions over the period 1870-2009.

C1173: **State space modeling of seasonal Gegenbauer processes with long memory**
Presenter: **Gnanadarsha Dissanayake**, University of Sydney, Australia
Co-authors: Shelton Peiris, Tommaso Proietti

The approximation of a seasonal Gegenbauer autoregressive moving average (GARSMA) process, characterised by long memory, using a finite order moving average model is considered. The approximating model is cast in state space form so that the parameters are estimated by pseudo maximum likelihood via the Kalman filter. Alternatively, an autoregressive approximation can be considered. A crucial issue is establishing the order of this approximation that is optimal for the purposes of estimating the unknown parameters and forecasting. An extensive Monte Carlo experiment is executed to address this issue. Finally, the proposed state space methodology is applied to real data, and compared with other similar time series methods proposed in the literature. Based on this study, we conclude that the proposed state space methodology provides a solution to efficient modelling and optimal forecasting of GARSMA processes.

C362: **Testing for neglected nonlinearity in economic time series: radial basis function network model**
Presenter: **Jinu Lee**, Queen Mary University of London, United Kingdom

In recent years, nonlinear models have been widely used to explain relationship among economic variables instead of a traditional linear model. The aim is concerned with testing for nonlinearity in economic time series. We revisit a neural network test using a radial basis function as an activation function in hidden units in order to improve its performance regarding to size and power. Particularly, we suggest a cross validation method to address how to choose the parameters of radial basis function which is neglected in the literature yet known to affect the accuracy of its approximation to arbitrary functional form. Monte Carlo experiments are conducted to present that the tuned basis function parameters can serve a significant complement to improve the size and power of the test for the conditional mean of univariate time series. A comparison with popular nonlinearity tests is also made to demonstrate the superiority of the new neural network test. The usefulness of the test is shown applying to empirical economic data.

CS100 Room M2 EMPIRICAL APPLICATIONS IN MACROECONOMICS AND TIME SERIES ANALYSIS Chair: Barbara Rossi

C070: **Emerging markets diversification benefits and FX risks in a globalizing world**
Presenter: **Milan Nedeljkovic**, National Bank of Serbia, Serbia and Montenegro

Over the past three decades emerging market (EM) economies have become increasingly integrated into world capital markets. We study time evolution of the integration of the emerging equity and currency markets and its implication for the optimal portfolio selection. Using a linear factor asset pricing framework we propose a time-varying measure of the global exposure of emerging markets - local (in time) R2 from the nonparametric regression on underlying factors. The measure explicitly takes into account time-variation in factor loadings and idiosyncratic volatilities and circumvents the pitfalls associated with the use of correlation coefficients in measuring asset return co-movements. The estimated exposure to global factors significantly varies between individual currency and equity markets and over time. We show that the exposures are priced risk factor in the EM equity and currency cross section. Optimal portfolio for CRRA investor loads positively on the low exposure assets and produces higher Sharpe ratios in comparison to passive and active portfolio strategies with currency risk hedging.

C073: **Measuring output gaps in real time by use of business tendency surveys**
Presenter: **Daniel Kaufmann**, Swiss National Bank, Switzerland
Co-authors: Rolf Scheufele

The information content of business tendency surveys is investigated to provide a reliable output gap measure in real time. To summarize the information of a large data set of sectoral business tendency surveys we extract a small number of common factors by a principle components estimator. The estimator is able to deal with mixed survey frequencies and missing observations at the beginning and end of the sample period. We show that these survey-based factors are highly correlated with standard output gap measures (HP filter or production function approach). There are three main advantages of our factor model. First, it is only marginally affected by data revisions which is extremely helpful in situations where the state of the economy is highly uncertain. Second, survey data are more timely available than quarterly variables that are required to calculate standard output gap measures. Third, our factor model provides a monthly output gap estimate. In a real-time out-of-sample exercise we show that a Phillips curve with our survey based output gap measure provides more precise inflation forecasts than standard output gap measures.

C102: **Forecasting FOMC target changes**
Presenter: **Michael Owyang**, Federal Reserve Bank of St Louis, United States
Co-authors: Travis Berge

Most interest rate rules are continuous functions of deviations of output from its potential and expected inflation from its target. In practice, central banks move the target rate in discrete increments and base their decisions on a wide-range of data. We estimate a dynamic ordered probit model of movements in the federal funds rate. In our model, monetary policy reacts to a large dataset that is summarized by a small set of dynamic factors. We then use the model for out-of-sample forecasting and evaluate these forecasts using methods unique to problems of classification.

C123: **Inference for ROC curves based on estimated predictive indices: A note on testing AUC = 0.5**
Presenter: **Robert Lieli**, Central European University and the National Bank of Hungary, Hungary
Co-authors: Yu-Chin Hsu

The purpose is to use the area under an ROC curve (AUC) to test the hypothesis that a predictive index combined with a range of cutoffs performs no better than pure chance in forecasting a binary outcome. We show that if the predictive index is an estimated, rather than fixed, function of the underlying covariates, then testing AUC = 1/2 based on standard asymptotic results (such as the limiting distribution of the Mann-Whitney U-statistics) can produce misleading results. The problem arises because first stage estimation overfits the model for the index in a way that artificially boosts the in-sample performance of the classifiers. Under these circumstances the bootstrap also fails to properly approximate the null distribution of the sample AUC.

C121: **Macroeconomic uncertainty indices**
Presenter: **Tatevik Sekhposyan**, Texas A&M University, Canada
Co-authors: Barbara Rossi

New indices to measure macroeconomic uncertainty are proposed. The indices measure how unexpected a realization of a representative macroeconomic key variable is relative to a forecast made a year before. The ex-ante likelihood of the realization is evaluated using density forecasts. We use density forecasts based on DSGE models, the Survey of Professional Forecasters, as well as densities pooled from a range of simple forecasting models. We further compare the new indices with those proposed in the literature.

CS46 Room I2 MACRO AND FORECASTING Chair: Fotis Papailias

C1145: **The time-frequency foundations of the Taylor rule**
Presenter: **Manuel Martins**, Universidade do Porto, Portugal
Co-authors: Luis Aguiar-Conraria, Soares Maria Joana

In 1993, John Taylor first proposed what became celebrated as the Taylor rule. The Taylor rule describes the Fed's monetary policy as a simple relation between the federal funds rate and the inflation rate and the output gap. The Taylor rule gained more importance recently, because there is a bill being discussed in congress, Federal Reserve Accountability and Transparency Act of 2014, that requires the Federal Reserve to provide Congress with a clear rule to describe the course of monetary policy. The text in the Bill precisely describes the classic Taylor rule formula. We provide the first assessment of the Taylor rule in the time-frequency domain. We rely on wavelet analysis, which seems particularly well-suited to study a subject such as this. On the one hand, the policy rule may not be constant over time. On the other hand, it is likely that the effects are different for different frequencies. We reach several important conclusions. Among several other conclusions, our results indicate that while monetary policy is effective as a tool to fight inflation, it is ineffective to fight recessions in economic activity.

C1100: **Smoothing macroeconomic and financial time series**
Presenter: **Dimitrios Thomakos**, University of Peloponnese, Greece

The problem of smoothing trending and non-stationary time series is considered. Two alternative methods for trend extraction, based on singular spectrum analysis and the discrete cosine transform are examined in terms of (a) their applicability and competitive performance (vis-a-vis popular

methods) and (b) methods for selecting the degree of smoothing. It is shown that both methods can easily perform on par or outperform existing methods in terms of the properties of the filters they use and the resulting residuals after smoothing. Empirical illustrations using simulated and real world data show how these methods can be applied in practice.

C1194: **Improved financial conditions indexes**

Presenter: **Fotis Papailias**, Queens University Belfast, United Kingdom

Co-authors: George Kapetanios, M. Marcellino

The performance of Financial Condition Indexes (FCIs) in forecasting key macroeconomic variables is investigated. A wide range of carefully selected financial indicators include Rates and Spreads, Stock Market Indicators and Macroeconomic Quantities. The results provide evidence that FCIs are particularly useful in forecasting GDP growth, Consumption growth, Industrial Production growth and the Unemployment Rate.

C1269: **The fundamental properties of time varying AR models with non stochastic coefficients**

Presenter: **Menelaos Karanasos**, Brunel University, United Kingdom

Co-authors: Alexandros Paraskevopoulos, Stavros Dafnos

The aim is to examine the problem of representing the dynamics of low order autoregressive (AR) models with time varying (TV) coefficients. The existing literature computes the forecasts of the series from a recursion relation. Instead, we provide the linearly independent solutions to TV-AR models. Our solution formulas enable us to derive the fundamental properties of these processes, and obtain explicit expressions for the optimal predictors. We illustrate our methodology and results with a few classic examples amenable to time varying treatment, e.g, periodic, cyclical, and AR models subject to multiple structural breaks.

C1008: **Local Bayesian estimation and forecasting with time-varying parameter DSGE models**

Presenter: **Katerina Petrova**, Queen Mary University London, United Kingdom

Co-authors: George Kapetanios, Liudas Giraitis, Ana Galvao

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 suitable for estimation of a DSGE model which can accommodate time-variation in all parameters of the model. There are two distinct advantages in allowing the structural parameters to vary over time. The first is that the method allows to address the question of constancy of these parameters and, in particular, whether there is evidence of regime changes (e.g. the Great Moderation or the appointment of Paul Volcker) or of misspecification in the design of DSGE models. To demonstrate the usefulness of the method to detect misspecification and structural change, the work presents an empirical application of the method to the Smets and Wouters model. The second advantage of the method is that, allowing the parameters to vary and conditioning on their most recent values when generating future projections, is expected to improve its forecasting record. To illustrate the time-varying model's forecasting performance in comparison to the fixed parameter DSGE model and other competing models, a pseudo out-of-sample forecasting assessment is presented.

Saturday 6.12.2014 16:45 - 18:00 Parallel Session F – ERCIM

ES03 Room I1 STATISTICAL MODELLING WITH R **Chair: Andreas Alfons**

E876: **Methods and tools for the simulation of synthetic populations**
Presenter: **Matthias Templ**, Vienna University of Technology, Austria
Co-authors: Bernhard Meindl

Synthetic population data are of use for teaching survey sampling methods, for comparing methods in design-based simulation studies and for microsimulation studies. Synthetic population data can be generated from tables, from survey samples, from population data and from a mix of macro and microdata. A methodological and computational framework for the simulation of synthetic population data is presented. We focus on methods that take care of generating not only individuals (agents) but also take care that these agents are grouped into some structure, typically households. The methods are efficiently and user-friendly implemented in the synthPop R package. The methods are finally applied to simulate the Austrian synthetic population on more than 20 variables using information from the European Statistics on Income and Living Conditions data as well as various tables from the Austrian Census. The results show that the model-fitting methods performs very well.

E912: **Sparse precision matrix estimation under elementwise contamination**
Presenter: **Viktoria Oellerer**, KU Leuven, Belgium
Co-authors: Christophe Croux

The dependency structure of multivariate data can be analyzed using the covariance matrix. In many fields the precision matrix, this is the inverse of the covariance matrix, is even more informative (e.g. gaussian graphical model, linear discriminant analysis). As the sample covariance estimator is singular in high-dimensions, it cannot be used to obtain a precision matrix estimator. In this scenario, the graphical lasso is one of the most popular estimators, but it lacks robustness. Most robust procedures assume that at least half of the observations are absolutely clean. However, often only a few variables of an observation are contaminated. An example is the high-dimensional independent contamination model, where small amounts of contamination lead to a large number of contaminated observations. Downweighting entire observations would then result in a loss of information, creating problems if the number of observations is small. We formally prove that initializing the graphical lasso with an elementwise robust covariance estimator leads to an elementwise robust, sparse precision matrix estimator computable in high-dimensions. Clever choice of the initial covariance estimator leads to a high breakdown point and positive definiteness of the final precision matrix estimator. The performance of the estimator is analyzed in a simulation study.

E968: **Robust sparse canonical correlation analysis**
Presenter: **Ines Wilms**, KU Leuven, Belgium
Co-authors: Christophe Croux

Canonical correlation analysis (CCA) is a multivariate method which describes the associations between two sets of variables. The objective is to find linear combinations of the variables in each data set having maximal correlation. A method for Robust Sparse CCA is discussed. Sparse estimation produces canonical vectors with some of their elements estimated as exactly zero. As such, their interpretability is improved. We also robustify the method so that it can cope with outliers in the data. Outliers are a common problem in applied data analysis. If the presence of outliers is ignored, then the estimation performance of standard estimation methods deteriorates drastically. To estimate the canonical vectors, we convert the CCA problem into an alternating regression framework. Sparse canonical vectors that are not attracted by outliers are obtained using the robust Sparse Least Trimmed Squares estimator. We illustrate the good performance of the Robust Sparse CCA method in several simulation studies and an empirical application.

ES13 Room B1 HIGH-DIMENSIONAL CAUSAL INFERENCE **Chair: Marloes Maathuis**

E275: **Efficient score-based estimation of causal models by greedy search**
Presenter: **Alain Hauser**, Bern University of Applied Sciences, Switzerland

Causal models based on directed acyclic graphs (DAGs) encode conditional independence relations among the involved random variables, but also predict effects of interventions, that is, perturbations of the system in which one or several random variables are forced to specific values. Data measured under different interventions ("interventional data") are ubiquitous in many fields of the life sciences, raising the demand for appropriate efficient estimation methods. Many score-based approaches (e.g., penalized maximum-likelihood estimation, or Bayesian inference) have been proposed for causal inference. Due to the large search space, greedy search is commonly used to make structure learning computationally feasible. Such score-based greedy search methods face two main challenges: specifying the traversal of the search space, and efficiently evaluating proposed moves. The first problem is relevant especially when causal models are only identifiable up to Markov equivalence (e.g., in the linear-Gaussian case, or in the case of unrestricted discrete causal models), and becomes more complex in the presence of interventional data. The second problem is crucial even when additional modeling assumptions make causal models fully identifiable. Solutions to both problems are addressed and their effectiveness is illustrated with simulation results.

E845: **How good is my graph estimate?**
Presenter: **Jonas Peters**, ETH Zurich, Switzerland
Co-authors: Peter Buhlmann, Joris Mooij

Causal discovery methods estimate a system's underlying causal graph from observational data. But even if the true structure is known, it is not clear how to evaluate the performance of this estimator. In many cases, we use the causal graph to infer the system's behavior under specific interventions, i.e. manipulations of some of the variables. Different graphs may then result in different predicted behaviors. The structural intervention distance (SID) quantifies these differences and thereby defines a (pre-)metric between graphs. In real systems we often do not know the underlying graph but have access to some data under interventional settings (e.g. knock-out experiments). We discuss how one could exploit these data for evaluating graph estimates and compare causal discovery techniques on real data sets.

E1052: **Estimation of directed acyclic graphs from partial orderings**
Presenter: **Ali Shojaie**, University of Washington, United States

Directed acyclic graphs (DAGs) are commonly used to model causal relationships among random variables. The problem of estimation of DAGs is both computationally and statistically challenging, and in general, the direction of edges may not be estimable from observational data alone. On the other hand, given a causal ordering of the variables, the problem can be solved efficiently, even in high dimensional settings. We consider an intermediate problem, where only a partial causal ordering of variables is available. We discuss a general estimation procedure for discovering DAGs with arbitrary structure from partial orderings. We then present an efficient penalized estimation framework for estimation of high dimensional, sparse directed acyclic graphs.

ES33 Room L1 STATISTICAL METHODS IN HIGH DIMENSIONS **Chair: Francesco Giordano**

E674: **An extremes-based double clustering procedure for financial returns**
Presenter: **Giovanni De Luca**, University of Naples Parthenope, Italy
Co-authors: Paola Zuccolotto

The recent financial crisis has renewed the interest of researchers in developing new approaches for portfolio selection, in particular replacing the traditional linear correlation coefficient as a measure of the association between the returns of two assets. A recent stream of literature has shown that the use of measures of association between extremely low or extremely high returns of two assets can provide a useful insight for portfolio selection. We propose a method for selecting the portfolio that maximizes the risk of joint positive extreme returns among all the possible portfolios that minimize the risk of joint negative extreme returns. The method is implemented after a double clustering procedure. First, we cluster the assets into groups characterized by high lower tail dependence, secondly we cluster the same assets into groups with high upper tail dependence. The choice of the portfolio is carried out into two steps. In the first step we consider all the possible portfolios selecting one asset from each of the lower tail dependence-based groups. Then, we choose that portfolio that minimizes a heterogeneity index according to the position of the assets in the upper tail dependence-based clusters. We show that such a criterion ensures the selection of portfolios with a better performance with respect to the traditional methods. Moreover, it can work also with high dimensional data.

E563: **Relevant covariates in high dimensional regression: the case of business failure**
Presenter: **Marialuisa Restaino**, University of Salerno, Italy
Co-authors: Alessandra Amendola, Francesco Giordano, Maria Lucia Parrella

In the context of high dimensional regression LASSO and its variants it is suitable to deal with the variable selection. When we have a binary variable (for example bankrupted vs. non-bankrupted) many approaches have been explored, such as GLM-LASSO, in order to identify the relevant covariates. We propose a new approach that can be organized in two steps. First, we identify some continuous covariates as proxies for the binary dependent variable. Second, we apply a nonparametric method to make variable and model selection in order to have the relevant covariates and the type of "relation" (linear and nonlinear). We evaluate the performance of the proposed method on a sample of firms that have experienced the bankruptcy and for which a set of financial ratios has been observed.

E633: **A multiple testing procedure in high dimensional nonparametric regression based on empirical likelihood**
Presenter: **Francesco Giordano**, University of Salerno, Italy
Co-authors: Soumendra Nath Lahiri, Maria Lucia Parrella

In the context of nonparametric regression, variable selection is usually done through some multiple testing procedure. We propose to use one based on the Empirical Likelihood (EL) technique and a variant of the local polynomial estimator. The main advantage of this choice is that we do not need to estimate the nuisance parameter, the variance of the error process, which would be difficult in a multivariate high dimensional context. This represents a big improvement over the usual nonparametric procedures available in the literature. Another advantage is that we can relax the assumption of Gaussianity for the error process. A peculiarity of our proposal which deserves attention is the particular implementation of the empirical likelihood technique. There are two innovative aspects compared with the other implementations of EL to the nonparametric kernel estimators. Firstly, it is known that the use of the EL for the analysis of the kernel based estimators is affected by the bias problem, so that a correction is necessary and usually performed through the undersmoothing technique. In our procedure, this problem is avoided because we use a variant of the local polynomial estimator which is unbiased under the null hypothesis. Secondly, the analysis of the asymptotic statistical properties of the EL procedure is done in a novel framework, where the bandwidths are fixed and not tending to zero as the sample size grows. This makes our analysis non-standard and the EL estimator more efficient.

ES36 Room M1 STATISTICS IN FUNCTIONAL AND HILBERT SPACES Chair: Gil Gonzalez-Rodriguez

E346: **A clustering method for functional data**
Presenter: **Enea Bongiorno**, Universita degli Studi di Milano, Italy
Co-authors: Aldo Goia

An unsupervised classification algorithm for curves is proposed. It extends a density based multivariate cluster approach to the functional framework. In particular, the modes of the small-ball probability are used as starting points to build the clusters. A simulation study and an application to real data are proposed.

E396: **Efficient recursive estimation of the geometric median in Hilbert spaces: new results**
Presenter: **Antoine Godichon**, Universite de Bourgogne, France

With the development of automatic sensors, it is more and more usual to study large samples of observations taking values in high dimensional spaces such as functional spaces. In this framework, the geometric median, which is a generalization of the real median for metric spaces, is an interesting robust location parameter. Two recursive estimators will be considered. Some results were given on the almost sure convergence of the stochastic gradient estimator and on the asymptotic normality of the averaged version. First, some new results on the L^2 rates of convergence of the algorithms are given. Finally, using some exponential inequalities for the martingale differences, non asymptotic confidence intervals will be given.

E886: **PLS regression for multivariate functional data**
Presenter: **Cristian Preda**, University of Lille, France
Co-authors: Gilbert Saporta

PLS methodology is considered for regression with multivariate functional predictor. The basis expansion approximation of the multivariate curves provides the PLS approximation of the linear regression function. A simulation study illustrates the method.

ES38 Room E1 INFERENCE BASED ON THE LORENZ AND GINI INDEX OF INEQUALITY Chair: Francesca Greselin

E330: **Gini and Lorenz in action: measuring the lack of increasingness in functions with applications in education and beyond**
Presenter: **Ricardas Zitikis**, University of Western Ontario, Canada
Co-authors: Danang Qoyyimi

In a variety of applications, solving problems frequently relies on the assumption that certain functions are increasing, which needs to be verified given data. We suggest and discuss an index for measuring the lack of increasingness in functions: the index takes on the value zero whenever functions are increasing and on positive values otherwise. We illustrate the utility of the index when assessing student performance in a variety of courses, which also clarify how the index can be used for optimizing investment portfolios and tackling other problems. The earlier literature on the topic is rather scarce: we give an overview of it together with an explanation of how the topic falls under the Gini-Lorenz framework.

E543: **Decomposition of family incomes by income sources, geographical areas and the issue of negative income values**
Presenter: **Achille Vernizzi**, University of Milan, Italy
Co-authors: Maria Giovanna Monti, Emanuela Raffinetti, Elena Siletti

Negative income often appears as an unfamiliar concept, however its presence in real surveys may lead to troubles in applying the classical inequality measures. We propose to extend the approach for the Gini index decomposition by including even incomes with negative values. The problem of negative income is here solved through a suitable new normalization ensuring that the adopted inequality indices lie inside their standard ranges. The introduced decomposition is applied to Household Income and Wealth survey of the Bank of Italy. Specifically, we first consider family income distributions and decompose them by income sources. The total net income was given as the sum of six sources: income from employment,

self-employment, pensions, transfers, capital gain and financial capital gain. Secondly, we analyse the effects of income sources on the overall inequality within and across the three main Italian geographical areas (North Italy, Central Italy, South Italy and Islands). Finally, investigations about the empirical distribution of the involved inequality indices and the related inferential issues are led thorough simulation studies based on re-sampling techniques.

E825: **Assessment and comparison of survival models using concentration indices**
Presenter: **Chiara Gigliarano**, Department of Economics and Social Sciences, Italy
Co-authors: Silvia Figini, Pietro Muliere

Predictive accuracy of survival models is usually assessed using Harrell's C index and its extensions. However, these are measures of concordance between model predictions and status indicator, based only on ranking and not on the differences in survival times. We introduce a novel approach to assess the discriminative ability of survival models taking into account differences in survival times based on a restricted version of the Gini concentration index. Equality in the concentration of predicted failure times between groups may reveal a weak discriminatory power, while differences in concentration between groups may suggest that the survival model has high predictive accuracy. Application on real data are also provided, showing the novel contribution of the methods proposed.

ES53 Room Q1 BAYESIAN SEMI AND NONPARAMETRIC ESTIMATION IN COPULA MODELS Chair: Brunero Liseo

E438: **Scalable Bayesian model averaging through local information propagation**
Presenter: **Li Ma**, Duke University, United States

The aim is to show that a probabilistic version of the classical forward-stepwise variable inclusion procedure can serve as a general data-augmentation scheme for model space distributions in (generalized) linear models. This latent variable representation takes the form of a Markov process, thereby allowing information propagation algorithms to be applied for sampling from model space posteriors. In particular, we propose a sequential Monte Carlo method for achieving effective unbiased Bayesian model averaging in high-dimensional problems, utilizing proposal distributions constructed using local information propagation. We illustrate our method - called LIPS for local information propagation based sampling - through real and simulated examples with dimensionality ranging from 15 to 1,000, and compare its performance in estimating posterior inclusion probabilities and in out-of-sample prediction to those of several other methods - namely, MCMC, BAS, iBMA, and LASSO. In addition, we show that the latent variable representation can also serve as a modeling tool for specifying model space priors that account for knowledge regarding model complexity and conditional inclusion relationships.

E379: **Bayesian model selection of high-dimensional vine copulas**
Presenter: **Luciana Dalla Valle**, University of Plymouth, United Kingdom
Co-authors: Claudia Czado

Vine copulas are multivariate models for highly dependent variables, constructed using bivariate pair copulas as building blocks. Vines have been applied to several high-dimensional problems due to their extreme flexibility, that allows the specification of various types of non-linear dependencies among variables. However, for high-dimensional datasets, the complexity of the vine structure becomes huge and inference and model selection are more problematic. We propose a novel approach to estimate the parameters of high-dimensional vines and, at the same time to select the best vine model inducing sparsity in the vine structure. Inference is conducted under the Bayesian framework, using the Reversible Jump Markov Chain Monte Carlo algorithm.

E597: **Adaptive importance sampling methods for the multivariate skew-Student distribution and skew-t copula**
Presenter: **Antonio Parisi**, University of Rome Tor Vergata, Italy
Co-authors: Brunero Liseo

A multivariate skew-t model is described under a Bayesian perspective, with a particular emphasis on prior elicitation and posterior sampling: a set of weakly informative priors on the model parameters is proposed and, since standard MCMC methods can hardly explore the whole posterior density, a Monte Carlo sampler is tailored for this model. Simulation results are provided for model selection using multivariate (skew)-normal and (skew)-t models. A constraint between the shape and the scale parameters makes difficult the extension of the model to dimensions higher than two. A skew-t copula model is employed to overcome this problem: using a vine copula representation, it is possible to decompose densities of arbitrary dimensions in the product of marginals and bivariate conditional densities.

ES58 Room D1 HEURISTIC OPTIMIZATION AND ECONOMIC MODELLING Chair: Peter Winker

E702: **No such thing like perfect hammer: comparing different objective function specifications for optimal control**
Presenter: **Dmitri Blueschke**, University of Klagenfurt, Austria
Co-authors: Ivan Savin

Policy makers constantly face optimal control problems: what controls allow us to achieve certain targets in, e.g., GDP growth or inflation? Conventionally, this is done by applying certain linear-quadratic optimization algorithms to dynamic econometric models. Several algorithms extend this baseline framework to nonlinear stochastic problems. However, those algorithms are limited in a variety of ways including, most importantly, restriction to local best solutions only and the inflexibility of quadratic objective functions. We run an extensive analysis of the objective functions for optimal control framework by means of classical and heuristic optimization. Furthermore, we introduce several new objective function forms for optimal control problems, which are aimed at being useful additions to the standard least mean of squares approach and allow for more robust and context-based optimal control solutions.

E934: **Simulating particle swarm optimization for estimating likelihood function of ARMA(1,1) model**
Presenter: **Basad Al-sarray**, Franche Comte University, France

Estimation of likelihood function parameters of an ARMA(1,1) model is considered by using Particle Swarm Optimization (PSO). It depends on two operators: population size and velocities. Results depending on designing the experiments to simulate PSO algorithm for optimizing likelihood function by setting different values of sample size, model's parameters and PSO operators are provided.

E973: **A heuristic for completing covariance and correlation matrices**
Presenter: **Martin van der Schans**, Ortec-Finance, Netherlands

A heuristic is presented for completing partially specified covariance (and correlation) matrices. In many statistical and econometric models, covariance matrices play a prominent role and are part of the model specification. In practical applications, however, commonly not all covariances are known due to, for example, data limitations. As an example, consider a case in which private equity returns (for which data is limited) are modeled though specifying a distribution conditional on the forecast of other asset classes. In such cases, a partially specified covariance matrix can arise which must be completed in order to use the model at hand. The already specified covariances generically imply a dependence between variables amongst which a covariance is unspecified. A criterion that can then be used to complete the covariance matrix is to introduce as little extra dependence between variables as possible which is equivalent to entropy maximization. Such completion problems can be solved with global optimization algorithms which are, unfortunately, too slow and memory consuming for large matrices. We present a heuristic for completing partially specified covariance (and correlation) matrices that is: fast (also for large matrices), approximates the maximum entropy criterion and also corrects for potential inconsistencies in the initially specified covariances.

ES78 Room G1 DIRECTED AND UNDIRECTED GRAPH MODELS Chair: Annie Qu

E467: **Second-order inference for high-dimensional time series**
Presenter: **Xiaohui Chen**, University of Illinois at Urbana-Champaign, United States
Co-authors: Mengyu Xu, Wei Biao Wu

In the "big data" era, high-dimensional datasets have been increasingly seen in a broad spectrum of research fields such as in neuroscience and bioinformatics. Current development of high-dimensional data analysis in statistics and machine learning communities primarily focuses on i.i.d. observations with sub-Gaussian tails. We discuss some second-order structured statistical estimation and inference problems for high-dimensional time series data: (i) covariance matrix and its functionals; (ii) time-varying graphs and change-point detection. Real applications and examples such as classification of task-related fMRI signals are provided.

E870: **Classification with unstructured predictors with an application to sentiment analysis**
Presenter: **Annie Qu**, University of Illinois at Urbana-Champaign, United States
Co-authors: Junhui Wang, Xiaotong Shen, Yiwen Sun

Unstructured data refers to information that lacks certain structures and cannot be organized in a predefined fashion. Unstructured data often involve words, texts, graphs, objects or multimedia types of files that are difficult to process and analyze with traditional computational tools and statistical methods. Ordinal classification for unstructured predictors with ordered class categories are explored, where imprecise information concerning strengths between predictors is available for predicting class labels. However, imprecise information is expressed in terms of a directed graph, with each node representing a predictor and a directed edge containing pairwise strengths between two nodes. One of the targeted applications for unstructured data arises from sentiment analysis, which identifies and extracts the relevant content or opinion of a document concerning a specific event of interest. Statistically, we integrate the imprecise predictor relations into linear relational constraints over classification function coefficients, where large margin ordinal classifiers are introduced, subject to many quadratically linear constraints. Moreover, the proposed classifiers are applied in sentiment analysis using binary word predictors. Computationally, we implement ordinal support vector machines and ψ-learning through a scalable quadratic programming package based on sparse word representations. Theoretically, we show that utilizing relationships among unstructured predictors improves prediction accuracy of classification significantly. Finally, we illustrate an application for sentiment analysis using consumer text reviews data from tripadvisor.com.

E967: **Regression trees for longitudinal data (LongCART) and their application in the biomarker neuroimaging study**
Presenter: **Jaroslaw Harezlak**, Indiana University Fairbanks School of Public Health, United States
Co-authors: Madan Kundu

Mixed model methodology has proven to be extremely useful in longitudinal data settings. Usually, we assume that a mean longitudinal model conditional on baseline covariates is applicable to the entire population. However in a heterogeneous population the variation in the longitudinal trajectories can be hard to explain. This is usually the case in observational studies with many possible predictors. In such cases, a group-averaged trajectory could mask important subgroup differences. Our aim is to identify and characterize subgroups based on the combination of baseline covariates with differential longitudinal behavior. We extend the CART methodology to identify such subgroups via evaluation of a goodness of fit criterion at all possible splits of partitioning variables. In such an approach, we encounter the problem of multiple testing. We ameliorate this problem by performing a single test identifying the parameter stability of a longitudinal model at all values of a partitioning variable. We propose a tree construction algorithm and obtain asymptotic results for the instability tests. Simulation study is used to evaluate the numerical properties of our LongCART method. Finally, we apply the LongCART to study the changes in the brain metabolite levels of chronically infected HIV patients.

ES97 Room P1 CHANGE-POINTS IN TIME SERIES I Chair: Daniel Vogel

E520: **Robust change-point tests for time series**
Presenter: **Herold Dehling**, Ruhr-University Bochum, Germany
Co-authors: Martin Wendler, Roland Fried

Recent developments in the area of robust change-point tests for time series are presented. Given the data $X_1,\ldots,X_n$, we test the hypothesis of stationarity against the alternative of a level shift at an unknown point in time. Our tests are based on common two-sample tests, such as the Wilcoxon test and the Hodges-Lehmann test. Specifically, we investigate the test statistics $\max_{1\le k\le n-1}\sum_{i=1}^{k}\sum_{j=k+1}^{n}1_{\{X_i\le X_j\}}$, and $\max_{1\le k\le n-1}\text{median}\{(X_j-X_i):1\le i\le k<j\le n\}$. We derive the asymptotic distribution of these test statistics, and more generally of two-sample empirical U-processes and U-quantiles of dependent data, both under the null hypothesis as well as under local alternatives.

E489: **Structural changes in time series of counts**
Presenter: **Marie Huskova**, Charles University, Czech Republic
Co-authors: Sarka Hudecova, Simos Meintanis

Sequential procedures for detecting structural changes in time series with discrete observations are developed and studied. Special emphasis is given to the models of integer autoregression (INAR) and Poisson autoregression (PAR). The statistics utilize the empirical probability generating functions. For both models structural changes in distribution are considered, but also comments for the classical problem of parameter change are included. The asymptotic properties of the proposed test procedures are studied under the null hypothesis as well as under alternatives. A Monte Carlo power study on bootstrap versions of the new methods is also included.

E617: **Piecewise quantile autoregressive modeling for non-stationary time series**
Presenter: **Alexander Aue**, University of California, Davis, USA
Co-authors: Rex Cheung, Thomas Lee, Ming Zhong

The purpose is to present a new methodology for the fitting of non-stationary time series that exhibit nonlinearity, asymmetry, local persistence and changes in location, scale and shape of the underlying distribution. In order to achieve this goal, we perform model selection in the class of piecewise stationary quantile autoregressive processes. The best model is defined in terms of minimizing a minimum description length criterion derived from an asymmetric Laplace likelihood. Its practical minimization is done with the use of genetic algorithms. If the data generating process follows indeed a piecewise quantile autoregression structure, we show that our method is consistent for estimating the break points and the autoregressive parameters. Empirical work suggests that the proposed method performs well in finite samples.

ES103 Room N1 TAIL DEPENDENCE AND MARGINALS WITH HEAVY TAILS Chair: Fabio Spizzichino

E745: **Portfolio optimisation under switching dependence**
Presenter: **Mauro Bernardi**, Sapienza University of Rome, Italy

A portfolio optimisation problem is considered where the asset returns follow a Hidden Markov process and the component density is chosen to be the multivariate Laplace distribution to account for asymmetric dependence structures and extreme events often observed in international equity markets. In this context, we provide and compare different risk measures, generalising Markowitz mean-variance portfolio theory. Application on a set of European Sector indexes is considered to show that the right specification of the conditional distribution of the Markov process as well as the

choice of the risk measure in the portfolio optimisation process result in optimal portfolios with better characteristics in terms of their risk-return profile.

E834: **Singular copulas and tail dependence**
Presenter: **Fabrizio Durante**, Free University of Bozen-Bolzano, Italy
In many problems of risk aggregation the specification of the dependence structure in the tails is one of the major issues.We discuss different tail specifications for copulas having a singular component. Specifically, two cases will be considered. First, we present a multivariate generalization of copulas whose support is contained in a hairpin set, which is, in two dimensions, the union of the graphs of two increasing homeomorphisms. As we will see, these constructions are related to the determination of the Frechet class of random vectors with given partial information about some of the order statistics. Second, we provide existence results for copulas with a prescribed singular component by means of the Markov kernel representation. Such a construction has a natural interpretation in multivariate stochastic systems of lifetimes, since it models situations when joint defaults can occur with a non-zero probability.

E807: **Large deviations of the interference in wireless communication models**
Presenter: **Giovanni Luca Torrisi**, CNR, Italy
Interference from other users limits the capacity and the connectivity of wireless networks. Simple models of wireless ad hoc networks in which node locations are described by various classes of point processes and signals are random will be described. Large deviations principles for the interference will be presented under light tail and heavy tail conditions on the signals.

ES106 Room A1 ROBUST CLUSTERING WITH MIXTURES Chair: Antonio Punzo

E843: **Robust mixture modelling using skewed multivariate distributions with variable amounts of tailweight**
Presenter: **Florence Forbes**, INRIA, France
Co-authors: Darren Wraith
The family of location and scale mixtures of Gaussians has the ability to generate a number of flexible distributional forms. It nests as particular cases several important asymmetric distributions like the Generalised Hyperbolic distribution. The Generalised Hyperbolic distribution in turn nests many other well-known distributions such as the Normal Inverse Gaussian (NIG) whose practical relevance has been widely documented in the literature. In a multivariate setting, we propose to extend the standard location and scale mixture concept into a so called multiple scaled framework which has the advantage of allowing different tail and skewness behaviours in each dimension of the variable space with arbitrary correlation between dimensions. Estimation of the parameters is provided via an EM algorithm with a particular focus on NIG distributions. Inference is then extended to cover the case of mixtures of such multiple scaled distributions for application to clustering. Assessments on simulated and real data confirm the gain in degrees of freedom and flexibility in modelling data of varying tail behaviour and directional shape.

E1213: **Robust latent class analysis through outlier detection and modelling**
Presenter: **Brian Francis**, Lancaster University, United Kingdom
Co-authors: Fulvia Pennoni, Silvia Pandolfi, Francesco Bartolucci
Outlier detection in mixture models where there may be multiple outliers contaminating the main data is considered. Methods in the literature for continuous data have grown and now include fitting mixtures of t-distributions rather than normal distributions or robustifying through modifying the likelihood in various ways. The focus in contrast is on discrete data and considers the approach of obtaining better estimates for the main underlying structure in latent class models in large data sets by detecting them and fitting them as separate mixture or latent class components. We propose a method that uses latent class outliers as seeds for new groups, and we propose a new algorithm for fitting latent class models where some of the latent classes are small. We show that often such latent class solutions are better than those obtained from random start points.

E943: **Estimation of Gaussian mixture models via mixted nuclear/ℓ_∞/ℓ_1-norm penalization**
Presenter: **Stephane Chretien**, Universite de Franche Comte, France
The problem of unsupervised classification of high dimensional data, e.g. images, time series, is considered. Such problems have been attracting much interest in the mathematical learning community for a long time. In the non-parametric setting, several techniques such as k-means algorithms, have been used extensively and active research is devoted to improving the performance and analysis of such methods. In the parametric setting, mixture models are also widely used. One of the most famous algorithms for maximizing the likelihood is EM and its multiple variants. Despite the good practical performances of all the existing methods, either parametric or non-parametric, most of them share the same embarrassing drawback of being based on minimizing a nonconvex criterion and convergence to a global optimum cannot be guaranteed in most applications. The goal is to introduce a fast, robust and efficient methodology, based on convex minimization, for performing non-parametric classification. Our approach is based on a simple low rank matrix model for the data sample. We use a mixture of convex norms surrogates for the sparsity and the rank recently developed for the Variable Selection (ℓ_1-norm) and the Matrix Completion (nuclear norm) problems. A detailed theoretical analysis will be presented as well as promissing simulation results.

ES116 Room H1 INTELLIGENT DATA ANALYSIS FOR HIGH DIMENSIONAL SYSTEMS Chair: Debora Slanzi

E643: **Screening designs for factors with many levels in drug discovery**
Presenter: **Philip Brown**, University of Kent, United Kingdom
Co-authors: Martin Ridout
Main effects designs for a few factors but at very many levels as arise in drug discovery are considered. With just two factors, as is the case for two chemical substitution positions, two types of design are investigated, termed *sawtooth* and *dumbbell*. For three factors, generalised cyclic sawtooth designs are considered. The designs are compared using various optimality criteria and on a published real pharmaceutical dataset. Links to microarray loop designs are explored and robustness to missing data quantified.

E689: **Intelligent control system model-based optimisation for energy saving**
Presenter: **Matteo Borrotti**, European Centre for Living Technology, Italy
Co-authors: Luca Sartore, Debora Slanzi
A novel approach involving linear/non-linear statistical models and multi-objective optimisation techniques is proposed for the reduction of energy consumption in office buildings. With the aim of simultaneously achieve optimal users' comfort levels and containment of related cost, linear stochastic time series models and nonlinear feed forward neural network models are considered together with Pareto Front optimisation for developing efficient control strategy in building automation systems. The approach is evaluated using simulations and numerical studies.

E703: **A multi-objective procedure for designing optimal static daylighting devices**
Presenter: **Davide De March**, Ca Foscari University of Venice, Italy
Co-authors: Irene Poli
In accordance with EU building efficiency directives, many efforts have to be posed on transforming existing residential buildings into nearly zero-energy buildings. Several energy improvements might be accomplished by using traditional approaches to optimise single components of the building, but the complexity and the high dimensionality of these systems impose new methodological approaches. We develop a Multi-objective Particle Swarm Optimisation (Mo-PSO) for static daylighting devices aiming to optimise simultaneously the thermal performance of a

residential building by reducing its energy demand and improving thermal comfort. We present the Mo-PSO, based on four variables Pareto front, for optimising the external shading devices of a residential building in Madrid, where a proper shading and control design can significantly increase the thermal performance and the user's comfort.

ES125 Room F1 MISSING COVARIATE INFORMATION IN SURVIVAL ANALYSIS Chair: Thomas Scheike

E062: **How to deal with missing covariate data in survival analysis**
Presenter: **Torben Martinussen**, University of Copenhagen, Denmark
Co-authors: Thomas Scheike, Klaus Holst

Missing covariate data often occurs in practice when analyzing survival data. Multiple imputation has been suggested as one way of dealing with this problem. When doing multiple imputation it is important to also use the response variable in the imputation algorithm, but for survival data it is not clear what the response really is when the survival time is censored. In this talk I will take another route based on the observed hazard function derived from assuming the Cox model for the full data setting. This leads to some recursive estimating functions for the unknown parameters that can be used for estimation. In the case where the observed covariates are discrete one can derive an estimator of the target parameters without having to model the covariate distribution, which is otherwise needed for existing methods. Large sample results are given and the method is applied to a real data example.

E314: **A comparison of multiple imputation via chained equations and general location model for AFT models with missing covariates**
Presenter: **Lihong Qi**, University of California Davis, United States
Co-authors: Yulei He, Rongqi Chen, Ying-Fang Wang, Xiaowei Yang

Missing covariates are common in biomedical studies with survival outcomes. Multiple imputation is a practical strategy for handling this problem with various approaches and software packages available. We compare two important approaches: multiple imputation by chained equation (MICE) and multiple imputation via a general location model (GLM) for accelerated failure time (AFT) models with missing covariates. Through a comprehensive simulation study, we investigate the performance of the two approaches and their robustness toward violation of the GLM assumptions and model misspecifications including misspecifications of the covariance structure and of the joint distribution of continuous covariates. Simulation results show that MICE can be sensitive to model misspecifications and may generate biased results with inflated standard errors while GLM can still yield estimates with reasonable biases and coverages in these situations. MICE is flexible to use but lack of a clear theoretical rationale and suffers from potential incompatibility of the conditional regression models used in imputation. In contrast, GLM is theoretically sound and can be rather robust toward model misspecifications and violations of GLM assumptions. Therefore, we believe that GLM shows the potential for being a competitive and attractive tool for tackling the analysis of AFT models with missing covariates.

E339: **Competing risks regression with missing data in the prognostic factors**
Presenter: **Federico Ambrogi**, University of Milan, Italy

For the medical studies involving competing risks, one often wishes to estimate and model the cumulative incidence probability, the marginal probability of failure for a specific cause. Recently, several new methods have been developed to directly model the cumulative incidence probability of a specific cause of failure. The key issue here is how to deal with incomplete data due to the fact that observations are subject to right-censoring. We refer to a simple problem in which one covariate, say Z, is always observed and the other, say X, is sometimes missing. There has been considerable focus on handling missing covariates and there are several suggestions for dealing with the simpler survival data where there are not several causes of death. For survival data the key suggestions are multiple imputation techniques that typically aim for the modelling of the hazard function. An alternative is the IPCW techniques for survival data. Even though the competing risks framework is very common practice, there are no studies dealing with the problem of missing covariate information in competing risks regression. We present some results regarding multiple imputation and IPCW techniques applied to the direct binomial regression model through some simple simulations.

ES127 Room C1 BIOSTATISTICS, EPIDEMIOLOGY AND TWIN STUDIES Chair: Ayse Ulgen

E302: **Event and event-free concordance in twin studies**
Presenter: **Soeren Moeller**, University of Southern Denmark, Denmark
Co-authors: Thomas Scheike, Klaus Holst, Jacob B. Hjelmborg

The risk of an event in a twin given the occurrence of the event in the co-twin, the casewise concordance, is the measure reported in a large number of twin studies of dichotomous traits. We review the classical concordance model and extend it to the cases in which multiple discrete outcomes and events over time are present. The casewise concordance is suitable for studying how within pair dependence might change over time and we present a novel method that allows for estimating the risk of an event in a twin up to a certain time given that the co-twin experienced the event before that time. Furthermore, we investigate a setting in which the outcome of interest is the absence of any event until some time point, and in which the casewise concordance describes the probability of one twin not having had an event yet, given that the co-twin did not experience an event until now. To take appropriate account for competing risks we have to switch from the risk scale to the hazard scale, on which we can estimate the event-free casewise concordance.

E1270: **gammaMAXT: a fast multiple-testing correction algorithm**
Presenter: **Francois Van Lishout**, University of Liege, Belgium
Co-authors: Francesco Gadaleta, Jason H. Moore, Louis Wehenkel, Kristel Van Steen

The purpose of the maxT algorithm is to control the family-wise error rate (FWER) when assessing significance of multiple tests jointly. However, the requirements in terms of computing time and memory of this procedure are proportional to the number of investigated hypothesis. The memory issue has been solved by Van Lishout's implementation of maxT, which makes the memory usage independent from the size of the dataset. This algorithm is implemented in *MBMDR-3.0.3*, a software that is able to identify genetic interactions, for a variety of SNP-SNP based epistasis models in an effective way. However, that implementation turned out to be less suitable for genome-wide interaction analysis studies, due to the prohibitive computational burden. We present *gammaMAXT*, a novel algorithm which is part of *MBMDR-4.2.2*. We show that, in the abscence of interaction effects, test-statistics produced by the MB-MDR methodology follow a mixture distribution with a point mass at zero and a shifted gamma distribution for the top 10% of the strictly positive values. We show that the *gammaMAXT* algorithm has a power comparable to maxT and maintains FWER, but requires less computational resources and time. *MBMDR-4.2.2* can be downloaded at http://www.statgen.ulg.ac.be.

E1272: **High quality seawater in Cyprus: a fecal contamination survey**
Presenter: **Ayse Ulgen**, University of Southern Denmark, Denmark
Co-authors: Leon Cantas, Evelthon Iacovides, Frank van der Meulen

The fecal contamination level in seawater of Cyprus is determined. Traditional bacteria culturing methods were used to determine the colony forming units. Log of the total concentrations of bacteria according to regions and taking seasonal variations into account have been analyzed. Microbiological seawater quality was found to be excellent according to the European Union standards that need to be conserved at that level for an optimal environment and public health. More studies are needed to investigate the micro-diversity and composition of coastal seawater flora of the whole of Cyprus.

ES153 Room O1 MULTIVARIATE STATISTICS I **Chair: Christian Hennig**

E1127: **Estimation of factor correlation in penalized likelihood factor analysis**

Presenter: **Kei Hirose**, Osaka University, Japan

Co-authors: Michio Yamamoto

The problem of sparse estimation via a lasso-type penalized likelihood procedure in a factor analysis model is considered. Typically, the model estimation is done under the assumption that the common factors are orthogonal (uncorrelated). However, the lasso-type penalization method based on the orthogonal model can often estimate a completely different model from that with the true factor structure when the common factors are correlated. In order to overcome this problem, we propose to incorporate a factor correlation into the model, and estimate the factor correlation along with parameters included in the orthogonal model by maximum penalized likelihood procedure. The proposed method can provide sufficiently sparse solutions, and be applied to the data where the number of variables is larger than the number of observations. Monte Carlo simulations are conducted to investigate the effectiveness of our modeling strategies. The usefulness of the proposed procedure is also illustrated through the analysis of real data.

E1132: **Global optimization of squared distance multidimensional scaling through the nuclear norm penalty**

Presenter: **Patrick Groenen**, Erasmus University Rotterdam, Netherlands

Multidimensional scaling (MDS) is a standard tool to visualize dissimilarities between pairs of objects by Euclidean distances between points in a low dimensional space. Often the dimensionality is low (say up to 3) to allow for an easy visualization. However, it is known that imposing such low rank constraints on the configuration can yield local minima. We investigate the use of the nuclear norm penalty in combination with squared distance MDS using the S-Stress loss function. It can be proved that adding the nuclear norm penalty reduces the rank of the configuration while the problem remains convex over a convex constraint and, therefore, has a minimum that is global. The usefulness of this approach is studied in particular in three cases: (a) regular MDS, (b) sparse and structured dissimilarities, (c) approximations of least-squares MDS.

E116: **Common principal components estimation: A dynamical system approach**

Presenter: **Nickolay Trendafilov**, Open University, United Kingdom

The common principal components (CPC) model is a generalization of the standard principal component analysis (PCA) for several covariance matrices. The classic CPC algorithm first finds the common components, and then, the individual variances are calculated for each of the covariance matrices. We adopt a dynamical system approach to present the maximum likelihood estimation of the CPCs and its least squares counterpart in a unified way. This facilitates the comparison of the two types of estimators through their optimality conditions. Additionally, this approach makes it possible to propose a new procedure for calculating the individual eigenvalues (variances) for each covariance matrix without finding the CPCs. Once the individual eigenvalues are found, the CPCs can be readily obtained by solving Procrustes problems. Numerical illustrations on well-known data are considered.

Sunday 7.12.2014 08:45 - 10:25 Parallel Session F – CFE

CS11 Room A2 BAYESIAN NONLINEAR ECONOMETRICS Chair: Roberto Casarin

C313: **Combining survey and Bayesian VAR forecasts of US macro variables: Evidence from entropic tilting**
Presenter: **Francesco Ravazzolo**, Norges Bank, Norway
Co-authors: Fabian Krueger, Todd Clark

Macroeconomic expert surveys often provide useful information beyond that contained in econometric models. However, no standard approach has yet emerged in order to combine these two distinct sources of information. Existing approaches include combination, interpolation methods, as well as using surveys in the process of estimating an econometric model. The aim is to explore entropic tilting in order to exploit various types of information contained in the Survey of Professional Forecasters (SPF). We consider baseline forecasts from various Bayesian Vector Autoregressive (BVAR) specifications with stochastic volatility and tilt them to survey mean forecasts, forecast variances as well as projected probabilities of negative GDP growth. Also, we analyze important implementation choices, such as whether to apply tilting to a univariate forecast distribution or to a multivariate distribution for several variables or forecast horizons. We find that all three types of survey information (means, variances and probabilities) can improve the point and density forecasts of the BVARs. Second, the gains are particularly large at short horizons. Third, as a means of incorporating survey information into time series models, entropic tilting performs similar to much more complex strategies which entail reparametrizing the model to match the survey information.

C728: **Multiple bi-dimensional SV models for volatility matrix estimation. The case of 5D-italian banks' return**
Presenter: **Andrea Pierini**, Roma Tre University, Italy
Co-authors: Roberto Casarin, Alessia Naccarato

The use of bidimensional Stochastic Volatility models is applied to a multivariate time series of 5 banks' return among the blue chips, in order to construct an estimation of their variance-covariance matrix. As the dimension of the problem is usually large, it is infeasible to tackle this matrix estimation by a direct application of a multivariate SV model or any other method of dimension n. By considering only bi-dimensional SV models, a solution is proposed which is feasible while considering the more natural dynamics between the variables. The financial dataset require that the variance-covariance estimation is time varying as well as takes into consideration the outliers. In this sense the SV model are able to represent the needed characteristics. However, in order to obtain the estimation, the MCMC method is needed, in that the flexibility of the SV brings with itself a computational burden.

C729: **A Bayesian test for multimodality with applications to DNA and economic data**
Presenter: **Nalan Basturk**, Erasmus University Rotterdam, Netherlands
Co-authors: Lennart Hoogerheide, Peter de Knijf, Herman K. van Dijk

In several applications the data comes from a non-standard, multimodal distribution. In such cases, standard exploratory data analysis can be misleading since possible multimodality is not taken into account. A Bayesian approach is proposed for detecting the number of distinct modes in the data. The proposed method is also suitable to estimate quantiles of the data distribution. A mixture of shifted Poisson distributions and a mixture of normal distributions are proposed to estimate the probability of multimodality in discrete and continuous data, respectively. The method is illustrated with simulation studies and applied to three datasets, namely DNA tandem repeats data, number of defaulted installments in a financial institution in Spain and cross-country economic growth data. The results are compared with those of the standard frequentist tests.

C772: **A general Bayesian MIDAS regression approach with application to data frequency selection**
Presenter: **Antonietta Mira**, University of Lugano, Switzerland
Co-authors: Eric Ghysels, Reza Solgi

A general Bayesian MIDAS regression is proposed in which the lag coefficients are modeled nonparametrically. In this model the lag coefficients are decreasing or hump-shaped without being forced to follow any specific parametric form. We argue that, in some applications, the proposed framework may provide a more flexible setting in comparison with parametric MIDAS regression models (for instance with Almon or beta lag coefficients). In the second part we address the issue of frequency selection in MIDAS regressions in a Bayesian setting. We demonstrate that the Bayesian model selection tools can be used to select the true sampling frequency of the right hand side variables.

CS14 Room N2 RECENT DEVELOPMENTS IN VOLATILITY MODELLING Chair: Christian Conrad

C426: **Risk forecasting in (T)GARCH models with uncorrelated dependent innovations**
Presenter: **Helmut Herwartz**, Georg-August-University Goettingen, Germany
Co-authors: Benjamin Beckers, Moritz Seidel

(G)ARCH type models have proven quite useful for the dynamic modeling and forecasting of risk attached to speculative asset returns. While the symmetric and conditionally Gaussian GARCH model has been generalized in a magnifold of directions, for purposes of (quasi) maximum likelihood estimation model innovations are throughout presumed to stem from an underlying iid distribution. From an empirical perspective we notice that GARCH implied model innovations are likely at odds with the commonly held iid assumption for a cross section of 18 stock market indices. Drawing on this observation on the one hand, and noticing the heterogeneity of actual dependence patterns on the other hand, we follow two (complementary) strategies to evaluate the conditional distributions of consecutive GARCH innovations, a nonparametric approach and a semiparametric model class of standardized copula distributions. Modeling higher order dependence patterns is found to improve standard (threshold) GARCH implied conditional value-at-risk and expected shortfall out-of-sample forecasts that rely on the notion of iid innovations.

C531: **Misspecification testing in GARCH-MIDAS models**
Presenter: **Melanie Schienle**, Leibniz University Hannover, Germany
Co-authors: Christian Conrad

The aim is to develop a misspecification test for the multiplicative two-component GARCH-MIDAS model previously suggested. In the GARCH-MIDAS model a short-term unit variance GARCH component fluctuates around a smoothly time-varying long-term component which is driven by the dynamics of a macroeconomic explanatory variable. We suggest a Lagrange Multiplier statistic for testing the null hypothesis that the macroeconomic variable has no explanatory power. Hence, under the null hypothesis the long-term component is constant and the GARCH-MIDAS reduces to the simple GARCH model. We provide asymptotic theory for our test statistic and investigate its finite sample properties by Monte Carlo simulation. Our test statistic can be considered as an extension of the Lundbergh and Terasvirta test for evaluating GARCH models.

C1291: **Transformed polynomials for modeling conditional volatility**
Presenter: **Marco Bazzi**, University of Padua, Italy
Co-authors: Francisco Blasques, Andre Lucas, Siem Jan Koopman

A flexible model for filtering time-varying conditional volatilities is proposed. In particular, we make use of novel transformed polynomial functions to update the unobserved time-varying conditional volatility parameter. The flexible updating equation is shown to approximate arbitrarily well any continuous function and to have known convergence rates of approximation on Holder spaces. We derive conditions for strict stationarity, ergodicity, fading memory and filter invertibility. We also establish the consistency and asymptotic normality of the parametric and semi-nonparametric ML

estimator. A Monte Carlo study studies the finite sample properties of the estimator. Finally, a number of applications shows the good performance of the model in empirically relevant settings.

C872: **Modelling returns and volatilities during financial crises: a time varying coefficient approach**
Presenter: **Stavroula Yfanti**, Brunel University, United Kingdom
Co-authors: Menelaos Karanasos

It is examined how the most prevalent stochastic properties of key financial time series have been affected during the recent financial crises. In particular we focus on changes associated with the remarkable economic events of the last two decades in the mean and volatility dynamics, including the underlying volatility persistence and volatility spillovers structure. Using daily data from several key stock market indices we find that stock market returns exhibit time varying persistence in their corresponding conditional variances. Furthermore, the results of our bivariate GARCH models show the existence of time varying correlations as well as time varying shock and volatility spillovers between the returns of FTSE and DAX, and those of NIKKEI and Hang Seng, which became more prominent during the recent financial crisis. Our theoretical considerations on the time varying model which provides the platform upon which we integrate our multifaceted empirical approaches are also of independent interest. In particular, we provide the general solution for low order time varying specifications, which is a long standing research topic. This enables us to characterize these models by deriving, first, their multistep ahead predictors, second, the first two time varying unconditional moments, and third, their covariance structure.

CS20 Room H2 BANKS AND THE MACROECONOMY: EMPIRICAL MODELS FOR STRESS TESTING Chair: Rochelle M. Edge

C369: **Individual-bank income and expense forecasts - How multi-response PLS methods compare**
Presenter: **Rochelle Edge**, Federal Reserve Board, United States
Co-authors: Luca Guerrieri

The aim is to consider the possible usefulness of multiple-response partial least squares (PLS) regression to generate conditional forecasts of bank Net Interest Margins and other components of bank Pre-Provisioning Net Revenue. Our focus is on forecasting individual banks' key income and expense variables, particularly, for those banks that participate in the Federal Reserve's annual stress tests. We focus on these variables since their forecasts - conditional on macroeconomic conditions - are an important component of the stress tests. A key characteristic of PLS that differentiates it from other data-reduction methods is its ability to extract factors from a large dataset that efficiently summarize the original dataset, but also take account of the factors correlation with the dependent variables. Our motivation for considering multiple-response PLS regression methods stems from our earlier research, which found single-response PLS to be competitive with the best performing (albeit still relative poor) alternative models in forecasting aggregate NIMs and individual-bank NIMs. Multiple-response PLS also appears to represent a reasonable way to generate forecasts of a variable for a large number of banks when macroeconomic conditions (that are the same for all banks) are the main variables in the model.

C603: **Vector autoregressive models for scenario design: a horse-race comparison**
Presenter: **Jeremy Ching-Wai Chiu**, Bank of England, United Kingdom

Various vector autoregressive (VAR) models in the context of scenario design in the United Kingdom are evaluated. The motivation is to generate distributions of scenarios for use in top-down stress-testing models - such as the Bank of England's RAMSI - and so explore the implied distribution of outcomes for the banking system. Seeking to address the potentially non-linear relationships between macro variables and to tackle the problem that the data sample includes relatively few severe recessions, the paper studies a range of VAR models which improve on the commonly-used Gaussian VAR models with Minnesota-type priors. In particular, we investigate the following four models: (i) a BVAR model with fat-tailed errors and stochastic volatility; (ii) a regime switching VAR model; (iii) a BVAR model with time-varying parameters and stochastic volatility; (iv) a panel BVAR model with priors calibrated to international financial crisis data. Recursive in-sample forecasts of major macroeconomic variables, especially during the Great Recession period, are evaluated across the four models. This horse-race comparison will shed light on the capability of each approach to produce a distribution of plausible - including adverse - economic scenario projections, which will in turn provide policy makers with a promising tool to assess the resilience of the banking system to a range of economic scenarios.

C620: **Stress testing and interest rate risk**
Presenter: **Rhys Bidder**, Federal Reserve Bank of San Francisco, United States
Co-authors: Raffaella Giacomini

A general method for conducting stress testing using exponential tilting is considered. The framework allows for a broad class of nonlinear models, the presence of parameter uncertainty and the incorporation of extra-model information or "judgment" - issues of particular relevance when constructing severe and unusual scenarios. We illustrate the method with two applications. First, we consider the interest rate risk faced by the Federal Reserve due to its increased holdings of long-term bonds resulting from the unconventional policy of recent years. Second, we consider the risks faced by a hypothetical, but representative, financial institution. Our analysis makes use of term structure models that incorporate the zero lower bound on the short rate and other nonlinearities. Through the incorporation of macro factors in our term structure models we are able to make statements regarding the joint behavior of the macro-economy and interest rates. This allows the construction of direct and reverse stress test scenarios in terms of macroeconomic aggregates. As an example, we explore the implications of different assumptions over the timing and pace of departure from the zero lower bound and the ultimate normalization of policy.

C801: **Stress-test scenarios for the euro area: a large Bayesian VAR methodology**
Presenter: **Jerome Henry**, European Central Bank, Germany
Co-authors: Banbura Marta, Domenico Giannone, Michele Lenza, Michele Modugno

EU-wide stress tests have regularly been conducted since 2010 when the financial crisis turned into a sovereign one, especially affecting the euro area. Macro-financial scenarios that are employed for this purpose need to be assessed in terms of 'stress' relevance and internal consistency. We run such an analysis using a large size B-VAR for the euro area - comprising 42 macro-financial variables. The model is validated using 2007-09 data, ie the great recession seen as a "natural" experiment of stress-testing. While the stress in this period was severe, results show an environment broadly consistent with macroeconomic regularities, showing that non-linearities and structural breaks were limited, contrary to conventional wisdom about crises times. Conditional on publicly available data released to document the stress scenarios, the whole set of macro-financial variables is then projected using the B-VAR, demonstrating that these scenarios while severe remained plausible, ie depicted an environment broadly consistent with macroeconomic regularities. The exercise also reveals that credit and monetary aggregates significantly react to macroeconomic stress, questioning the so-called assumption of 'static' balance sheets, as used in the EU-wide stress tests, whereby banks keep both asset and liability size and structure constant over the stress-test horizon. The employed B-VAR could then be valuably employed to document such second round effects of stress-test scenarios.

CS30 Room I2 MIXTURE MODELS FOR FINANCIAL AND MACROECONOMIC TIME SERIES Chair: Markus Haas

C491: **Testing for misspecification in GARCH-type models**
Presenter: **Thomas Chuffart**, Aix-Marseille University, France
Co-authors: Emmanuel Flachaire, Anne Peguin-Feissolle

A large number of non-linear conditional heteroskedastic models have been proposed in the literature since the seminal works of Engle and Bollerslev. Practitioners need tools to choose the best specification.We propose a LM test based on a Taylor expansion to distinguish between

(G)ARCH models and an unknown nonlinear GARCH model. We derive the test statistic and we correct it with a bootstrap procedure. This new test can be seen as general misspecification test of a large set of GARCH-type univariate models. We investigate the size and the power of this test through Monte Carlo experiments and we compare it to other existing tests. Our test performs well when we are in the presence of regime mixture processes.

C573: **Asset returns density forecasting with MCD algorithms**
Presenter: **Marco Gambacciani**, University of Zurich and Swiss Financial Institute, Switzerland
Co-authors: Marc Paolella

A new approach for multivariate modeling of asset returns is proposed which accounts for all the major stylized facts, and also lends itself to portfolio optimization. It is based on a two-component multivariate normal mixture model, estimated using a new variation of the minimum covariance determinant (MCD) method. An empirical application demonstrates the viability of the proposed method, in terms of estimation speed and out of sample density prediction.

C276: **Modeling stock market returns with mixtures of skew-normal distributions**
Presenter: **Markus Haas**, University of Kiel, Germany

Markov normal mixture models are widely applied in modeling financial returns because they offer enormous flexibility and often admit an intuitive interpretation of the component densities. For example, in applications to stock index returns, researchers typically identify a component with high expected return and low volatility and a second component with the opposite pattern; thus these components can be interpreted as bull and bear market regimes, respectively. A potential disadvantage is that, with Gaussian regimes, any deviation from a symmetric density is automatically attributed to the bull/bear market asymmetry in expected returns. A more general model would allow the component densities to display asymmetries as well, thus allowing to assess the actual importance of both possible sources of skewness. This naturally leads to finite mixtures of skew-normal (SN) distributions. Empirical results for a large number of index return series suggest that the two sources of unconditional skewness are called for, i.e., the bull/bear market asymmetry as well as within-regime asymmetry. Mixtures of SN distributions feature both of these. Since skewness is an important property to account for in risk management, mixtures of SN distributions represent an attractive alternative to the widely used Gaussian mixture models.

CS48 Room P2 NOWCASTING AND FORECASTING MACRO-ECONOMIC TRENDS Chair: Gianluigi Mazzi

C163: **A nowcasting library in JDemetra+ for reading and visualizing news**
Presenter: **David de Antonio Liedo**, National Bank of Belgium, Belgium
Co-authors: Jean Palate

The aim is to document the computational methods underlying the new JDemetra+ library for the real-time analysis of macroeconomic data. Although the econometric systems supported in the original version are based on the literature of dynamic factor models, which are suitable to capture the strong comovements in macroeconomic time series, the analysis is valid for any model that has a linear state-space representation. News, which is defined as the difference between the data releases and the values expected by the model, will induce changes in the model based expectations for the current and future states of the economy.

C179: **Nowcasting U.S. headline and core inflation**
Presenter: **Edward Knotek**, Federal Reserve Bank of Cleveland, United States
Co-authors: Saeed Zaman

Forecasting future inflation and nowcasting contemporaneous inflation are difficult. We propose a new and parsimonious model for nowcasting headline and core inflation in the U.S. price index for personal consumption expenditures (PCE) and the consumer price index (CPI). The model relies on relatively few variables and is tested using real-time data. The model's nowcasting accuracy improves as information accumulates over the course of a month or quarter, and it easily outperforms a variety of statistical benchmarks. In head-to-head comparisons, the model's nowcasts of CPI inflation outperform those from the Blue Chip consensus, with especially significant outperformance as the quarter goes on. The model's nowcasts for CPI and PCE inflation also significantly outperform those from the Survey of Professional Forecasters, with similar nowcasting accuracy for core inflation measures. Across all four inflation measures, the model's nowcasting accuracy is generally comparable to that of the Federal Reserve's Greenbook.

C581: **Forecasting German key macroeconomic variables using large dataset methods**
Presenter: **Inske Pirschel**, Kiel Institute for the World Economy, Germany
Co-authors: Maik Wolters

The forecasting performance of three alternative large scale approaches is studied using a dataset for Germany that consists of 123 variables in quarterly frequency. These three approaches handle the dimensionality problem evoked by such a large dataset by aggregating information, yet on different levels. We consider different factor models, a large Bayesian vector autoregression and model averaging techniques, where aggregation takes place before, during and after the estimation of the different models, respectively. We find that overall the large Bayesian VAR and the Bayesian factor augmented VAR provide the most precise forecasts for a set of eleven core macroeconomic variables, including GDP growth and CPI inflation, and that the performance of these two models is relatively robust to model misspecification. However, our results also indicate that in many cases the gains in forecasting accuracy relative to a simple univariate autoregression are only moderate and none of the models would have been able to predict the Great Recession.

C686: **Forecasting in a DSGE model with banking intermediation: evidence from the US**
Presenter: **Alessia Paccagnini**, Bicocca University, Italy
Co-authors: Roberta Cardani, Stefania Villa

The aim is to examine the forecasting performance of a standard dynamic stochastic general equilibrium (DSGE) model augmented with a banking sector using Bayesian methods for the US economy. It focuses on a comparative evaluation of the out-of-sample predictive performance of the estimated DSGE model and non-structural benchmarks, such as hybrid DSGE models (in the form of a DSGE-VAR), and tested against a restricted DSGE model (without the banking sector). For each econometric specification we present point and density forecasts of real gross domestic product, harmonized consumer price index, federal funds interest rate and net worth of banks for the out-of-sample testing period 2001Q1-2008Q4. Our exercise provides a useful evaluation of DSGE models for policy oriented analysis.

CS73 Room G2 REGIME SWITCHING, FILTERING, AND PORTFOLIO OPTIMIZATION Chair: Joern Sass

C884: **Regime switching in continuous time and filter-based volatility**
Presenter: **Joern Sass**, University of Kaiserslautern, Germany
Co-authors: Vikram Krishnamurthy, Elisabeth Leoff

A continuous time regime switching model, where the observation process is a diffusion whose drift and volatility coefficients jump governed by a continuous time Markov chain, can explain some of the stylized facts of asset returns, even in this simple linear and non-autoregressive form. But due to the switching volatility, in continuous time the underlying Markov chain could be observed and no filtering is needed (in theory). Therefore, if in finance explicit theoretical results are obtained, they may not provide a good approximation for the discretely observed model in which we

have to filter. On the other hand, a continuous time hidden Markov model (HMM), where only the drift jumps and the volatility is constant, allows for explicit calculations but has no such good econometric properties. We first discuss estimation, model choice and portfolio optimization in both models. To combine useful aspects of both models, we then look at a HMM where the volatility depends on the filter for the underlying Markov chain. We analyze its relation to Markov switching models and, using examples from portfolio optimization, we illustrate that we can still get quite explicit results and that these provide a good approximation to the discretely observed model.

C875: **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. The study builds on this growing literature by extending recent results obtained from the Generalized Dynamic Factor Model (GDFM), thus avoiding the limiting assumptions imposed by the use of static factor models. The comparison of GDFM analyses for time periods before, during, and after the 2007-2009 financial crisis provides consistent evidence that commonality both within and across market liquidity dimensions increases following periods of extreme liquidity conditions. Furthermore, a measure of market liquidity constructed by extracting the largest common factor across multiple liquidity dimensions is shown to be significantly priced in the cross-section of asset returns.

C879: **Generalized Pareto processes and liquidity**
Presenter: **Sascha Desmettre**, University of Kaiserslautern, Germany
Co-authors: Johan de Kock, Frank Thomas Seifried

Liquidity risk in fund management has become an important topic over the last years and the employment of an appropriate liquidity risk management has become mandatory by European guidelines. In that regard, recent results show that the generalized Pareto distribution is suitable for the modeling of liquidty risks by computing the POT-quantile. We introduce the autoregressive generalized Pareto (ARGP) process which is able to cover the time structure of fund redemption data. The ARGP process possesses thereby two regimes; one covering periods of normal redemptions and one covering periods of exceptionally high redemptions. We give the formal defintion of an autoregressive process with GPD tails, prove its stationarity, compute its joint distribution, estimate its model parameters and apply it to real data.

C894: **Expert opinions and optimal portfolio strategies under partial information**
Presenter: **Ralf Wunderlich**, BTU Cottbus-Senftenberg, Germany

Optimal portfolio strategies are investigated for utility maximizing investors in a market where the drift is driven by an unobserved Markov chain. Information on the state of this chain is obtained from stock prices and expert opinions in the form of signals at random discrete time points. These signals we model by a marked point process with jump-size distribution depending on the current state of the hidden Markov chain. We use stochastic filtering to transform the original problem into an optimization problem under full information where the state variable is the filter for the Markov chain. This problem is studied with dynamic programming techniques and with regularization arguments. For the case of power utility we present results from the numerical solution of the dynamic programming equation.

CS76 Room M2 TOPICS IN FINANCIAL ECONOMETRICS Chair: Leopold Soegner

C247: **Adaptive control variables for estimating functionals of diffusion processes**
Presenter: **Klaus Poetzelberger**, WU Vienna, Austria

The toolbox of Monte Carlo simulation provides - easy to implement - procedures for estimating functionals of stochastic processes such as the c.d.f. of marginal distributions, boundary crossing probabilities, c.d.f.'s of first exit times or prices of certain path dependent options. There exist procedures which generate paths on a discrete time grid from the exact distributions, but most procedures are biased in the sense that even on the discrete time grid, the distribution from which the samples are drawn is an approximation. The MSE (mean squared error) is then the sum of the squared bias and a variance term. If N univariate random variables are used, n discrete paths of lengths $m = N/n$ are generated, the variance is of order $1/n$, but the MSE is of order $1/N^{\gamma}$ with $\gamma < 1$. Naive applications of MC often have a MSE of order $1/\sqrt{N}$ only! The aim is to present as variance reduction technique the method of adaptive control variables. 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. Examples of applications such as estimating boundary crossing probabilities for k-dimensional Brownian motion or diffusions and results on approximation rates are given.

C255: **A predictive Bayesian model averaging approach on firm default probabilities**
Presenter: **Laura Vana**, WU Vienna University of Economics and Business, Austria
Co-authors: Bettina Gruen, Paul Hofmarcher, Kurt Hornik

The importance of identifying key factors to predict corporate firm failures has become more evident in the aftermath of the global financial crisis. Statistical models in credit risk propose a range of different financial ratios for predicting a company's ability to meet its financial obligations. One such model is Altman's prominent Z-score model, which predicts bankruptcy using five financial ratios chosen from four different categories (i.e., liquidity, profitability, etc.). However, due to lack of theory, no preferred set of ratios is indicated to build a simple, but well performing predictive model. We address this problem of model choice and selection of financial ratios for inclusion in the model by collecting an extensive list of ratios from Compustat North America and comparing the ability to predict rating-implied default probabilities of parsimonious models where not only the total number of regressors is restricted, but also their assignment to different ratio categories. Thus the structure of the considered models closely resembles Altman's Z-score model. Bayesian model averaging based on a predictive model performance criterion provides the framework for comparing these models and allows us to assess the importance and robustness of the different financial ratios for a fair comparison to the Z-score model.

C811: **How to sample from a distribution when only the characteristic function is known**
Presenter: **Filippo Macaluso**, University of Lugano, Switzerland
Co-authors: Antonietta Mira, Paul Schneider

A novel efficient simulation-based procedure is developed to sample from a multivariate distribution when only its characteristic function is known, as often happens in financial applications. To achieve this goal, we combine two strands of the statistical literature: The first one is concerned with the approximation of the density ratio of the original target to an auxiliary measure by orthonormal polynomial series in weighted L^2 spaces, the second relates to simulation-based methods where the target density is not available and thus an approximation is used.

C857: **Evaluation of expected shortfall forecasts**
Presenter: **Justinas Pelenis**, Institute for Advanced Studies, Vienna, Austria

The aim is to discuss and address the issue of the evaluation of competing expected shortfall (conditional value-at-risk) forecasts. We plan to provide a critical summary and review of existing methods for the comparison of expected shortfall forecasts and provide new directions in possible evaluation methods of forecasts. Additionally, we will consider whether any of the evaluation methods could successfully be used for estimation purposes as well.

CS82 Room B2 BOOTSTRAPPING TIME SERIES AND PANEL DATA Chair: Jean-Pierre Urbain

C915: **Bootstrap-based tests for multiple structural changes in linear models with endogenous regressors**
Presenter: **Adriana Cornea-Madeira**, University of York, United Kingdom
Co-authors: Otilia Boldea, Alastair Hall

The aim is to derive the limiting distributions of break-point tests in models estimated via two-stage least squares, when the first-stage is unstable. We show that the presence of breakpoints in the first-stage renders the asymptotic distribution of the second-stage break-point tests non-pivotal. To obtain critical values, we propose bootstrapping these tests via the recursive-design wild bootstrap and fixed-design wild bootstrap, and prove their asymptotic validity. We provide simulation results that show good finite sample performance of our procedures, even in cases where the tests remain pivotal. Via an application to the US interest rate reaction function that illustrates our methods, we confirm recent findings of instability in the Taylor rule.

C1082: **Estimating unobservable common trends in small samples using panel cointegration methods**
Presenter: **Stefano Fachin**, Rome Sapienza, Italy
Co-authors: Francesca Di Iorio

Non stationary panel models allowing for unobservable common trends have recently become very popular, for instance in the stream of literature devoted to the analysis of cross country growth differentials. In this type of models the latent trend coincides with the residual traditionally identified as total factor productivity. However, standard methods, which are based on factor extraction or models augmented with cross-section averages, require large sample sizes, not always available in practice.In these cases we propose the simple and robust alternative of augmenting the panel regression with common time dummies. The underlying assumption of additive effects can be tested by means of a panel cointegration test designed for the null hypothesis of cointegration in no units of the panel, with no need of estimating a general interactive effects model. An application to modelling labour productivity growth in the four major European economies (France, Germany, Italy and UK) illustrates the method.

C1121: **Bootstrap simultaneous confidence bands for time-varying coefficient models**
Presenter: **Marina Friedrich**, Maastricht University, Netherlands
Co-authors: Stephan Smeekes

Simultaneous confidence bands for nonparametric estimators of time-varying coefficient regression models are developed. For this purpose an existing sieve bootstrap method for deterministic trends is extended to time-varying coefficient models. The autoregressive sieve is based on nonparametric local constant or local linear estimation of the coefficient curves. The asymptotic properties of the method are investigated. An extensive simulation study and an empirical application show that the proposed method performs well in finite samples in the presence of serial correlation.

C1181: **Block bootstrap theory for multivariate integrated and cointegrated processes**
Presenter: **Carsten Jentsch**, University of Mannheim, Germany
Co-authors: Efstathios Paparoditis, Dimitris Politis

Some asymptotic theory is developed for applications of block bootstrap resampling schemes to multivariate integrated and cointegrated time series. It is proved that a multivariate, continuous-path block bootstrap scheme applied to a full rank integrated process, succeeds in estimating consistently the distribution of the least squares estimators in both, the regression and the spurious regression case. Furthermore, it is shown that the same block resampling scheme does not succeed in estimating the distribution of the parameter estimators in the case of cointegrated time series. For this situation, a modified block resampling scheme, the so-called residual based block bootstrap, is investigated and its validity for approximating the distribution of the regression parameters is established. The performance of the proposed block bootstrap procedures is illustrated in a short simulation study.

CS96 Room E2 ECONOMETRIC MODELS FOR MIXED FREQUENCY DATA Chair: Joerg Breitung

C029: **Large Bayesian mixed-frequency vector autoregressions**
Presenter: **Thomas Goetz**, Maastricht University, Netherlands
Co-authors: Alain Hecq

Granger causality testing is analyzed in a mixed-frequency VAR where the difference in the variables' sampling frequencies is large. To tackle the parameter proliferation issue for realistic sample sizes we propose a Bayesian mixed-frequency VAR, for which we extend a previous dummy observation approach to a mixed-frequency setup, presenting an alternative to classical Bayesian estimation techniques. Starting from prior beliefs we show precisely how to choose the moments of the prior distributions when faced with mixed frequencies in the data. Subsequently, we compare this method to a common low-frequency model as well as to the unrestricted VAR in terms of their Granger non-causality testing behavior using Monte Carlo simulations. Finally, the techniques are illustrated in an empirical application involving daily realized volatility and monthly business cycle fluctuations.

C031: **Estimating vector autoregressions with mixed frequency data**
Presenter: **Joerg Breitung**, University of Bonn, Germany
Co-authors: Natalia Soldatenkova

Several alternative approaches are proposed for estimating mixed frequency Vector Autoregressions. Previous work employs Bayesian approaches or ML estimation techniques based on the Kalman filter. An important drawback of all these approaches is the prohibitive computational burden even for a moderate number of variables and lags. We propose simple estimators for the VAR parameters and the missing values based on a non-linear LS method and a simple iterative estimator. The idea is to replace the full information EM step in the Kalman recursions by simpler estimators of the missing values. Specifically we show that the updating step of the Kalman filter involves a nonlinear least-squares problem that can be used to construct a simplified limited information ML estimator. Monte Carlo simulations suggest that the computationally convenient alternative estimators perform comparably to the ML estimator, although the Kalman filter technique tends to produce slightly lower standard errors. We also study the problem of estimating the MIDAS-VAR when the low frequency variables result from linear transformations of high-frequency observations, which naturally arises whenever the low-frequency time series is a flow variable. The leading example is quarterly growth of GDP that results as a particular linear combination of (unobserved) monthly observations.

C034: **MIDAS regressions with time-varying parameters**
Presenter: **Christian Schumacher**, Deutsche Bundesbank, Germany

Mixed-data sampling (MIDAS) regressions allow the estimation of dynamic equations that explain a low-frequency variable by high-frequency variables and their lags. To account for temporal instabilities in this relationship, an extension to MIDAS is discussed with time-varying parameters (TVP-MIDAS), which follow random-walk processes. The model does also allow for the estimation of stochastic volatility. The resulting non-linear functional forms in the MIDAS regression necessitate the use of non-linear filtering techniques. The Particle Filter with parameter learning is used to estimate the time-varying parameters of the model. Simulations with time-varying DGPs help to assess the properties of the estimation approach. In the applications, we show how TVP-MIDAS can be used to investigate the stability of business cycle indicators over time.

C040: **Density forecasts with MIDAS models**
Presenter: **Knut Are Aastveit**, Norges Bank, Norway
Co-authors: Claudia Foroni, Francsco Ravazzolo
The existent MIDAS approach is extended to obtain density forecast. Firstly, we compute density forecasts from different MIDAS models. In particular, we consider the classical MIDAS models and the unrestricted version. We provide Monte Carlo simulations that help comparing the different MIDAS models in terms of their forecasting performance. We find that the results in terms of point and density forecasts are coherent. Moreover, the results are not clearly indicating a superior performance of one of the models under scrutiny, especially when the persistence of the low frequency variable is low. Some differences are instead more evident when the persistence is high, for which the AR-UMIDAS seems to perform better. Finally, we carry out an empirical exercise. We evaluate point and density forecasts computed for the US output growth, by exploiting typical monthly series.

CP01 Room First floor Hall POSTER SESSION Chair: Francisco de Asis Torres-Ruiz

C327: **State space models for hourly electricity prices**
Presenter: **Paolo Chirico**, University of Turin, Italy
Structural time series models are not trendy in modelling electricity prices. Actually, scholars prefer using models of the ARIMA-GARCH class in nearly every case. That is clear if the target of analysis is the only prediction of price and volatility. However, ARIMA modelling is not suitable for explaining seasonality that plays a main role in the dynamics of the electricity prices, particularly of hourly prices. On the other hand, structural time series models in state space form make that very well. Two structural time series models for hourly electricity prices have been performed: (i) a local level model with hourly effects for the expected hourly price; (ii) a seasonal stochastic volatility model for the hourly volatility of the prices. The estimations of the models have been got by Kalman filters, and the results have pointed out some interesting findings: (i) the hourly effects change little over time; (ii) the prices volatility present an hourly pattern with high persistence; (iii) the seasonal stochastic model is better performing than similar EGARCH models.

C391: **The study of multi-regimes switching copula models**
Presenter: **Anna Czapkiewicz**, Polish Academy of Science, Poland
Co-authors: Pawel Jamer
The knowledge about financial market dependance enables us to diversify the risk of investments. In the literature there have been proposed models in which the dynamic dependence between financial time series have been considered, such as the dynamic conditional correlation model or models based on the copulas theory, such as the regime switching copula models. Discussed in the literature the regime switching copula model generally considers only two regimes. The aim is to verify the usefulness of the switching Copula-GARCH model with more regimes to describe the dependencies between financial markets. For the multi-regimes model contruction purposes the normal copula, t-Student copula and some Archimedean copulas are taken into consideration. To assess the suitability of the model with various copulas the proper tests are applied. For selecting among competing models, also for non-nested ones, some procedures are carried out.

C1106: **On the modelling and forecasting multivariate realized volatility: generalized heterogeneousautoregressive (GHAR) model**
Presenter: **Frantisek Cech**, Charles University in Prague, Czech Republic
Co-authors: Jozef Barunik
A methodology is introduced for dynamic modelling and forecasting of realized covariance matrices based on generalization of the heterogeneous autoregressive model (HAR) for realized volatility. Multivariate extensions of popular HAR framework leave substantial information unmodeled in residuals. We propose to employ a system of seemingly unrelated regressions to capture the information. The newly proposed generalized heterogeneous autoregressive (GHAR) model is tested against natural competing models. In order to show the economic and statistical gains of the GHAR model, portfolio of various sizes is used. We find that our modeling strategy outperforms competing approaches in terms of statistical precision, and provides economic gains in terms of mean-variance trade-off. Additionally, our results provide a comprehensive comparison of the performance when realized covariance and more efficient, noise-robust multivariate realized kernel estimator, is used. We study the contribution of both estimators across different sampling frequencies, and we show that the multivariate realized kernel estimator delivers further gains compared to realized covariance estimated on higher frequencies.

C1056: **Using regression models to classify national football teams according to the number of goals**
Presenter: **Jose Rodriguez-Avi**, University of Jaen, Spain
Co-authors: Maria Jose Olmo-Jimenez, Antonio Conde-Sanchez, Antonio Jose Saez-Castillo, Ana Maria Martinez-Rodriguez
An important indicator about the quality of a football team is the number of goals scored and/or received. In turn, this number depends on several variables such as the type of match (official, friendly...), the type of competition (continental, World championship...) and the rival team (World championship team...), among others. The number of goals is a count data variable, so a methodology based on regression models for count data is proposed in order to explain the behavior of several national football teams and to classify them. Specifically data about World Championship teams and about teams of second and third level are fitted using the following regression models: Poisson regression model, negative binomial regression model, generalized Waring regression model and complex biparametric regression model. All the fits are compared to each other using the Akaike information criterion. So, the effect of "the field factor" or the difference of goals scored if the match has been won or lost by the team, for instance, can be estimated. Moreover, the analysis of significant covariates provides additional information of interest about the quality of the team.

C1176: **Change in trading rules and its impact on the distributional properties of commodity futures**
Presenter: **Yoshinori Kawasaki**, The Institute of Statistical Mathematics, Japan
Co-authors: Yoshimitsu Aoki
Until April 2009, Tokyo Commodity Exchange (TOCOM) limited price movement in such a way that the commodity-wise caps and floors were determined dependent on recent price levels. In May 2009, TOCOM changed the transaction rules and introduced a circuit breaker system. This institutional conversion has entirely altered the distributional properties of the price difference of commodity futures. The bottom line is that the tail-heaviness in the downside has been observed under the new rules. Under price limitation regime, we fit autoregressive models to the censored price difference series which inevitably exploits MCMC technique. Generally, normal distribution seems appropriate and no severe asymmetry is found. Under the circuit breaker system, considering the asymmetry in the empirical distribution of price difference series, we fit low order autoregressive models driven by (1) normal, (2) asymmetric Laplace, and by (3) the half-exponential half-normal distribution. In our analysis also done by MCMC, the half-exponential half-normal distribution fits best for the most of commodity futures in TOCOM.

C1193: **GARCH model for income time series data with income inequality**
Presenter: **Haruhisa Nishino**, Chiba University, Japan
Income inequality is a great issue also in the Japanese economy as in the world economy. An approach is taken by use of a parametric model for analyzing income inequality. Lognormal distribution, which is used, is better fitted to Japanese income data and useful to extract inequality from income data because its scale parameter only represents inequality. GARCH models are proposed including income inequality for income time series data by use of lognormal distribution. The GARCH model is suitable for modelling the scale parameter of lognormal distribution. A

joint distribution from selected order statistics enables us to construct a likelihood of the GARCH model from quantile data. The proposed model has an inequality structure and a time series structure. It is useful for analyzing persistent income inequality and a dynamic change of income inequality. The models are estimated from Japanese quantile income data (Family Income and Expenditure Survey by Ministry of Internal Affairs and Communications, Statistics Bureau, Japan) and the various GARCH models are compared and examined. The historical change of the Japanese income inequality is finally shown.

C1261: **Quantile absolute prediction error measure in claim frequency mixed model**

Presenter: **Alicja Wolny-Dominiak**, University of Economics in Katowice, Poland

Co-authors: Tomasz Zadlo, Wojciech Gamrot

Prediction theory finds numerous successful applications in economic decision making. Properties of different predictors are compared using sample statistics which measure various aspects and dimensions of prediction quality such as goodness-of-fit, risk or accuracy. For example, to compare relevance of two models, one can use among others the coefficient of determination, residual variance, Bayesian Information Criterion or Akaike's Information Criterion. To compare risks of alternative portfolios, variance, semivariance, value at risk or conditional value at risk are often used. In some applications such as estimating index numbers the whole distribution of sample statistics may be of interest. One of the most popular prediction accuracy measures is the mean squared error of prediction and its modifications such as heteroskedasicity adjusted mean squared error. We propose the quantile measure of prediction accuracy and the estimation procedure for a parametric model with fixed and random effects with an application in claim frequency prediction, which is important in ratemaking problem.

Sunday 7.12.2014 08:45 - 10:25 Parallel Session G – ERCIM

ESI03 Room Sala Convegni TIME SERIES MODELING AND COMPUTATION Chair: Roland Fried

E210: **Testing for independence between functional time series**
Presenter: **Lajos Horvath**, University of Utah, United States
Co-authors: Gregory Rice

Frequently econometricians are interested in verifying a relationship between two or more time series. Such analysis is typically carried out by causality and/or independence tests which have been well studied when the data is univariate or multivariate. Modern data though is increasingly of a high dimensional or functional nature for which finite dimensional methods are not suitable. We develop a methodology to check the assumption that data obtained from two functional time series are independent. Our procedure is based on the norms of empirical cross covariance operators and is asymptotically validated when the underlying populations are assumed to be in a class of weakly dependent random functions which include the functional ARMA, ARCH and GARCH processes.

E861: **Cointegration and phase synchronization: bridging two theories**
Presenter: **Rainer Dahlhaus**, University of Heidelberg, Germany
Co-authors: Jan Neddermeyer

A new multivariate time series model with VEC-state oscillators is presented for oscillators with random phases. In particular the phases may be synchronized. The model is a nonlinear state space model where the phase processes follow a vector error correction model used in econometrics to model cointegration. We demonstrate the relevance of this model for phase synchronization. In that way we bridge the theories of cointegration and phase synchronization which have been important theories in econometrics and physics, respectively. The common ground of both theories is that they describe the fluctuation of some multivariate random process around an equilibrium. We demonstrate how the methods from cointegration can be applied to phase synchronization. In particular, we consider a unidirectionally coupled Rossler-Lorenz system and identify the unidirectional coupling, the phase synchronization equilibrium and the phase shifts with cointegration tests.

E1073: **Outliers detection in unobserved dynamic harmonic regression models**
Presenter: **Antonio Garcia-Ferrer**, Universidad Autonoma Madrid, Spain
Co-authors: Marcos Bujosa, Pilar Poncela

Recent empirical results have shown that most seasonal adjustment methods are heavily influenced by the presence of different types of outliers. In these circumstances, the estimated seasonally adjusted series still show considerable seasonal evidence. While this situation has been successfully addressed in ARIMA based methodologies like SEATS/TRAMO or variants of X-12, this is not the case in unobserved DHR models. We analyze the effects of outliers in these models, both at the identification and estimation stages by using simulated and real data. Both in the case of the simple trend + noise model, as well as in the complete seasonal DHR model, preliminary results show that while single outliers do not affect the model's identification, they do considerably affect the parameter estimates.

ES12 Room F1 MULTIVARIATE SURVIVAL ANALYSIS AND MULTI-STATE MODELS Chair: Jacobo De Una-Alvarez

E154: **The illness death model under left truncated and right censored data**
Presenter: **Micha Mandel**, The Hebrew University of Jerusalem, Israel
Co-authors: Bella Vakulenko-Lagun

Left truncated data arise when a lifetime variable T and an independent truncation variable L are observed only if $L < T$. There are several ways to perform statistical inference under this setting. One can (i) condition on the event $\{L < T\}$ only, (ii) condition on the event $\{L < T\}$ and on L, or (iii) condition on the event $\{L < T\}$, on L and on all the history up to time L. When covariates are time independent, the latter two approaches are exactly the same. However, the situation becomes more complicated when multi-state models are considered, as approaches (ii) and (iii) differ. More specifically, using appraoch (iii) in the illness-death model leads to loss of important information as subjects truncated in the illness state do not contribute to estimation of functionals related to the healthy state. This information can be exploited using approaches (i) and (ii), but new and more complicated estimation methods are required. We discuss regression models under various assumptions and show that estimators obtained using approaches (i) and (ii) perform better than estimators obtained under (iii). The methods are applied to ICU data collected in a cross-sectional design, where the illness state corresponds to blood-stream infection.

E186: **Nonparametric estimation of transition probabilities in the non-Markov illness-death model: a comparative study**
Presenter: **Jacobo de Una-Alvarez**, Universidad de Vigo, Spain
Co-authors: Luis Meira-Machado

Multi-State models can be successfully used for modelling complex event history data. In these models the estimation of the transition probabilities is of particular interest since they allow for long-term predictions of the process. These quantities have been traditionally estimated by the Aalen-Johansen estimator, which is consistent if the process is Markovian. Several non-Markovian estimators have been proposed in the recent literature, and their superiority relative to the Aalen-Johansen estimator has been demonstrated in situations in which the Markov condition is strongly violated. However, the existing estimators have the drawback of requiring the support of the censoring distribution to contain the support of the life distribution, which is not often the case. We propose two new sets of estimators for estimating the transition probabilities, and some asymptotic results are derived. The proposed estimators are consistent regardless of the Markov condition and the referred assumption on the censoring support. We explore the finite sample behavior of the estimators through simulations. The main conclusion is that the proposed estimators may be much more efficient than the competing estimators. Real data illustration from a clinical trial on colon cancer is included.

E705: **Direct modeling of regression effects for transition probabilities in the progressive illness-death model**
Presenter: **Leyla Azarang**, University of Vigo, Spain
Co-authors: Thomas Scheike, Jacobo De Una-Alvarez

Multi-state models are often used to represent the individuals' progress along a certain disease. The estimation of transition probabilities is an important goal in such a setting. The progressive illness-death model is an important multi-state model which has many applications in medical research. We introduce direct binomial regression to model the covariate effects on transition probabilities in censored progressive illness-death model. To this end, a binomial regression model with time-varying effects is considered, where the sample is restricted to individuals whose sojourn time in state healthy and/or total survival time meet some conditions.

E706: **Methods for testing the Markov condition in the illness-death model: a comparative study**
Presenter: **Mar Rodriguez Girondo**, Leiden University Medical Center, Netherlands
Co-authors: Jacobo de Una-Alvarez

Illness-death models are often used in biomedicine for describing survival data when an intermediate event of interest may be observed during follow-up. However, usual estimators (e.g. Aalen-Johansen transition probabilities) may be systematically biased in non-Markovian situations. Despite non-Markovian estimators for transition probabilities and related curves are available, including the Markov information in the construction of the estimators allows for variance reduction. The practical superiority of non-Markovian estimators has been demonstrated in situations where

the Markov condition is "strongly" violated, however, the gain is questionable for weak deviations or semi-Markovian situations, where the variance of non-Markovian estimators may dominate the bias. On the other hand, several non-parametric methods for testing the Markov condition in the illness-death model have been recently proposed, with different assumptions and power to detect deviations from the future-past independence null hypothesis. We discuss several characterizations of the Markov condition and we evaluate, by means of intensive simulations, three different strategies for illness-death data at hand: a) using Markovian estimators, b) using non-Markovian estimators, or c) decision making according to different tests for Markovianity. We illustrate our findings through the analysis of a data set from stem cell transplant in leukemia.

ES26 Room O1 ADVANCES IN CLUSTER ANALYSIS Chair: M. Brigida Ferraro

E333: **Factor clustering and visualization for categorical data**
Presenter: **Francesco Palumbo**, University of Naples Federico II, Italy
Co-authors: Alfonso Iodice D'Enza, Michel van de Velden

A set of n multivariate statistical units described by p attributes is generally arranged in an $n \times p$ matrix X, whose rows are geometrically represented as points in the $\Re^p$ dimensional space. Two individuals are similar when they are 'close' to each other. Then a cluster is a subset of individuals that are close to each other: cluster analysis aims to identify the best partition of data. As p increases, clustering the units becomes more and more challenging, due to the so-called *curse of dimensionality*: distances between any pair of p-dimensional points tend to converge, then cluster analysis becomes an unfeasible task. In the particular case of categorical data, p depends on the total number of categories characterizing the considered variables. In addition, high dimensionality limits the effectiveness of commonly used cluster visualization tools. A popular approach to deal with such dimensionality issues is to sequentially apply dimension reduction and clustering. However, a more enhanced approach consists in combining dimension reduction and clustering: methods implementing such an approach are referred as factor clustering. The aim is to present some recent advances in factor clustering for categorical data. Furthermore, exploiting the low dimensional solution it presents some consistent visualization tools to assist the interpretation of the clusters.

E436: **Flexible parametric bootstrap for testing homogeneity against clustering and assessing the number of clusters**
Presenter: **Christian Hennig**, UCL, United Kingdom

Many cluster analysis methods deliver a clustering regardless of whether the dataset is indeed clustered or homogeneous, and need the number of clusters to be fixed in advance. Validation indexes such as the Average Silhouette Width are popular tools to measure the quality of a clustering and to estimate the number of clusters, usually by choosing the number of clusters that optimizes their value. Such indexes can be used for testing the homogeneity hypothesis against a clustering alternative by exploring their distribution, for a given number of clusters fitted by a given clustering method, under a null model formalising homogeneous data. The same approach can be used for assessing the number of clusters by comparing what is expected under the null model with what is observed under different numbers of clusters. Many datasets include some structure such as temporal or spatial autocorrelation that distinguishes them from a plain Gaussian or uniform model, but cannot be interpreted as clustering. The idea is to specify a null model for data that can be interpreted as homogeneous in the given application, which captures the non-clustering structure in the dataset by some parameters, which are estimated from the data, and then bootstrapping a cluster validity index can be used for testing homogeneity against a clustering alternative and for assessing the number of clusters. Several applications will be presented.

E737: **Mixtures of hidden truncation hyperbolic distributions**
Presenter: **Paula Murray**, University of Guelph, Canada
Co-authors: Ryan Browne, Paul McNicholas

A novel formulation of the hyperbolic distribution is introduced which we refer to as the hidden truncation hyperbolic (HTH) distribution. The HTH distribution encapsulates several distributions as special or limiting cases including the skew-t and multivariate-t distributions. We incorporate the HTH distribution into a mixture model for the purpose of model-based clustering. An expectation-maximization (EM) algorithm for the estimation of model parameters is presented. We present clustering results for both real and simulated data and discuss the ability of the HTH distribution to model asymmetric data as compared to other skewed distributions which have appeared in the literature.

E553: **Mixtures of skewed distributions with hypercube contours**
Presenter: **Brian Franczak**, McMaster University, Canada
Co-authors: Cristina Tortora, Ryan Browne, Paul McNicholas

Mixture models with components that have skewed hypercube contours are developed via a generalization of the multivariate shifted asymmetric Laplace density. Specifically, we develop mixtures of multiple scaled shifted asymmetric Laplace distributions. The component densities have two unique features: they include a multivariate weight function, and the marginal distributions are also asymmetric Laplace. We used these mixtures of multiple scaled shifted asymmetric Laplace distributions for clustering applications, but they could equally well be used in the supervised or semi-supervised paradigms. The expectation-maximization algorithm is used for parameter estimation and the Bayesian information criterion is used for model selection. Simulated and real data is used to illustrate the approach and, in some cases, to visualize the skewed hypercube structure of the components.

ES31 Room I1 APPLICATIONS IN ENGINEERING AND ECONOMICS Chair: Carolina Garcia-Martos

E534: **Variable importance assessment and prediction using regression trees: Application to electricity markets**
Presenter: **Jose Manuel Mira-McWilliams**, Technical University of Madrid, Spain
Co-authors: Camino Gonzalez, Isabel Juarez

The purpose is electricity price forecasting through regression tree methods. Electricity price forecasting has become the focus of considerable interest in a deregulated energy market. It is essential for suppliers and buyers to rely on reasonable forecasts for the elaboration of bidding strategies to mitigate losses and increase benefits. Three forms of regression tree models have been applied: CART, Bagging and Random Forests, which have been used in two directions: first, to identify the variables dominating the marginal price of the commodity and second, for short-run (one hour ahead and day ahead) electricity price forecasting for the Spanish-Iberian Market. Regression trees are a computationally efficient alternative to traditional statistical models, particularly suitable to accommodate non-linearities and variability in local behavior. The proposed models include the main features of the market: load, hydro and thermal generation, wind energy production, of strategic interest in the Spanish market, as well as lagged prices and hour, day, month and year indicators. To build the models, hourly data of 22 variables from 2000-2010 have been used. The results show the effectiveness of the proposed ensemble of tree-based models which emerge as a competitive and promising alternative to existing methods.

E661: **Volatility transmission between European Energy markets and CO2 prices**
Presenter: **Pilar Munoz**, Universitat Politecnica de Catalunya, Spain

Detecting and estimating volatility transmission in the European Energy Markets is vital for competitiveness among these markets and, of course, for analyzing the financial risk they generate. We propose to first study the univariate volatility of the most important European power markets: EXPEXSPOT (France, Germany, Austria (Phelix), Switzerland (Swissix)); OMIE (Iberian Peninsula); Nord Pool Spot (Nordic countries) and GME (Italy), as well as the transmission of volatility between those markets. Analyzing volatility transmission between these energy markets is the starting point for diversifying an investment portfolio. Finally we will analyze how CO2 prices affect these energy markets, as well as the "macro"

relationships between these markets and the Euro/USD exchange rate. The obtained data set contains business data and ranges from January, 2, 2008 to June, 30, 2014.

E765: **Optimal trading for wind power producers under uncertainty, using bootstraping techniques**
Presenter: **Eduardo Caro**, Technical University of Madrid, Spain
Co-authors: Beatriz Martin-Sierra, Carolina Garcia-Martos, Juan Miguel Morales Gonzalez

In liberalized electricity markets, both conventional and renewable electricity producers must trade in the day-ahead market in the same way, although the tomorrow's generation for renewable producers may not be accurately forecasted. In this case, the balancing market is used to mend the differences between the offered quantity and the power finally produced. Optimal trading for stochastic producers under uncertainty are analyzed, using bootstrapping techniques for the computation of the optimal quantity to be offered in the day-ahead electricity markets which maximizes the expected monetary value.

E854: **Extracting common trends from scram rates of nuclear power plants using Dynamic Factor Analysis**
Presenter: **Carolina Garcia-Martos**, Universidad Politecnica de Madrid, Spain
Co-authors: Camino Gonzalez-Fernandez, Jose Mira-McWilliams

The focus is in the evaluation of one of the performance indicators of commercial nuclear power plants, belonging to the initiate cornerstone event: the unpplanned scrams per 1000 critical hours. We estimate the scram rates by using Hierarchical Bayesian Models (HBM) and we analyze the trends in this indicator, both for Spanish and North American Power Plants. We use dimensionality reduction techniques for multivariate time series that consists of extracting unobserved common factors as well as modelling them as Vector ARIMA models.

ES32 Room G1 DIRECTIONAL STATISTICS Chair: Toshihiro Abe

E779: **Limit theorems on the spherical Poisson space**
Presenter: **Giovanni Peccati**, Luxembourg University, Luxembourg

Some new results and techniques are reviewed, allowing one to study the asymptotic behaviour of U-statistics based on Poisson measures defined on a sphere. Applications to cosmological data analysis will be discussed.

E375: **Normal approximations for wavelet coefficients on spherical Poisson fields**
Presenter: **Domenico Marinucci**, University of Rome Tor Vergata, Italy
Co-authors: Claudio Durastanti, Giovanni Peccati

The purpose is to compute explicit upper bounds on the distance between the law of a multivariate Gaussian distribution and the joint law of wavelet/needlet coefficients based on a homogeneous spherical Poisson field. In particular, we develop some results from previous works, based on Malliavin calculus and Stein's methods, to assess the rate of convergence to Gaussianity for a triangular array of needlet coefficients with growing dimensions. Our results are motivated by astrophysical and cosmological applications, in particular related to the search for point sources in Cosmic Rays data.

E780: **Gaussian approximations for nonlinear statistics on spherical Poisson spaces**
Presenter: **Claudio Durastanti**, University of Tor Vergata - Rome, Italy
Co-authors: Solesne Bourguin, Domenico Marinucci, Giovanni Peccati

The aim is to show how it is possible to assess the rate of convergence to Gaussianity for triangular arrays of U-statistics, built from wavelet coefficients evaluated on a homogeneous spherical Poisson field of arbitrary dimension. For this purpose, we exploit the Stein-Malliavin approach introduced in a previous seminal paper; we focus in particular on statistical applications covering evaluation of variance in nonparametric density estimation and Sobolev tests for uniformity.

E595: **Isotropy and homogeneity test on the sphere**
Presenter: **Gilles Fay**, Ecole Centrale Paris, France
Co-authors: Claire Lacour, Thanh Mai Pham Ngoc

A short survey is given of existing nonparametric methods for goodness-of-fit for the distribution of directional data on the sphere (including isotropy tests) and for homogeneity (two-sample) tests for the same kind of data (points on the sphere). We propose new procedures and results based on multiple tests, in a nonparametric framework. Those problems are motivated by applications to astrophysics, especially in the field of ultra-high energy cosmic rays (see e.g. Pierre Auger Observatory experiment), or neutrinos experiments (such as Antares or IceCube).

ES46 Room E1 GENERALIZED ADDITIVE MODELS FOR LOCATION, SCALE AND SHAPE Chair: Thomas Kneib

E1223: **Bayesian structured additive distributional regression**
Presenter: **Stefan Lang**, University of Innsbruck, Germany

The aim is to propose a generic Bayesian framework for inference in distributional regression models in which each parameter of a potentially complex response distribution and not only the mean is related to a structured additive predictor. The latter is composed additively of a variety of different functional effect types such as nonlinear effects, spatial effects, random coefficients, interaction surfaces or other (possibly non-standard) basis function representations. To enforce specific properties of the functional effects such as smoothness, informative multivariate Gaussian priors are assigned to the basis function coefficients. Inference is then based on efficient Markov chain Monte Carlo simulation techniques where a generic procedure makes use of distribution-specific iteratively weighted least squares approximations to the full conditionals. We study properties of the resulting model class and provide detailed guidance on practical aspects of model choice including selecting an appropriate response distribution and predictor specification.

E415: **Boosting beyond the mean and permutation tests**
Presenter: **Andreas Mayr**, Friedrich-Alexander-University Erlangen-Nuremberg, Germany
Co-authors: Matthias Schmid

Measurement errors of clinical devices can be separated into systematic bias and random error. We propose a new method to asses both based on generalized additive models for location, scale and shape (GAMLSS) in combination with permutation tests. More precisely, we extend a recently proposed boosting algorithm for GAMLSS to provide a test procedure to compare measurements from different clinical devices. We carried out a large-scale simulation study to provide empirical evidence that our method is able to identify possible sources for systematic bias as well as random errors under different conditions. Finally, we apply our approach to compare measurements of skin pigmentation from two different devices in an epidemiological study.

E297: **Geoadditive stochastic frontier analysis - a distributional approach**
Presenter: **Nadja Klein**, Georg-August-University Goettingen, Germany
Co-authors: Helmut Herwartz, Thomas Kneib

A flexible Bayesian distributional approach to inefficiency modelling is proposed that offers a structural perspective on eventual determinants of inefficiency in the spirit of generalised additive models for location, scale and shape. Therefore, we explicitly derive the distribution for the compound error term of the normally distributed noise component and the truncated normally distributed inefficiency component. The latter is

augmented multiplicatively to incorporate different effect types of explanatory variables such as spatial dependence structures and location-specific unobserved spatial heterogeneity also for the inefficiency term of the stochastic frontier model. Inference becomes feasible with Markov chain Monte Carlo simulation techniques for which we derive model-specific iteratively weighted least squares proposals. As an empirical illustration we analyse the efficiency of farms located in 62 (out of 71) districts of England and Wales. As a particular result, spatial patterns of inefficiency are clearly supported by model diagnostics. Neglecting common performance patterns of farms located in the same region induces systematic biases to efficiency scores.

E1075: **Simultaneous inference in structured additive conditional copula regression models**
Presenter: **Thomas Kneib**, University of Goettingen, Germany
Co-authors: Nadja Klein

While most regression models focus on explaining aspects of one single response variable alone, interest has recently shifted towards simultaneously studying multiple response variables as well as their dependence structure. A particularly useful tool for pursuing such an analysis are copula-based regression models since they enable the separation of the marginal response distributions and the dependence structure summarised in a specific copula model. However, so far copula-based regression models have mostly been relying on two-step approaches where the marginal distributions are determined first whereas the copula structure is studied in a second step after plugging in the estimated marginal distributions. Moreover, the parameters of the copula are mostly treated as a constant not related to covariates. We therefore propose simultaneous Bayesian inference for both the marginal distributions and the copula using computationally efficient Markov chain Monte Carlo simulation techniques. In addition, we replace the commonly used linear predictor by a generic structured additive predictor and also allow to make the copula parameters covariate-dependent. To facilitate Bayesian inference, we construct proposal densities for a Metropolis Hastings algorithm relying on quadratic approximations to the full conditionals of regression coefficients, thereby avoiding manual tuning and thus providing an attractive adaptive algorithm.

ES49 Room B1 STATISTICAL INFERENCE IN HIGH DIMENSIONS Chair: Soumendra Lahiri

E199: **False discovery rate control for serially dependent test statistics**
Presenter: **Karl Gregory**, Texas A and M University, United States
Co-authors: Soumendra Lahiri, Veerabhadran Baladandayuthapani

Simultaneous testing of many hypotheses requires interpreting the values of many (very often) dependent test statistics. Where the dependence among the test statistics is induced by a few latent factors, a recently proposed method for estimating and removing these effects is available if the covariance matrix of the test statistics is known. The Benjamini-Hochberg procedure may then be applied to the factor-adjusted test statistics. We adapt the factor adjustment methodology to the context of serially dependent test statistics, in which the latent factors may be more readily estimated by assuming a Toeplitz structure for the covariance matrix of the test statistics or by directly defining the factors as harmonic functions. We offer a proof that the removal of factor effects from the test statistics boosts the power of the Benjamini-Hochberg procedure and demonstrate these gains in a simulation study. The gains in power are demonstrated in simulation studies, and the procedure is illustrated in an analysis of real two-sample data in which the equal means hypothesis is tested for a large number of variables admitting an ordering and exhibiting serial dependence. The dependence adjustment results in greater power across a range of false discovery rate thresholds.

E1159: **Boxplot and clustering using minimum volume ellipsoid**
Presenter: **Saswata Sahoo**, Samsung Research Lab, India
Co-authors: Soumendra Lahiri

The minimum volume ellipsoid covering of a collection of points in $\mathbb{R}^p$ is considered and based on it a boxplot of multivariate data is developed. The developed boxplot is used to detect outliers. We propose a clustering algorithm based on minimum volume ellipsoid which can be implemented in a distributed file system to run parallelly on big data.

E1206: **De-sparsified conservative lasso: uniform confidence regions for high dimensional models**
Presenter: **Mehmet Caner**, North Carolina State University, United States
Co-authors: Anders Bredahl Kock

Uniform confidence regions in high dimensional parameter case are developed. It is proposed to use the conservative lasso estimator to form the basis of a new estimator which is formed by getting a "relaxed inverse" of Gram matrix. The first step, conservative lasso is used mainly to prevent false zeros, unnecessary deletions of relevant variables with small coefficients as can be seen in lasso. False deletion of relevant-small variables is an important problem. Conservative lasso is an offshoot of adaptive lasso, but unlike adaptive lasso, it has data dependent weights on all coefficients. Adaptive lasso uses weights only from coefficients that are not deleted from lasso. In the second step we form a new estimator by de-sparsifying the conservative lasso. Since this new estimator has standard least squares limit and entails full set of coefficients, we can use these properties to get individual confidence sets for each coefficient, by inverting a simple t-test. When we derive our results, we provide a new maximal inequality which will provide sharper bounds with independent data. We derive honest confidence intervals for our coefficients. Also it has been shown that confidence intervals contract at an optimal rate. The contributions over de-sparsified lasso are as follows. The data set allows for conditional heteroskedasticity. There is also variance-covariance matrix estimation in the high dimensional case. Also oracle inequality, l1, l2 error bounds are introduced for conservative lasso.

E769: **On the residual empirical process based on the ALASSO in high dimensions and its functional oracle property**
Presenter: **Soumendranath Lahiri**, North Carolina State University, United States
Co-authors: Arindam Chatterjee, Shuva Gupta

Post variable-selection inference is considered in a high dimensional penalized regression model based on the ALASSO method of Zou. It is shown that under suitable sparsity conditions, the residual empirical process based on the ALASSO provides valid inference methodology in very high dimensional regression problems where conventional methods fail. It is also shown that the ALASSO based residual empirical process satisfies a functional oracle property, i.e., in addition to selecting the set of relevant variables with probability tending to one, the ALASSO based residual empirical process converges to the same limiting Gaussian process as the OLS based residual empirical process under the oracle. The functional Oracle property is critically exploited to construct asymptotically valid confidence bands for the error distribution function and prediction intervals for unobserved values of the response variable in the high dimensional set up, where traditional non-penalized methods are known to fail. Simulation results are presented illustrating finite sample performance of the proposed methodology.

ES54 Room N1 WHAT IS NEW IN MODELING AND DESIGN OF EXPERIMENTS Chair: Raul Martin-Martin

E262: **Randomization tests in response-adaptive designs**
Presenter: **Jose Moler**, Universidad Publica de Navarra, Spain
Co-authors: Arkaitz Galbete, Fernando Plo

Due to its particular sampling process, the assumption of a population model in a clinical trial may be inappropriate. A plausible alternative is the use of randomization tests, which do not require any parametric hypothesis. We compare randomization tests for several adaptive randomization procedures using the usual methods for computing such tests: asymptotic normality, exact distributions and Monte Carlo simulation. We focus on some inferential issues when response-adaptive designs are applied.

E161: **Optimal design of experiments for dose-ranging studies based on joint continuous and binary endpoint**
Presenter: **Vladimir Dragalin**, Johnson and Johnson, United States

The overall goal of dose-ranging studies is to establish the existence, nature and extent of dose effect. Recently, a great effort has been made to investigate and develop designs and methods for efficient learning about dose response and more accurate and faster decision making on dose. The vast majority of the methodological developments in this area are devoted to a single endpoint. In the Phase I dose finding studies this is a binary safety endpoint, while in the Phase II studies this is most often a continuous efficacy endpoint. However, in order to evaluate the benefit-risk of a new drug and efficiently determine the recommended dose for the confirmatory stage of drug development, the designs that use both efficacy and safety endpoints are more appropriate. The focus is on adaptive model-based optimal designs of experiments modeling both efficacy and safety endpoints. The proposed dose allocation rule is based on a combination of D- and c-optimality criteria in optimal design of experiments.

E032: **Robust model-based sampling designs**
Presenter: **Douglas Wiens**, University of Alberta, Canada

The addressed problem is to draw a sample, from which to estimate a population total. The data are completely known covariates, to which the unknown response variable is related. Difficulties to be overcome are that the relationship between these variables is only approximate, and perhaps erroneously, specified; similarly the variance/covariance structure of the data must be anticipated at the design stage. We derive minimax designs, and a genetic algorithm for computing the designs.

E109: **Max-min optimal discriminating designs for several statistical models**
Presenter: **Chiara Tommasi**, University of Milan, Italy
Co-authors: Jesus Lopez-Fidalgo, Raul Martin-Martin

Different optimality criteria have been considered for model identification. Most of the proposals assume the Normal distribution for the response variable and thus they are discriminating optimality criteria for regression models. Differently, a max-min approach is followed to discriminate among competing statistical models (i.e. probability distribution functions). More specifically, k different statistical models (plausible for the data) are embedded in a more general model, which includes them as particular cases. The proposed optimal design maximizes the minimum KL-efficiency to discriminate each rival model from the extended one. Some theoretical results are given in order to prove an equivalence theorem and an algorithm is derived from it, which is useful to compute max-min KL-efficiency designs. Finally, two illustrative examples are provided and some future line of research are described.

ES56 Room A1 ROBUST STATISTICAL MODELING Chair: Marco Riani

E217: **Approximate Bayesian computation with robust estimating equations**
Presenter: **Laura Ventura**, University of Padova, Italy
Co-authors: Nicola Sartori, Erlis Ruli

A novel approach is illustrated for developing robust posterior distributions using Approximate Bayesian Computation (ABC) methods with robust estimating equations. This is formally motivated by the use of unbiased estimating functions as automatic informative summary statistics in ABC. Examples with both the Tsallis score and classical robust M-estimating functions are illustrated.

E340: **On robust generalized linear models with a diverging number of parameters**
Presenter: **Marco Avella Medina**, University of Geneva, Switzerland
Co-authors: Elvezio Ronchetti

The asymptotic properties of M-estimators for generalized linear models (GLM) when the number of covariates diverges are studied. More precisely, we establish asymptotic normality for a class of Malow's type M-estimators, prove consistency of the estimated covariance matrix and reexamine likelihood ratio type and Wald type tests based on these estimators. In the derivation of the results the dimension p is allowed to increase with the sampe size n in the order $p = o(n)$. This extends a recent equivalent asymptotic normality result from logistic regression to general maximum likelihood GLM and robust GLM. Furthermore, our analysis greatly improves on existing results for tests in GLM that only allow for $p = o(n^{1/5})$. Finally, we discuss the relationship between our work and some results in high dimensional regression.

E388: **Monitoring robust regression**
Presenter: **Domenico Perrotta**, EC, Joint Research Centre, Italy
Co-authors: Anthony Atkinson, Andrea Cerioli, Marco Riani

How to monitor very robust regression is shown by looking at the behaviour of residuals and test statistics as we smoothly change the robustness of parameter estimation from a breakdown point of 50% to non-robust least squares. The resulting procedure provides insight into the structure of the data including outliers and the presence of more than one population. Monitoring overcomes the hindrances to the routine adoption of robust methods, being informative about the choice between the various robust procedures. Methods tuned to give nominal high efficiency fail with our most complicated example. We find that the most informative analyses come from S estimates combined with Tukey's biweight or with the optimal ρ functions. For our major example with 1,949 observations and 13 explanatory variables, we combine robust S estimation with regression using the forward search, so obtaining an understanding of the importance of individual observations, which is missing from standard robust procedures. We discover that the data come from two different populations. They also contain six outliers.

E809: **A plug-in approach to sparse and robust Principal Component Analysis**
Presenter: **Luca Greco**, University of Sannio - Benevento, Italy
Co-authors: Alessio Farcomeni

A technique for sparse and robust principal component analysis is developed, that builds on the sPCA algorithm. It is structured into two main steps: first, a robust estimate of the covariance matrix is obtained, then this estimate is plugged-in in the sPCA algorithm and sparseness of the robust loadings is induced by using the elastic net. This approach provides an intuitive, general and flexible extension of sparse principal component analysis to the robust setting, allowing for any robust estimate of covariance to be used. This technique is also valid in high dimensions and when the dimensionality exceeds the number of observations by combining the sPCA algorithm with ROBPCA. The methodology provides a valuable alternative to sparse and robust PCA based on projection pursuit.

ES69 Room M1 ANALYSIS OF COMPLEX FUNCTIONAL DATA Chair: Hans-George Mueller

E475: **Detecting changes in mean functions for a functional data sequence**
Presenter: **Jeng-Min Chiou**, Academia Sinica, Taiwan
Co-authors: Yen-Hung Chen, Yu-Ting Chen

Detecting changes in the mean functions of a sequence of functional data has wide applicability. We propose a least squares segmentation approach to detecting multiple changes in mean functions for a functional data sequence, including the total number and the positions of the change-points. The least squares segmentation stage recursively detects multiple change-point candidates for any given number, uniquely identifying the change-point candidates that are statistically consistent with the true ones if they exist. These candidates are assured to be the genuine change-points

by hypotheses testing for statistical significance. As compared with the classical partial-sum-based methods extending from single to multiple change-point detections, the proposed method demonstrates better performance in our numerical results.

E671: **Growth dynamics of adult stem cells**
Presenter: **Tatyana Krivobokova**, Georg-August-Universitaet Goettingen, Germany
Co-authors: Stephan Huckemann, Florian Rehfeldt
The aim is to elucidate the structure and organization of acto-myosin stress fibers of adult stemm cells and their temporal evolution. The data are given as life cell microscopic movies of human mesenchymal stem cells on substrates of different elasticities observed for up to 48 hours under physiological conditions at 37 degrees of Celsius. At the first step, extraction and meaningful description of a filament process of a stem cell has been performed. Subsequently, time evolution of several parameters of the filament process is studied employing FDA techniques, which take an autoregressive noise process into account.

E1125: **Two dimensional representation of functional data for outlier detection**
Presenter: **Ana Arribas-Gil**, Universidad Carlos III de Madrid, Spain
Co-authors: Juan Romo
A two dimensional representation of functional data focusing on shape features is proposed that allows us to characterize shape outliers. Given a sample of curves, shape outliers can be defined as those curves presenting a different shape from the rest of the sample. These outlying curves may not take atypical values and thus be masked among the rest of the curves, which makes them difficult to detect. Our method relies on the combination of two depth measures for functional data whose relationship is investigated. A 2D-plot of one of these measures against the other results in a graphic in which points corresponding to shape outliers lie far away from the majority of the data. The use of this visualization tool, the outliergram, is illustrated through several examples. Moreover, we present an algorithm for shape outlier detection and analyze its performance.

E522: **A domain-selective functional ANCOVA for the study of knee movement**
Presenter: **Alessia Pini**, Politecnico di Milano, Italy
Co-authors: Konrad Abramowicz, Sara Sjostedt de Luna, Lina Schelin, Johan Strandberg, Simone Vantini
A procedure is presented for performing a domain-selective ANCOVA on functional data, and apply it to the analysis of knee movements. We model in a semi-parametric framework the effect of factors and continuous covariates on a functional response, and test for the significance of factors and covariates. All tests are based on the Interval Testing Procedure, i.e., a non-parametric procedure for testing functional data that selects the intervals of the domain where statistically significant effects are detected. To exemplify the procedure, we apply it to human movement data of a study on Anterior Cruciate Ligament (ACL) injury. We compare the knee movement patterns over time during a one-leg hop of three groups of individuals: ACL-injured individuals treated with surgery and physiotherapy; ACL-injured individuals treated with physiotherapy alone; uninjured controls. We include in the ANCOVA model additional information on subjects (e.g., BMI, age, ...) as covariates. The final result of the procedure is a test for the differences between the three groups, and a test for the significance of the effect of each additional covariate. For each test, a selection of the time intervals presenting significant effects is provided.

ES75 Room C1 RECENT ADVANCES IN GENETIC ASSOCIATION STUDIES Chair: Taesung Park

E184: **MOSGWA, a new software tool for model selection in GWAS**
Presenter: **Florian Frommlet**, Medical University Vienna, Austria
Co-authors: Bernhard Bodenstorfer, Malgorzata Bogdan
MOSGWA is a new package for genome-wide association studies. MOSGWA identifies SNPs associated with quantitative traits or disease phenotypes (case-control studies) using generalized linear models. The number and locations of significant associations are estimated based on modified versions of the Bayesian Information Criterion, which allow us to incorporate the prior knowledge on the expected number of causal genes, or alternatively allow for control of the Family Wise Error Rate or False Dicovery Rate. The package takes the data in the PLINK format and returns the estimated generalized regression model, including the identified SNPs and the estimates of their effect size, as well as a mixed model approach to account for population structure. MOSGWA is currently available as a stand alone program, though an R package based on MOSGWA is in preparation.

E687: **Methodological aspects in integromics**
Presenter: **Kristel Van Steen**, University of Liege, Belgium
Co-authors: Kyrylo Bessonov, Ramouna Fouladi
The enthusiasm of having access to vast amounts of information resources comes with a caveat. In contrast to single omics studies, integrated omics studies are extremely challenging. These challenges include protocol development for standardizing data generation and pre-processing or cleansing in integrative analysis contexts, development of computationally efficient analytic tools to extract knowledge from dissimilar data types to answer particular research questions, the establishment of validation and replication procedures, and tools to visualize results. However, from a personalized medicine point of view the anticipated advantages are believed to outweigh any difficulty related to "integromics". The strong interest in the topic has already resulted in the emergence of new integrative cross-disciplinary techniques based on for instance kernel fusion, probabilistic Bayesian networks, correlation networks, statistical data-dimensionality reduction models, and clustering. In this contribution, we will highlight the key steps involved in omics integration efforts and will summarize main analytic paths. We will then zoom in on a novel integrated analysis framework (based on genomic MB-MDR). This framework will be used as a red thread to discuss main issues, pitfalls and merits of integrated analyses. Unprecedented opportunities lie ahead.

E709: **Multi-bin multi-marker method for genome-wide association analysis using clusters of SNPs in linkage disequilibrium**
Presenter: **Yun Joo Yoo**, Seoul National University, Korea, South
Co-authors: Shelley Bull, Sun Ah Kim, Lei Sun
Multi-SNP region-based analysis for genome-wide association study can be performed by global tests constructed from the multiple regression analysis for each region. Quadratic tests such as the Wald test may have large df as many as the number of SNPs, resulting in a power disadvantage in certain situations whereas linear-combination global tests may suffer loss of power when the regression coefficients have opposing effects. The multi-bin linear combination test is a hybrid of quadratic and linear tests constructed by restricting the linear combination within bins of closely correlated SNPs and obtaining the quadratic sum over bins. By using weights constructed from the inverse covariance matrix, a multi-bin linear combination test can be powerful when SNPs that are highly positively correlated with the causal SNP are grouped together. To construct bins of highly correlated SNPs, we develop an efficient clustering algorithm. It is combined with an algorithm to correct the coding of genotypes such that most of the pairwise correlation between SNPs is positive. Using numerical power calculations and simulated data based on data from the 1000 genomes project, we confirm that multi-bin linear combination tests can be powerful over various disease model scenarios.

E805: **A comparative study on multifactor dimensionality reduction methods for the survival phenotype for detecting gene-gene interaction**
Presenter: **Seungyeoun Lee**, Sejong University, Korea, South
Co-authors: Min-Seok Kwon, Yongkang Kim, Taesung Park
Genome-wide association studies have extensively analyzed single SNP effects for a wide variety of common and complex diseases, but a large

portion of the genetic variants left unexplained. It has been proposed that these limitations are due to the analytical strategy that limits analyses to only single SNPs. Recently, it was commonplace to consider the challenges of identifying SNP-SNP interactions associated with the phenotype of interest. The multifactor dimensionality reduction (MDR) method has been proposed to identify multi-SNP effects by reducing a high-dimensional genetic level into one dimensional binary level such as high and low risk groups for the case-control study. Many modifications of MDR have been proposed for extending to the quantitative trait under the generalized linear model and defining a residual score as a new classifier. We compare the modified methods for the survival phenotype, in which the Cox-MDR method uses the martingale residual of a Cox model as a new classifier, and the AFT-MDR method uses the standardized residual of an accelerated failure time model as a new classifier. However, it has been found that the AFT-MDR method is very sensitive to censoring fraction and has low power under the moderate censoring fraction in the simulation study. Therefore, we consider the modifying methods for AFT-MDR to improve the power and compared these methods with a variety of scenarios by simulation study.

ES96 Room D1 SOCIAL NETWORK DATA ANALYSIS Chair: Maria Prosperina Vitale

E608: **Scientific collaboration dynamics in a national scientific system**
Presenter: **Anuska Ferligoj**, University of Ljubljana, Slovenia
Co-authors: Luka Kronegger, Franc Mali, Tom Snijders, Patrick Doreian

The collaboration structures and dynamics of complete co-authorship network of all Slovenian researchers in the period from 1986 to 2010 were studied to identify the main factors driving collaboration and the main differences in the collaboration behaviour across scientific disciplines. The former are on two levels: research policy related external factors and internal factors affecting the motivations of scientists regarding compatibility, cultural proximity, academic excellence, position and status. Two approaches for modeling network dynamics were combined: the small world model and the mechanism of preferential attachment known also as the process of cumulative advantage. One dimension of the small world was measured by its clustering level and the preferential attachment was operationalized through collaboration of researchers within and across disciplines. For studying the effects of clustering level and preferential attachement for new collaboration ties stochastic actor based modelling on network dynamics was used. The complete longitudinal co-authorship networks for all six scientific areas and 72 scientific disciplines for 1996-2010 were analyzed. While the presence of clustering level was confirmed, preferential attachment is far more complex than one autonomous mechanism claim.

E651: **Ianus - a program for temporal network analysis based on calculus of temporal quantities**
Presenter: **Vladimir Batagelj**, University of Ljubljana, Slovenia
Co-authors: Selena Praprotnik

A temporal quantity is varying through time. In some time intervals can be also undefined. An approach to computing with temporal quantities based on semirings is presented. This calculus is used to develop an approach to the analysis of temporal networks with zero latency. The algorithms for computing temporal versions of traditional network analysis measures and structures such as degrees, betweenness, clustering coefficient, weak and strong connectivity, PathFinder skeleton, etc. are proposed. They are available as a Python library TQ (temporal quantities) and through a GUI called Ianus. Its use will be illustrated with analyses of some real-life temporal networks.

E713: **A density-based framework for clustering relational data**
Presenter: **Domenico De Stefano**, Ca Foscari University of Venice, Italy
Co-authors: Giovanna Menardi

In the analysis of social networks, clusters are typically regarded as collections of actors with dense local ties in their neighborhood. This notion of clusters as dense sets is the cornerstone on which a clustering approach for non-relational data hinges: modal clustering associates groups with the regions around the modes of the probability density function underlying the data. The conceptual analogy between the two notions of cluster is exploited to discuss an extension of modal clustering to detection of groups of nodes in social networks. Such extension, albeit appealing, is not obvious because of the different nature of the space where relational and non-relational data are defined. The need of relaxing the notion of density function, in fact, results in a greater flexibility, as different measures of density allow us to embrace a more comprehensive concept of network clustering which can be adapted to different notions of clusters.

E730: **Resampling regression models in the presence of network effects**
Presenter: **Maria Prosperina Vitale**, University of Salerno, Italy
Co-authors: Patrick Doreian, Michele La Rocca, Giovanni C. Porzio

Procedures extending resampling methods for network data are proposed in the literature to take into account interconnected relational data. In this case, given the set of actors in the observed network, a large number of bootstrap samples are drawn with replacement from the actors to derive standard errors for some descriptive statistics. The aim is to evaluate the extent to which existing resampling techniques for network data combined with resampling techniques for regression problems are able to better estimate the standard errors of regression coefficients in Network Autocorrelation Models. The coefficient standard errors reported by the lnam routine in SNA package rely on asymptotic approximations and may not be relevant especially for small sample sizes. This issue is tackled both by means of a simulation study and using a real empirical example.

ES108 Room Q1 BAYESIAN NONPARAMETRIC AND ROBUSTNESS Chair: Judith Rousseau

E351: **Robustifying model components: the choice of priors**
Presenter: **Havard Rue**, NTNU, Norway

The aim is to discuss the choice of prior distribution for the flexibility parameter when we robustify a model component by making it more flexible, like using Student-t instead of a Gaussian, and a spline model instead of a linear effect.

E584: **Asymptotics for Bayesian nonparametric hidden Markov models with finite state space**
Presenter: **Elodie Vernet**, Universite Paris Sud, France

Hidden Markov models (HMMs) have been widely used in diverse fields such as speech recognition, genomics, econometrics. Because parametric modeling of emission distributions may lead to poor results in practice, in particular for clustering purposes, recent interest in using non-parametric HMMs appeared in applications. Yet little thoughts have been given to theory in this framework. We present asymptotic results on Bayesian hidden Markov models with finite state space. Rates of convergence for the posterior are given. Then the last result is applied to discrete emission distribution and location-scale mixtures of Gaussians by Dirichlet process.

E800: **Bayesian empirical likelihood for quantile regression**
Presenter: **Xuming He**, University of Michigan, United States
Co-authors: Yunwen Yang

The Bayesian approach to quantile regression has drawn attention in recent years. Taking the empirical likelihood into a Bayesian framework, we show that the resultant posterior from any fixed prior is asymptotically normal; its mean shrinks toward the true parameter values, and its variance approaches that of the maximum empirical likelihood estimator. A more interesting case can be made for the Bayesian empirical likelihood when informative priors are used to explore commonality across quantiles. Regression quantiles that are computed separately at each percentile level tend to be highly variable in the data sparse areas (e.g., high or low percentile levels). Through empirical likelihood, the proposed method enables

us to explore various forms of commonality across quantiles for efficiency gains. A theoretical framework of shrinking priors is used to better understand the power of the proposed method.

E828: **Robustness of statistical decisions to model approximation**
Presenter: **Chris Holmes**, University of Oxford, United Kingdom

Bayesian decision theory, the procedure of ranking actions according to their expected utility, assumes perfect specification in the joint model consisting of the sampling distribution (likelihood) and prior. If the model is an approximation then so are the predictions of expected utility. We discuss a nonparametric Bayesian framework to assess stability of decisions to model misspecification. In particular we consider the Kullback-Leibler neighbourhood of probability distributions around the approximating model and show how using Monte-Carlo methods one can assess the local-minimax utility within the KL ball, as well as sample the distribution of expected utilities from across all models (distributions) within the ball. This allows for assessment of sensitivity analysis and robustness of conclusions to model assumptions.

ES114 Room H1 LATENT VARIABLES AND FEEDBACK IN GRAPHICAL MARKOV MODELS Chair: Giovanni Marchetti

E694: **On identifiability of causal effects in Bayesian networks**
Presenter: **Elena Stanghellini**, University of Perugia, Italy

Several authors proved the do-calculus to be a complete algorithm for determining, in a non-parametric Bayesian network (BN), which causal effects are identifiable. The theory does not address identifiability when assumptions about the nature of the hidden variables are made. We consider all possible BN's with five discrete variables, one of which is a direct confounder with state space of size r, with r finite. We establish the value k such that the map from the parameters of the joint distribution to the marginal distribution over the observable variables is generically k-to-one. When k is finite, the parameters are generically identifiable. In particular, when $k = r!$, i.e. the only source of non-identifiability is due to swapping of the labels of the confounder, we show that the causal effect is uniquely recovered by application of the do-calculation formula. One may be tempted to conclude that, when k is finite, causal effects can always be uniquely recovered from application of the do-calculus. However, when $k > r!$, instances of models are presented with generically identifiable parameters leading to conflicting results from application of the do-calculus. A Theorem on parameter equivalent models is also presented, and an example coming from applied research is discussed.

E750: **How structural equation models can arise from dynamical systems**
Presenter: **Joris Mooij**, University of Amsterdam, Netherlands
Co-authors: Dominik Janzing, Bernhard Scholkopf

Ordinary differential equations (ODEs) are a very popular and extremely useful mathematical modeling tool in many applied sciences (e.g., biology, chemistry, physics, etc.). ODEs are usually not thought of as causal models. On the other hand, structural equation models (SEMs), a different mathematical modeling framework mostly applied in the social and economical sciences, are usually interpreted in a causal way. We show that these apparently different modeling frameworks are actually quite closely related. The main result is that under certain conditions, equilibrium distributions of ODE systems can be directly mapped onto structural equation models, preserving the right semantics under interventions. This result sheds more light on the nature of SEMs, in particular in cases where causal feedback is present. It also shows that SEMs can be used as an alternative to ODEs when time series data is absent. We discuss some issues that arise when trying to extend this framework to allow for with latent variables.

E784: **High-dimensional causal inference with latent variables**
Presenter: **Marloes Maathuis**, ETH Zurich, Switzerland

The aim is to consider the problem of learning causal information between random variables in directed acyclic graphs when allowing arbitrarily many latent and selection variables. The FCI (Fast Causal Inference) algorithm has been explicitly designed to infer conditional independence and causal information in such settings. However, FCI is computationally infeasible for large graphs. We therefore discuss the new RFCI algorithm, which is much faster than FCI. We show some theoretical results on the algorithms, discuss issues with order-dependence, and also compare the algorithms in simulation studies. Finally, we briefly discuss causal reasoning in the presence of latent variables, using a generalized backdoor criterion.

E812: **Trek separation and latent variable models for multivariate time series**
Presenter: **Michael Eichler**, Maastricht University, Netherlands

In systems that are affected by latent variables conditional independences are often insufficient for inference about the structure of the underlying system. One common example is a system in which four observed variables X_1, X_2, X_3, and X_4 are conditionally independent given a fifth unobserved variable Y. While there are no conditional independences among the observed variables, they must satisfy the so-called tetrad constraints of the form $\rho_{X_iX_j}\rho_{X_kX_l} - \rho_{X_iX_l}\rho_{X_jX_k} = 0$. Recently, such additional non-Markovian constraints were discussed and a characterisation in terms of low-rank conditions on submatrices of the covariance matrix was provided. Graphically these general constraints can be identified by a new separation concept called trek separation. We discuss the extension of the results to the multivariate time series case. Because of the commonly present serial correlation, the results are not directly applicable. For instance, the above tetrad constraints do not hold if the variables $X_1, \ldots, X_4$ and Y (as time series) have non-zero auto-correlation. Graphically, this corresponds to the fact that any instances of the variables $X_1, \ldots, X_4$ cannot be separated by a single instance of Y. As an alternative, we consider mixed graphs in which each node corresponds to a complete time series. Such graphical descriptions for time series have been considered in previous works. We show that trek separation in such graphs corresponds to low-rank conditions on the spectral matrix of the process. In particular, we obtain a spectral version of the above tetrad constraints in terms of spectral coherences. We discuss tests for vanishing tetrad constraints in the frequency domain based on asymptotic results and on bootstrap techniques.

ES121 Room P1 STOCHASTIC MODELS FOR POPULATION DYNAMICS Chair: Manuel Mota

E082: **Stochastic modeling through two-sex branching processes in a random environment**
Presenter: **Manuel Molina**, University of Extremadura, Spain
Co-authors: Shixia Ma, Yongsheng Xing

Branching processes have played a major role in modeling population dynamics. We are especially interested in developing stochastic models to describe the demographic dynamics of populations with sexual reproduction. Significant efforts have been made to develop branching models based on the assumption that mating and reproduction change overtime in a predictable manner. However, in many biological populations with sexual reproduction, mating and reproduction change in a non-predictable manner. Such situations could be mathematically described by considering branching models in a random environment. We study a class of two-sex branching models in a random environment which take such mating and reproductive behaviors into account. For such a class, we investigate several properties and develop some specific software in R. As illustration, we present a simulated example in population dynamics.

E326: **Stochastic modeling of stress erythropoiesis using a two-type age-dependent branching process with immigration**
Presenter: **Ollivier Hyrien**, University of Rochester, United States
Co-authors: Scott Peslak, Nikolay Yanev, James Palis

The aim is to model the dynamics of immature (BFU-E) and mature (CFU-E) erythroid progenitors, which have markedly different kinetics of recovery, following sublethal total body irradiation using a two-type reducible age-dependent branching process with immigration. Properties of

the expectation and variance of the frequencies of both types of progenitors are presented. Their explicit expressions are derived when the process is Markovian, and their asymptotic behavior is identified in the age-dependent (non-Markovian) case. Analysis of experimental data on the kinetics of BFU-E and CFU-E reveals that the probability of self-renewal increases transiently for both cell types following sublethal irradiation. In addition, the probability of self-renewal increased more for CFU-E than for BFU-E. The strategy adopted by the erythroid lineage ensures replenishment of the BFU-E compartment while optimizing the rate of CFU-E recovery. Finally, our analysis also indicates that radiation exposure causes a delay in BFU-E recovery consistent with injury to the hematopoietic stem/progenitor cell compartment that gives rise to BFU-E. Erythroid progenitor self-renewal is thus an integral component of the recovery of the erythron in response to stress.

E382: **Simulation study of X-inactivation in heterozygous females through branching processes**
Presenter: **Manuel Mota**, University of Extremadura, Spain
Co-authors: Miguel Gonzalez

Early in development, one X-chromosome in each cell of the female embryo is inactivated. The knowledge of some issues, as the number of cells at the time of X-inactivation or the probabilities for a single cell to divide or die, can improve our understanding of certain diseases such as cancer or genetic disorders. However, the moment of X-inactivation in humans is difficult to observe directly. A mathematical model was previously developed using branching processes to deal with this heterozygous adult females problem. We generalize that research by considering the possibility of different “splitting rates" in cells for which different X-linked alleles have been inactivated. On the basis of real and/or simulated data, we use simulation-based methods to estimate the number of cells at the time of X-inactivation or the probabilities for a single cell to divide or die.

E797: **Skeletons of near-critical Bienayme-Galton-Watson processes**
Presenter: **Maria Conceicao Serra**, Minho University, Portugal
Co-authors: Serik Sagitov

Near-critical Bienayme-Galton-Watson processes are considered and a flexible way for building the skeletons is proposed. Skeletons of branching processes are usually defined as trees of lineages and characterized by an appropriate feature that ensures future reproduction success. In the supercritical case a natural choice is to look for the lineages that survive forever. In the critical case it was earlier suggested to distinguish the particles with the total number of descendants exceeding a certain threshold. These two definitions lead to asymptotic representations of the skeletons as either pure birth process (in the slightly supercritical case) or critical birth-death processes (in the critical case conditioned on exceeding a high threshold value). We adress the near-critical case. Each vertex of the family tree is independently marked with a small probability. Marking is seen as a rare event that leads to reproductive sucess. The branch connecting the root with a marked vertex is called a marked branch. The marked branches form a subtree of the family tree of the branching process and this will be called a skeleton. Such a skeleton is approximated, when marking is a rare event, by a birth-death process.

ES126 Room L1 RESAMPLING TESTS Chair: Arnold Janssen

E039: **Permutation tests for equivalence and noninferiority**
Presenter: **Fortunato Pesarin**, University of Padova, Italy
Co-authors: Luigi Salmaso

The notion of testing for equivalence and/or non-inferiority of two treatments is widely used in clinical trials, pharmaceutical experiments, bioequivalence and quality control. Traditionally, it is essentially approached within the intersection-union principle (IUP). According to this principle the null hypothesis is stated as the set of effects lying outside a suitably established interval and the alternative as the set of effects lying inside that interval. The non-inferiority problem is tackled by simply setting at the infinity one of the interval limits. The previous solutions are essentially based on likelihood techniques, which in turn are rather difficult to handle especially in multidimensional situations. The main goals are: (i) to go beyond the limitations of likelihood based methods, by considering a nonparametric setting within the permutation frame; (ii) to provide an insight into those problems by observing that they can rationally be approached by two quite different principles: one based on the IUP, the other based on Roy’s union-intersection principle (UIP); (iii) to provide a general multidimensional permutation solution.

E125: **A survey about permutation methods for randomly censored survival data**
Presenter: **Arnold Janssen**, Heinrich-Heine University Duesseldorf, Germany

Consider a two-sample testing problem for an exchangeable null hypothesis H_0 of survival distributions. When the data are randomly right censored then the exchangeability of the data structure may be lost. However, the method of studentized permutation tests can correct the underlying permutation variance of survival test statistics. Within this set up asymptotic semiparametric valid survival tests are constructed. They apply for instance to well-known weighted logrank test statistics.

E166: **A data-dependent multiplier bootstrap applied to transition probability matrices of inhomogeneous Markov processes**
Presenter: **Dennis Dobler**, University of Duesseldorf, Germany
Co-authors: Markus Pauly

The analysis of transition probability matrices of non-homogeneous Markov processes is of great importance (especially in medical applications) and it constantly gives rise to new statistical developments. While observations may be incomplete, e.g. due to random left-truncation and right-censoring, estimation of these matrices is conducted by employing the Aalen-Johansen estimator which is based on counting processes. However, results of weak convergence towards a Gaussian process cannot be utilized straightforwardly since the complicated limiting covariance structure depends on unknown quantities. In order to construct asymptotically valid inference procedures, we insert a set of bootstrap multipliers (from a large class of possible distributions) into a martingale representation of this estimator. A new aspect to this approach is given by the possibility to choose these multipliers dependent on the data, covering, for instance, the Wild bootstrap as well as the Weird bootstrap. In doing so, we gain conditional weak convergence towards a Gaussian process with correct covariance functions resulting in consistent tests and confidence bands. For small samples the performance in the simple competing risks set-up is assessed via simulation studies illustrating the type I error control and analyzing the power of the developed tests and confidence bands for several bootstrap multipliers.

E594: **Permutation and randomization tests of parameters**
Presenter: **Joseph Romano**, Stanford University, United States
Co-authors: EunYi Chung

Given independent samples from P and Q, two-sample permutation tests allow one to construct exact level tests when the null hypothesis is $P = Q$. On the other hand, when comparing or testing particular parameters θ of P and Q such as their means or medians, permutation tests need not control Type 1 error (or Type 3 error), even in arbitrarily large samples. Under very weak assumptions for comparing estimators, we provide a general test procedure whereby the asymptotic validity of the permutation test holds while retaining the *exact* rejection probability α in finite samples when the underlying distributions are identical. A quite general theory is possible based on a coupling construction, as well as a key contiguity argument for the multinomial and multivariate hypergeometric distributions. Time permitting, the results will be extended to multivariate settings and multiple testing, as well as tests of correlation. An important ingredient is the use of a statistic that is asymptotically distribution free. In the multivariate case, this leads to a bootstrap after permutation algorithm in order to achieve the desired results.

EP02 Room First floor Hall POSTER SESSION I Chair: Francisco de Asis Torres-Ruiz

E353: **Recursive filtering in sensor networks with random packet dropouts and random delays in the measurements**
Presenter: **Josefa Linares-Perez**, Universidad de Granada, Spain
Co-authors: Raquel Caballero-Aguila, Aurora Hermoso-Carazo

In recent decades, theoretical and practical research on estimation problems in multi-sensor systems, where sensor networks are used to obtain all the available information on the signal, is gaining increasing attention in multiple disciplines. Sensor networks usually produce both random communication packet losses and random delays, which could degrade the network performance. Furthermore, random parameter measurement matrices are becoming an active research area, since they arise in many situations involving networked systems with stochastic sensor gain degradation, measurement multiplicative noises or missing measurements. The work is concerned with the optimal least-squares linear estimation problem of discrete-time signals from noisy measurements when the measured outputs are perturbed by random parameter matrices and transmitted from different sensors to the processing center. It is assumed that one-step delays and packet dropouts may occur in the transmission and the phenomena of random delays and packet dropouts are modeled employing different sequences of Bernoulli random variables in each sensor. Based on the proposed model and using an innovation approach, a covariance-based linear filter is designed by a recursive algorithm very simple computationally and suitable for online applications. Numerical simulation examples are shown to illustrate the feasibility of the proposed algorithm.

E355: **Estimation using measured outputs with random parameter matrices and one-step correlated random delays**
Presenter: **Aurora Hermoso-Carazo**, Universidad de Granada, Spain
Co-authors: Raquel Caballero-Aguila, Josefa Linares-Perez

In many sensor network applications the measured outputs present uncertainties which cannot be described only by the usual additive disturbances and multiplicative noises must be included in the observation equations. These situations are a special case of systems with random parameter matrices, which have important practical significance and arise in many application areas. Covariance information is used to address the least-squares (LS) linear estimation problem of discrete-time signals for linear stochastic systems with measured outputs perturbed by random parameter matrices. Moreover, it is assumed that the multiple measurements are subject to one-step delays with different delay rates, and that the Bernoulli variables characterizing the measurement delays are correlated at consecutive sampling times. This correlation model covers interesting situations where the common assumption of independent delays is not realistic; for example, networked systems with stand-by sensors for the immediate replacement of a failed unit, thus avoiding the possibility of two successive delayed observations. A recursive LS linear filtering algorithm is derived by an innovation approach and its applicability is illustrated by a simulation example, where a scalar state process is estimated from delayed measurements coming from two sensors in systems with uncertain observations.

E460: **Robust zero-inflated Poisson regression model for quantitative trait loci mapping**
Presenter: **Abbas Moghimbeigi**, Hamadan University of Medical sciences, Iran
Co-authors: Ayoub Maghsoudi, Massoud Saidijam , Ali Reza Soltanian

The count trait may contain extra zeros relative to the Poisson distribution. Maximum likelihood (ML) estimation was used for fitting these models. It is well known that ML estimator is highly sensitive to the presence of outliers. We propose an alternative robust estimation approach, robust expectation-solution (RES) estimation. We propose the interval mapping approach to count trait with excess zeros and extreme values. We analyze the QTL effects through a ZIP regression model with robust estimation. This approach, implemented with the RES algorithm allows for a genome wide scan for the existence of QTL throughout the entire genome. Simulation studies are performed to evaluate the statistical behavior of the approach. An application of the RES estimation is considered to detect associations between the number of cholesterol gallstone and the genotype of markers.

E670: **Bayesian model selection methodology for road safety**
Presenter: **Bahar Dadashova**, Universidad Politecnica de Madrid, Spain
Co-authors: Blanca Arena-Ramirez, Jose Mira-McWilliams, Francisco Aparicio

Road accidents are a very relevant issue in many countries and macroeconomic models are very frequently applied by academia and administrations to reduce their frequency and consequences. The selection of explanatory variables and response transformation parameter within the Bayesian framework for the selection of the set of explanatory variables a TIM (two input model) procedure is proposed. The procedure also uses the DIC and pseudo- R2 goodness of fit criteria. The model to which the methodology is applied is a time series with explanatory variables, structural explanatory model, which is a regression model with Box-Cox transformation for the explanatory variables and response and AR autocorrelation structure for the response. The initial set of 28 explanatory variables can be classified in 9 groups: exposure, economic factors, driver characteristics and surveillance, vehicle characteristics, road infrastructure, legislation, weather conditions and calendar effects. The effects of these factors on the fatal accident frequency in Spanish roads, during 2000-2012, are assessed. In accordance with a previous work, the question of modelling a stochastic trend present in the data is also addressed.

E880: **Regression models for zero-truncated count data**
Presenter: **Ana Maria Martinez-Rodriguez**, University of Jaen, Spain
Co-authors: Antonio Conde-Sanchez, Maria Jose Olmo-Jimenez, Jose Rodriguez-Avi, Antonio Jose Saez-Castillo

The study of count data for which the value zero cannot occur is usually carried out by the truncated Poisson distribution. When there is overdispersion in addition to zero truncation, the truncated binomial negative distribution is used. In the context of regression analysis, zero-truncated Poisson and the zero-truncated binomial negative regression models have been developed. In both models the independent variables are introduced to explain the mean of the respective no-truncated distributions. For these distributions the zero has not been observed and, moreover, the combination of the explicative variables for which the dependent variable became zero is unknown. For this reason, it could be interesting to introduce the regressors in the mean of the truncated distribution, so the usual regression models are modified. Both situations are compared with several examples to see if there are differences in the regression coefficient estimates and in the goodness of fit.

E939: **Sample size issue of the Particle Monte Carlo EM algorithm**
Presenter: **Jing Cheng**, The Chinese University of Hong Kong, China
Co-authors: Ngai-hang Chan

The Particle Monte Carlo EM (or Particle MCEM) algorithm, which constitutes the surrogate to approximate the expectation in E-step with series of weighted particles, is a special variant of the MCEM algorithm. How to choose sample sizes in the iterations is a considerable issue when implementing Particle MCEM algorithm, since the prior rules were customized for MCEM algorithm based on MCMC samples whose properties are quite different from those of weighted particles. New developed Central Limit Theorems for particle filters and smoothers are applied to deal with this issue. An automated criterion based on the confidence interval of relative likelihood is presented to measure the necessity of boosting sample size at each iteration. A simulation study is presented to demonstrate that the proposed criterion could find a better trade-off between accuracy and computational cost compared with the other rules.

E992: **Pairwise comparisons of 3 normal population means by the Bonferroni correction factor**
Presenter: **Ahmet Sezer**, Anadolu University, Turkey
Co-authors: Berna Yazici, Evren Ozkip

In practice, it is often of interest to make inference about the difference between the means of two populations, which are assumed to be normal but with possibly different variances. When the variances are unknown, it is usually referred to as the Behrens-Fisher problem. In the literature associated with the Behrens-Fisher problem, there have been quite a few solutions proposed. The Bonferroni correction is an adjustment made to the significance level when several dependent or independent statistical tests are being performed simultaneously on a single data set. To perform a Bonferroni correction, nominal value is divided by the number of comparisons being made. We focus on the problem of pairwise comparison of three normal populations by the Bonferroni correction factor. We consider four methods to develop pairwise intervals: Welch method, Cochran-Cox method, method proposed by Singh and generalized approach method. An extensive simulation study is conducted to evaluate those methods in respect to their coverage probabilities and expected lengths.

E1053: **Estimation in stochastic systems with packet dropouts and cross-correlated measurement noises**

Presenter: **Maria Jesus Garcia-Ligero**, Universidad de Granada, Spain

Co-authors: Aurora Hermoso-Carazo, Josefa Linares-Perez

Usually in complex systems with multiple sensors, the signal estimation problem is addressed combining the information provided by the different sensors by means of two different techniques: centralized and distributed fusion. The distributed method is considered due to its computational advantages over the centralized one. In sensor network systems the loss of some measurements (packet dropouts) is usually unavoidable due to unreliable features of the networks; in each sensor, the packet dropouts are usually modeled by Bernoulli random variables which take values 0 or 1 depending on whether the measurement is lost or received. Traditionally, the assumption of independent measurement noises is considered; however, in many real situations, this assumption may not be realistic since some kind of correlation may exist. So, our aim is to address the least-squares linear estimation problem for systems with packet dropouts and measurement noises cross-correlated at the same sampling time and at consecutive sampling times in a multisensor environment. Assuming that the covariance functions of the signal and noise processes are known, local estimators from each sensor are derived and the distributed estimator is obtained as a linear combination of them by using least squares criterion.

E1085: **Estimating the parameters of the Richards diffusion process via metaheuristic procedures**

Presenter: **Francisco Torres-Ruiz**, University of Granada, Spain

Co-authors: Patricia Roman-Roman

The problem of maximum likelihood estimation is studied for the parameters of the Richards diffusion process. Since a complex system of equations appears, whose solution cannot be guaranteed via the classic numerical procedures, we suggest the use of metaheuristic optimization algorithms such as Simulated Annealing and Variable Neighborhood Search. One of the fundamental problems for the application of these methods is the space of solutions, since in this case it is continuous and unbounded, which could lead to unnecessary calculation and a long running time for the algorithms. To this end, some strategies are suggested for bounding the space of solutions and a description is provided for the application of the algorithms selected. In the case of the Variable Neighborhood Search algorithm, a hybrid method is proposed in which it is combined with Simulated Annealing. Some examples based on simulated sample paths are developed in order to test the validity of the bounding method for the space of solutions, and a comparison is made between the application of both methods.

E556: **Spatial shift-share analysis using modified AMOEBA procedure**

Presenter: **Luca Grassetti**, University of Udine, Italy

Co-authors: Gian Pietro Zaccomer

Recent studies in the geographical literature have introduced the spatial weights matrix to model the interactions between units for shift-share decomposition. The way in which the weights matrix is defined, goes, however, beyond the scope of these works. The only specification requirement is for the matrix to be row-standardised. Moving one step further, an initial "two-way" spatial decomposition has been proposed for the investigation of occupation growth rates by considering various a priori clusterisations. Previous works within this stream of literature also show that the choice of the weighting system has a substantial effect on the results of spatial shift-share analyses. The aim is to provide an integration of the "two-way" decomposition with a slightly modified AMOEBA clustering procedure (Last Order AMOEBA). The proposed solution allows for a clusterisation maximising the measure of local spatial autocorrelation and at the same time representing the spatial association in a set of units. The obtained results are compared to those found in the literature. In order to do so we rely on data on manufacturing employment from the official Italian Business Statistical Register of the Italian Region of Friuli-Venezia Giulia. All the results are obtained using R.

CSI03 Room Sala Convegni VAR MODELING, COINTEGRATION AND UNCERTAINTY Chair: Peter Winker

C126: **Structural vector autoregressions with smooth transition in variances - the interaction between U.S. monetary policy and the stock market**

Presenter: **Helmut Luetkepohl**, DIW Berlin and Freie Universitaet Berlin, Germany

Co-authors: Aleksei Netsunajev

In structural vector autoregressive analysis identifying the shocks of interest via heteroskedasticity has become a standard tool. Unfortunately, the approaches currently used for modelling heteroskedasticity all have drawbacks. For instance, assuming known dates for variance changes is often unrealistic while more flexible models based on GARCH or Markov switching residuals are difficult to handle from a statistical and computational point of view. Therefore we propose a model based on a smooth change in variance that is flexible as well as relatively easy to estimate. The model is applied to a five-dimensional system of U.S. variables to explore the interaction between monetary policy and the stock market. It is found that previously used conventional identification schemes in this context are rejected by the data if heteroskedasticity is allowed for. Shocks identified via heteroskedasticity have a different economic interpretation than the shocks identified using conventional methods.

C561: **Confidence bands for impulse responses: Bonferroni versus Wald**

Presenter: **Peter Winker**, University of Giessen, Germany

Co-authors: Anna Staszewska-Bystrova, Helmut Lutkepohl

In impulse response analysis estimation uncertainty is typically displayed by constructing bands around estimated impulse response functions. These bands may be based on frequentist or Bayesian methods. If they are based on the joint distribution in the Bayesian framework or the joint asymptotic distribution possibly constructed with bootstrap methods in the frequentist framework often individual confidence intervals or credibility sets are simply connected to obtain the bands. Such bands are known to be too narrow and have a joint confidence content lower than the desired one. If instead the joint distribution of the impulse response coefficients is taken into account and mapped into the band it is shown that such a band is typically rather conservative. It is argued that a smaller band can often be obtained by using the Bonferroni method. While these considerations are equally important for constructing forecast bands, we focus on the case of impulse responses. We also provide some ideas for finetuning of the bands.

C790: **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.

CS54 Room N2 CONTRIBUTIONS ON VOLATILITY MODELS AND ESTIMATION Chair: Helmut Herwartz

C688: **On the link between new stock listings and delistings and average cross-sectional idiosyncratic stock volatility**

Presenter: **Serguey Khovansky**, Northeastern University, United States

Co-authors: Oleksandr Zhylyevskyy

By applying a novel econometric methodology, estimates of average cross-sectional idiosyncratic stock volatility (AIVOL) at the annual frequency in the setting of an ICAPM model are obtained. We find positive contemporaneous associations between these estimates and the numbers of delisted stocks on the U.S. stock exchanges. The discovered relationships tend to be strongly statistically significant and robust to the choice of the calendar month on which an annual period falls. In addition, the relationships are robust to controlling for aggregate financial and economic variables that have been found by the previous literature to explain the dynamics of AIVOL. Our results suggest that the destruction of publicly traded firms (caused by technological and market forces and by entrepreneurial innovations) increases AIVOL among surviving firms and also increases the volatility of financial portfolios. Such an increase may reduce aggregate demand for stocks. Thus, the number of delisted stocks is an important variable for financial investors to account for, as they may experience negative effects of forces that contribute to economic growth in the long run.

C964: **Volatility forecasting with exogenous variables and latent information**

Presenter: **Chainarong Kesamoon**, Autonomous University of Barcelona, Spain

Co-authors: Joan del Castillo

Volatility forecasting is one of the most challenging research areas in financial mathematics. Accurate forecasting of volatility is desirable for risk management and derivatives valuation. Practitioners find difficulties forecasting volatility since it is varying over time, yet unobservable. Two widely used classes of models are GARCH and stochastic volatility (SV) models. GARCH models are simple in estimation while SV models are closely related to option pricing theory. We present a new stochastic volatility model that incorporates exogenous variables in the market such as trading volume, open, close, high and low prices. The stochastic process is specified by normal inverse Gaussian (NIG) distribution and the estimation is done by hierarchical likelihood (h-likelihood) method based on generalized linear model (GLM). The property of this SV model is tractable and the model can be easily extended. The estimation allows us to estimate the unobservable volatility. This latent information is relevant to use in volatility forecasting. We apply this model to forecast volatilities of three currency exchange rates from 2006 to 2010 covering the financial crisis of 2008.

C1059: **Does U.S. monetary policy affect crude oil future price volatility? An empirical investigation**

Presenter: **Antonio Scognamillo**, University of Salerno, Italy

Co-authors: Alessandra Amendola, Vincenzo Candila

Modeling crude oil volatility is of substantial interest for both energy researchers and policy makers. The aim is to investigate the impact of the U.S. Federal Reserve monetary policy on crude oil future price (COFP) volatility. By means of the generalized autoregressive conditional heteroskedasticity-mixed data sampling (GARCH-MIDAS) model, the Effective Federal Fund Rate (EFFR) - as a proxy of the monetary policy - is plugged into the GARCH(1,1) model. Strong evidence of an inverse relation between the EFFR and COFP volatility is found. This means that an expansionary monetary policy is associated with an increase of the COFP volatility. Conjecturing that the unusual behavior of the COFP in 2007-2008 was driven by a monetary policy shock, we test the presence of mildly explosive behavior in the prices. The sup Augmented Dickey-Fuller test (SADF) confirms the presence of a bubble in the COFP series that started in October 2007 and ended in October 2008. We expect that the COFP-EFFR association could be affected by such a bubble. Therefore, we apply the same experimental set-up to two sub-samples - before and after October 2007. Interestingly, the results show that EFFR influence on COFP volatility is greater in the aftermath of the bubble.

C898: **Apophenia: Data under-mining the volatility leverage-effect**
Presenter: **Alessandro Palandri**, University of Warwick, Denmark

The inverse relation between stock returns and their volatility, known as volatility leverage-effect (VLE), is documented as a strikingly robust empirical regularity. It is argued that existing explanations of the phenomenon either suffer from logical inconsistencies or have secondary implications that contradict empirical evidence. Robustness of the empirical findings is re-examined by conducting a thorough investigation of VLE in S&P500 data. Combining misspecification analysis with a novel approach to outlier detection reveals that the VLE relation is indeed very fragile. Implications range from the empirical validity of VLE itself to its use as a moment condition for structural models.

C998: **Quadratic variation estimation of an irregularly observed semimartingale with jumps and noise**
Presenter: **Yuta Koike**, Institute of Statistical Mathematics, Japan

The estimation of the quadratic variation of a semimartingale observed at a high-frequency is considered. High-frequency financial data are often modeled as discrete observations of a semimartingale, and the quadratic variation can be seen as a measure of the volatility of the corresponding asset, so its estimation has attracted attention in financial econometrics recently. The situation where the observation data are contaminated by microstructure noise and the observed semimartingale is allowed to have jumps is considered, and the estimation of the entire quadratic variation is discussed. In such a situation, a pre-averaged version of the realized variance estimator is considered as a natural estimator for the quadratic variation. The asymptotic mixed normality of this estimator is presented under the situation where the observation times show irregularity. In particular, the result shows that some standard methods for constructing confidence intervals of the estimator, which are derived under the regular sampling assumption, are still valid in many irregular sampling settings. This is different from the case of the realized variance estimator in a pure semimartingale setting.

CS12 Room P2 MODELLING UNCERTAINTY IN MACROECONOMICS Chair: Wojtek Charemza

C173: **What's the outlook: Combining narrative, numerical forecasts, and policy uncertainty**
Presenter: **Pierre Siklos**, Wilfrid Laurier University, Canada

Forecasters disagree about the economic outlook for many reasons. Inflation forecast disagreement is investigated for nine economies over the 1999-2013 period, five of which target inflation. The approach developed previously serves as the starting point. However, three important modifications are made. First, borrowing from the ensemble approach to inflation forecasting, quasi-confidence bands for forecast disagreement are generated. Second, one form of 'real time' data that is often critical such as monetary policy announcements by a central bank - generally written statements about the current and anticipated stance of monetary policy - are not directly incorporated in existing measures of forecast disagreement. Using DICTION, the tone of central bank narratives is quantified to determine whether this has an impact on forecast disagreement over a one year ahead horizon. Third, policy uncertainty is considered as a separate determinant of forecast disagreement. A variety of proxies, including the ones introduced by Bloom and Davis (http://www.policyuncertainty.com/), are used.

C191: **Uncertainty and economic activity: A global perspective**
Presenter: **Ambrogio Cesa-Bianchi**, Bank of England, United Kingdom
Co-authors: Hashem Pesaran, Alessandro Rebucci

The 2007-2008 global financial crisis and the subsequent anaemic recovery have rekindled academic interest in quantifying the impact of uncertainty on macroeconomic dynamics based on the premise that uncertainty causes economic activity to slow down and contract. We study the interrelation between financial markets volatility and economic activity assuming that both variables are driven by the same set of unobserved common factors. We further assume that these common factors affect volatility and economic activity with a time lag of at least a quarter. Under these assumptions, we show analytically that volatility is forward looking and that the output equation of a typical VAR estimated in the literature is mis-specified as least squares estimates of this equation are inconsistent. Empirically, we document a statistically significant and economically sizable impact of future output growth on current volatility, and no effect of volatility shocks on business cycles, over and above those driven by the common factors. We interpret this evidence as suggesting that volatility is a symptom rather than a cause of economic instability.

C220: **Fan charts vs. survey forecasts. How similar are they in Poland?**
Presenter: **Halina Kowalczyk**, National Bank of Poland, Poland
Co-authors: Ewa Stanislawska

Recent years were characterized by considerable uncertainty about macroeconomic development. It seems that the uncertainty was so high that it could not be regarded as a statistical one, which means that the analysis of the uncertainty could not be reduced to investigating errors of point forecasts. In this context, NBP SPF - containing questions about possible economic scenarios - gained special importance, as presented in the survey probability distributions which describe the scenario uncertainty. Therefore, it seems important to answer the question whether this different way of defining uncertainty in the NBP SPF is reflected in the features of distributions provided by the survey participants. More specifically, comparison of experts' probabilistic forecasts and fan charts published by the NBP, in terms of calibration and informativeness, is presented. Also, the coordination between the NBP projections and survey forecasts is examined.

C237: **Measuring global and country-specific macroeconomic uncertainty**
Presenter: **Xuguang Sheng**, American University, United States
Co-authors: Ezgi Ozturk

Using monthly survey data from the Consensus Forecast over the period of 1989-2013, new measures of global and country-specific macroeconomic uncertainty are proposed. With a panel of professional forecasters, we define the uncertainty of a typical macroeconomic variable as the sum of disagreement among individual forecasters and the perceived variability of future common shocks. We first estimate the variance of common shocks using GARCH and stochastic volatility models. We then construct an index of country-specific uncertainty by aggregating the uncertainty measures of a range of variables, including output growth, inflation, interest rates, unemployment rate, current account and government budget balance. Finally, we create a measure of global uncertainty by appropriately weighting uncertainty at the country levels. Our new measures prove useful in studying the impact of local and global uncertainty shocks both within and across countries.

C492: **Ex-post inflation forecast uncertainty and skew normal distribution: 'Back from the future' approach**
Presenter: **Svetlana Makarova**, University College London, United Kingdom
Co-authors: Wojciech Charemza, Carlos Diaz

Empirical evaluation of macroeconomic uncertainty and its use for probabilistic forecasting are investigated. A new distribution, in which parameters are interpretable in relation to monetary policy outcomes and actions, is proposed. This distribution is fitted to forecast errors obtained recursively for monthly and annual inflation for 38 countries. It is found that this distribution fits inflation forecast errors better than the two-piece normal distribution, which is often used for inflation forecasting. With the use of decomposition based on this new distribution it is possible to identify the ex-ante component of uncertainty in forecast errors. This component is free of the epistemic (potentially predictable) elements and, therefore, can be used as an approximation to the ex-ante uncertainty. This approach seems to be particularly attractive, as it allows for computation of ex-ante forecast error uncertainty without the use of a panel of professional forecasters' data.

CS23 Room Q2 A NOVEL PERSPECTIVE IN PREDICTIVE MODELLING FOR RISK ANALYSIS Chair: Chiara Gigliarano

C171: **Nonlinear CUB models**
Presenter: **Paola Zuccolotto**, University of Brescia, Italy
Co-authors: Marica Manisera

A general statistical model for ordinal or rating data, which includes some existing approaches as special cases, is proposed. The focus is on the CUB models and a new class of models, called Nonlinear CUB, which generalize CUB. In the framework of the Nonlinear CUB models, it is possible to express a transition probability, i.e. the probability of increasing one rating point at a given step of the decision process. Transition probabilities and the related transition plots are able to describe the state of mind of the respondents about the response scale used to express judgments. Unlike classical CUB, the Nonlinear CUB models are able to model decision processes with non-constant transition probabilities.

C734: **Non-parametric and functional approaches to analysing SME performance**
Presenter: **Jonathan Ansell**, The University of Edinburgh, United Kingdom

In most credit scoring models, underlay assumptions are that variables are normally distributed and linearly correlated with dependent variables. Those assumptions receive furious criticism, especially, during a financial crisis. To explore how regressors influence SMEs default events during 'credit crunch' in practice, we lose underlay assumptions, build generalized additive models (GAM) on a large UK SME data from 2007 to 2010. We use GAM to find whether explanatory variables have nonparametric influence besides linear effect. Segmentations according to firm age are considered. We find that nonparametric effects have different impacts on start-up SMEs and non-start-ups. Other statistical analysis has been given to demonstrate variables' nonparametric effects. Our research shows nonparametric effects are significant in building SME scorecards.

C799: **Gender differences in consumer credit risk**
Presenter: **Galina Andreeva**, University of Edinburgh, United Kingdom
Co-authors: Anna Matuszyk

Gender is not normally used in credit risk assessment due to legal and ethical considerations. Due to the lack of the data, there is limited empirical evidence as to significance of gender in risk modelling. The aim is to analyze a unique proprietary dataset with the aim of establishing the impact of removing gender from a credit scoring model on predictive accuracy and probability of being accepted for credit of men and women. Male and female risk profiles are compared. Observations are made to what extent gender can be proxied by other covariates in the model.

C127: **Recovery rates in consumer lending: Empirical evidence and the model comparison**
Presenter: **Jiri Witzany**, University of Economics in Prague, Czech Republic
Co-authors: Samuel Privara, Marek Kolman

The bank regulation embodied in the Basel II Accord has opened-up a new era in estimating recovery rates or complementary loss given default in the retail lending credit evaluation process. We investigate the properties of survival analysis models applied to recovery rates in order to predict loss given default for retail lending. We compare the results to standard techniques such as linear and logistic regressions and discuss the pros and cons of the respective methods. The study is performed on a real dataset of a major Czech bank.

C1237: **Parameterised intensity models with macroeconomic variables for credit cards**
Presenter: **Jonathan Crook**, University of Edinburgh, United Kingdom
Co-authors: Mindy Leow

Using a large dataset comprising of credit card loans from 2002 to 2010, and incorporating a mix of application variables, behavioural variables and macroeconomic variables, intensity models are used to estimate probabilities of delinquency and default. We estimate these predicted probabilities for each account for each type of transition over the duration time of the loan, and find interesting insights and differences between groups of accounts, as well as over time. By segmenting the dataset by selected application variables, we find that the predicted probabilities of delinquency and default for different groups of accounts have different trends. Next, we conduct two simulations. First, random distributions based on the properties of each observed transition rate are generated and compared against the predicted probabilities to get predicted transitions and events for each account over each duration time step. Based on the simulation and the corresponding predicted events, we are able to calculate distributions for the transitions and losses. Second we select vectors of values of behavioural and macroeconomic variables from historical distributions and predict hazard rates and, indexed on an EAD and LGD of 1.0, infer the value at risk of the 'expected loss' distribution.

CS26 Room Q2 STATISTICAL INFERENCE ON RISK MEASURES Chair: Ghislaine Gayraud

C195: **On the strong approximation of bootstrapped empirical copula processes with applications**
Presenter: **Salim Bouzebda**, Universite de Technologie de Compiegne, France

The aim is to discuss the strong invariance principle for the generalized bootstrapped empirical copula process with the rate of the approximation for multivariate empirical processes. We give a uniform-in-bandwidth consistency's result for kernel-type estimators of copula derivatives, which is of its own interest. Finally, we give some applications.

C376: **Estimating extreme quantiles under random truncation**
Presenter: **Gilles Stupfler**, Aix Marseille University, France
Co-authors: Laurent Gardes

The quantile function, or the VaR risk measure as it is called in financial risk management, is widely used as a means of understanding the tails of a probability distribution. The so-called extreme quantiles, whose exceedance probabilities are small, are particularly relevant in this context. The goal is to provide Weissman-type estimators of the extreme quantiles of a heavy-tailed random variable when it is randomly right-truncated. We state weak consistency and asymptotic normality results for our estimators and we illustrate their finite sample performance on a simulation study.

C414: **Posterior rates of Bayesian VaR and CoVaR**
Presenter: **Ghislaine Gayraud**, University of Technology of Compiegne, France
Co-authors: Mauro Bernardi, Lea Petrella

Within a quantile regression framework, the asymptotic behavior of posterior distributions of two related risk measures is discussed, namely the VaR and the CoVaR. In order to deal with this statistical problem from a Bayesian point of view, we consider the misspecified Asymmetric Laplace distribution, which is known to be a natural way for modelling Bayesian quantile regression. Assuming a nonparametric representation of the VaR and the CoVaR, we then state general conditions under which one is able to derive posterior rates. As an application, we apply our theoretical results to discuss particular appropriate priors.

C465: **Confidence intervals for dynamic Theil entropy of economic systems modeled by Birth-Death processes**
Presenter: **Philippe Regnault**, University of Reims Champagne-Ardenne, France
Co-authors: Guglielmo D'Amico

Theil entropy is a well-known static measure of concentration of - or inequality of repartition of - the wealth shared by a population of economic agents. We are interested in estimating a newly introduced dynamic version of Theil entropy, designed for economic systems involving a large number of independent agents with identical behavior, modelled by a continuous-time birth-death process. Several schemes of observation are

considered, including equidistant and non-equidistant discrete observations. Confidence intervals for dynamic Theil entropy are derived from the asymptotic properties of the studied estimators.

C488: **Randomly censored quantile regression estimation using functional stationary ergodic data**
Presenter: **Mohamed Chaouch**, United Arab Emirates University, United Arab Emirates
Co-authors: Salah Khardani

The asymptotic properties of the conditional quantile function of a randomly censored scalar response given a functional covariate are investigated when the data are sampled from a stationary and ergodic process. We consider a model in which the response variable is censored but not the covariate. Besides the infinite dimensional character of the data, we avoid the widely used strong mixing condition and its variants to measure the dependency and the very involved probabilistic calculations that it implies. Moreover, the mixing properties of a number of well-known processes are still open questions. Indeed, several models are given in literature where mixing properties are still to be verified or even fail to hold for the process they induce. Therefore, we consider, in our setting, the ergodic property to allow the maximum possible generality with regard to the dependence setting. As far as we know, the estimation of conditional quantile combining censored data, ergodic theory and functional data has not been studied in statistical literature. First of all, a kernel-type estimator of the conditional quantile under ergodic and random censorship assumptions is introduced . Then, a strong consistency rate as well as the asymptotic distribution of the estimator are established under mild assumptions. A simulation study that shows the performance of the estimator is also considered. Finally, an application to the peak electricity demand interval prediction with censored smart meter data is provided.

CS27 Room B2 INFERENCE FOR FINANCIAL TIMES SERIES Chair: Giuseppe Storti

C457: **State space modeling of fractional cointegration subspaces**
Presenter: **Roland Weigand**, IAB Nuremberg, Germany

The aim is to investigate a setup for fractionally cointegrated time series which is formulated in terms of latent integrated and short-memory components. It accommodates nonstationary processes with different fractional orders and cointegration of different strengths and is applicable in high-dimensional settings. A convenient parametric treatment is achieved by finite-order ARMA approximations in the state space representation. Monte Carlo simulations reveal good estimation properties for processes of different dimensions. In an application to realized covariance matrices, we find that orthogonal short- and long-memory components provide a reasonable fit and outstanding out-of-sample performance compared to several competitor methods.

C819: **Estimating multivariate symmetric stable distributions with independent components by means of indirect inference**
Presenter: **Giorgio Calzolari**, University of Firenze, Italy
Co-authors: Roxana Halbleib

Financial returns exhibit fat tails, which may be captured by α-stable distributions. We concentrate on multivariate symmetric stable distributions with independent components, which are of particular interest in modeling and estimating factor models with stable distributed idiosyncratic noises. While the simulation from such a distribution is straightforward, the estimation encounters difficulties. These difficulties are overcome by implementing the indirect inference estimation method with multivariate Student's t as auxiliary distribution.

C830: **Least squares estimation for GARCH (1,1) model with heavy tailed errors**
Presenter: **Giuseppe Storti**, University of Salerno, Italy
Co-authors: Arie Preminger

GARCH (1,1) models are widely used for modelling processes with time varying volatility. These include financial time series, which can be particularly heavy tailed. We propose a log-transform-based least squares estimator (LSE) for the GARCH (1,1) model. The asymptotic properties of the LSE are studied under very mild moment conditions for the errors. We establish the consistency, asymptotic normality at the standard convergence rate of *pn* for our estimator. The finite sample properties are assessed by means of an extensive simulation study. Our results show that LSE is more accurate than the quasi-maximum likelihood estimator (QMLE) for heavy tailed errors. Finally, we provide some empirical evidence on two financial time series considering daily and high frequency returns. The results of the empirical analysis suggest that in some settings, depending on the specific measure of volatility adopted, the LSE can allow for more accurate predictions of volatility than the usual Gaussian QMLE.

C663: **On asymptotic theory for ARCH(infinite) models**
Presenter: **Arie Preminger**, Emek Eezreel College, Israel
Co-authors: Christian Hafner

ARCH(infinite) models nest a wide range of ARCH and GARCH models including models with long memory in volatility. The existing literature on such models is quite restrictive in terms of existence of moments. However, the popular FIGARCH, one version of a long memory in volatility model, does not have finite second moments and rarely satisfies the moment conditions of the existing literature. The moment assumptions of a general ARCH(infinite) class of models are considerably weakened, and the theory for consistency and asymptotic normality of the quasi maximum likelihood estimator is developed.

C1154: **A flexible approach to volatility prediction in high-dimensional GARCH models**
Presenter: **Giancarlo Cesale**, University of Salerno, Italy
Co-authors: Giuseppe Storti

Modeling the volatility of large asset portfolios is a common challenge both in theoretical investigation and investment strategy. Due to the high number of parameters, classical MGARCH models describing joint conditional volatilities turn out to be computationally unfeasible in high dimensions. So they often need to be simplified by imposing some constraints on their dynamic structure. Nevertheless this parametric reduction inevitably causes an information loss and some degree of bias in the covariance and correlation estimators. The aim is to present alternative estimators of conditional covariance and correlation which may overcome the problem of information loss entailed by commonly used models (like scalar DCC or BEKK) without incurring in overwhelming computational issues. The approach we follow is based on the combined application of shrinkage approaches and regularization techniques . The effectiveness of the proposed approach is assessed by means of an application to a portfolio of US stocks.

CS63 Room E2 FORECASTING Chair: Pilar Poncela

C268: **The pairwise approach to model and forecast a large set of disaggregates with common trends**
Presenter: **Antoni Espasa**, Carlos III, Spain
Co-authors: Guillermo Carlomagno

The objective is to model and forecast all the components of a macro or business variable. The contribution concerns cases with a large number (hundreds) of components where multivariate approaches are not feasible. We extend in several directions the pairwise approach previously proposed and study its statistical properties. The pairwise approach consists on performing common features tests between the $N(N-1)/2$ pairs of series that exist in a group of N of them. Once this is done, groups of series that share common features can be formed. Next, all the components are forecast using single equation models that include the restrictions derived by the common features. We focus on discovering groups of components

that share single common trends. The asymptotic properties of the procedure are studied analytically. Monte Carlo evidence on the small samples performance is provided and a small samples correction procedure designed. A comparison with a DFM alternative is also carried out, and results indicate that the pairwise approach dominates in many empirically relevant situations. A relevant advantage of the pairwise approach is that it does not need common features to be pervasive. A strategy for dealing with outliers and breaks in the context of the pairwise procedure is designed and its properties studied by Monte Carlo. Results indicate that the treatment of these observations may considerably improve the procedure's performance when series are 'contaminated'.

C746: **Medium term inflation forecasts**
Presenter: **Eva Senra**, Universidad de Alcala, Spain
Co-authors: Pilar Poncela

US and euro area inflation forecasts are considered based on the corresponding Surveys of Professional Forecasters published by the Federal Reserve Bank of Philadelphia and the European Central Bank respectively. Both surveys provide point forecasts as well as probabilities for different outcomes. We aim to use these two information sets to produce a single one year ahead combined inflation forecast. Additionally we shall deal with the alternative measures of inflation uncertainty and their forecasting content.

C789: **The predictive content of co-movement in non-energy commodity price changes**
Presenter: **Lya Sierra**, Pontificia Universidad Javeriana, Colombia
Co-authors: Pilar Poncela, Eva Senra

The predictive content of the co-movement either of a large range of commodities, or the co-movement within a specific category of raw material prices is evaluated. Success is reported in using small scale factor models in forecasting the nominal price of non-energy commodity changes on a monthly basis. Therefore, communalities of commodities in the same category, estimated by the Kalman filter, can be useful for forecasting purposes. Notably, category communalities in oils and protein meals, as well as metals seem to substantially improve the forecasting performance of the random walk model. In contrast, co-movement in extensive data of commodity prices, estimated through Principal Components, has poor predictive power over non-energy commodity prices, compared to the small-scale factors.

C896: **Forecasting volatility measures through dynamic factor models of realized measures**
Presenter: **Pilar Poncela**, Universidad Autonoma de Madrid, Spain
Co-authors: Ana-Maria Fuertes

The aim is to present a comprehensive analysis of the ability of various factor models and related latent variable techniques to provide accurate fore casts of daily quadratic variation by combining a large set of realized volatility estimators. We consider estimators from four distinct families: realized variance, realized range, realized bipower variation and realized kernel, sampled at different intraday frequencies from 1 to 60 minutes. As a byproduct, we compare the forecast performance of combine-then-forecast versus forecast-then-combine approaches.

CS74 Room G2 REGIME CHANGE MODELING IN ECONOMICS AND FINANCE II Chair: Willi Semmler

C919: **International employment and the business cycle: new stylized facts with an application to the great moderation**
Presenter: **Benjamin Kay**, US Treasury, United States
Co-authors: Thomas Daula

The recent global financial crisis suggests the post-1984 Great Moderation has come to an abrupt end. How we obtained nearly 25 years of stability and why it ended are ongoing explorations. Two empirical regularities in US employment complicate the search: i) the decline in the pro-cyclicality of labor productivity with respect to output and labor input and ii) the increase in the volatility of labor input relative to output. We first consider whether these stylized facts are robust to statistical methodology. We find that the widely reported decline in the pro-cyclicality of labor productivity with respect to output is fragile. Using a new international data set on total hours previously constructed, we then consider whether these moments are stylized facts of the global Great Moderation. We document significant international heterogeneity. We then investigate whether the role of labor market frictions in the US as found in a previous work can explain the international results. We conclude that their stylized model does not appear to account for the differences with the US experience and suggest a direction for future research.

C925: **Cointegrated periodically collapsing bubbles in the exchange rate of BRICS countries**
Presenter: **Wilfredo Maldonado**, Catholic University of Brasilia, Brazil
Co-authors: Octavio Tourinho, Jorge de Abreu

The aim is to test the occurrence of rational bubbles in the exchange rate of Brazil, Russia, India, China and South Africa (the 'BRICS') against the US dollar. We assume that the fundamental exchange rate value follows a modified PPP relation, including the interest rate differentials. At each point in time the endogenous probability of collapse of the bubble is a non-linear logistic function of its size. The expected next period bubble size in the collapse regime is a linear function of its current size. The estimation uses the maximum likelihood procedure, and it passes the specification tests. The rational expectations hypothesis in the market for the forward exchange rate is also tested and accepted for 3 of the 5 countries. The hypothesis of two linear regimes (rather than the non-linear regimes we use) is rejected for 3 of the 5 countries. We show that the relative bubbles (with respect to its fundamental value) time series are integrated, and that they pass the Johansen's cointegration test. Finally, we estimate an error correction model to discuss the long term relation between the relative bubbles and the speed of adjustment of each country's relative bubble to shocks to the long term relation between them.

C951: **Non-parametric simulated ML estimation of the heterogeneous agent models**
Presenter: **Jiri Kukacka**, Charles University in Prague, Czech Republic
Co-authors: Jozef Barunik

The aim is to adapt and test a new method for empirical validation of heterogeneous agent models (HAMs) and estimate the very famous Brock and Hommes financial HAM. The methodology of the Non-parametric simulated maximum likelihood estimation is based on the work of Kristensen and Shin. For the HAM the analytical form of the density function does not exist and therefore we are not able to analytically derive the likelihood function to maximize the log-likelihood and estimate the model parameters. However, we are still able to simulate observations from that model to non-parametrically estimate the conditional density by kernel methods. The likelihood function is then constructed and we obtain NPSMLE via maximization over the parameter space. We test our methodology via Monte Carlo simulations, present the small finite sample properties results of the single parameter estimation of the intensity of choice (beta parameter of switching between trading strategies) and show that the consistency and asymptotic efficiency of the estimator is reached. We also present the results of the multiple parameter estimation. We suppose the methodology is likely to appear more general for other HAMs or ABMs in the future.

C1023: **Towards a consumer sentiment channel of monetary policy**
Presenter: **Eric Mayer**, University of Wuerzburg, Germany
Co-authors: Sebastian Debes, Johannes Gareis, Sebastian Rueth

The purpose is to investigate from an empirical and theoretical perspective the role of consumer confidence in the transmission of monetary policy shocks. Standard VAR based analysis suggests that an empirical measure of consumer confidence drops significantly after a monetary tightening and amplifies the impact of monetary policy on aggregate consumption. Using an appropriately augmented behavioral DSGE model, we provide

evidence that a consumer sentiment channel can account for the empirical findings. In an environment of heterogeneous belief, which gives rise to the notion of consumer sentiment, innovations to the Federal Funds rate impact on consumer confidence and thereby the broader economy.

C1069: **Are house price dynamics responsible for unemployment persistence?**
Presenter: **Ekkehard Ernst**, ILO, Switzerland
Co-authors: Faten Saliba

The ambiguous role of house prices and the dynamics of housing wealth for unemployment dynamics are analized. Whereas traditional models see an increase in house prices as a dynamic multiplier that contributes positively to business cycle swings, it is considered an additional transmission mechanism via the competitiveness channel. As house prices rise, wages tend to follow to make up for the loss in real disposable income, which limits employment creation. In addition, with rising house prices, the relative size of the construction sector – a low-productivity industry – tends to increase, lowering aggregate productivity growth, further dampening competitiveness. A stylised dynamic general equilibrium model with unemployment flows is estimated. Introducing different transmission mechanisms through which the housing market influences labour and macroeconomic dynamics, the size and direction of the housing market channel is being analysed. The estimation results show that housing shocks can have long-lasting negative effects on employment even though a housing boom can generate a short-lived stimulus on growth and employment. We also offer some policy advice simulating housing shocks under different types of labour and macro-prudential regulation.

CS92 Room O2 REAL-TIME MODELLING WITH DATA SUBJECT TO DIFFERENT COMPLICATIONS Chair: Peter Zadrozny

C365: **Weighted-covariance factor decomposition of VARMA models for forecasting**
Presenter: **Peter Zadrozny**, Bureau of Labor Statistics, United States
Co-authors: Baoline Chen

The aim is to develop and apply weighted-covariance factor decomposition (WCFD) as a new method for computing factors from time-series observations on a vector of variables for the purpose of forecasting the variables more accurately using parsimonious estimated univariate autoregressive moving-average (UARMA) models of the factors than can be done using estimated vector ARMA (VARMA) models of the vector of variables.

C378: **Micro information dynamics: Decomposing the forecasting power of aggregate indicators**
Presenter: **Klaus Wohlrabe**, Institute for Economic Research, Germany
Co-authors: Georg Strasser

Indicators based on surveys mix responses from a very heterogeneous set of respondents. We explore whether maintaining the full range of respondents is worthwhile, or if focusing on subsets of respondents improves the forecast performance. Further, indicators based on aggregate averages ignore an important part of information. We introduce additional measures derived from micro survey data. These include entropy, standard deviation, disagreement between firms and combinations of them. Furthermore we split the sample into different subcategories. Beside the standard ones (size and branches), we also categorize firms by their answering characteristics (switching behavior or consistent answering). We employ a large micro data set from the Ifo Business Survey. Our target variable is industrial production in Germany. We demonstrate that forecasting accuracy can be improved by using specific information sets. Furthermore, we uncover the sources of forecasting accuracy.

C639: **Outperforming IMF forecasts by the use of leading indicators**
Presenter: **Katja Drechsel**, Halle Institute for Economic Research, Germany
Co-authors: Sebastian Giesen, Axel Lindner

The aim is to analyze the performance of the IMF World Economic Outlook forecasts for world output and the aggregates of both the advanced economies and the emerging and developing economies. With a focus on the forecast for the current and the next year, we examine whether IMF forecasts can be improved by using leading indicators with monthly updates. Using a real-time dataset for GDP and for the indicators we find that some simple single-indicator forecasts on the basis of data that are available at higher frequency can significantly outperform the IMF forecasts if the publication of the Outlook is only a few months old.

C701: **Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections**
Presenter: **Marta Banbura**, European Central Bank, Germany
Co-authors: Domenico Giannone, Michele Lenza

An algorithm is described to compute the distribution of conditional forecasts, i.e. projections of a set of variables of interest on future paths of some other variables, in dynamic systems. The algorithm is based on Kalman filtering methods and is computationally viable for large vector autoregressions (VAR) and large dynamic factor models (DFM). For a quarterly data set of 26 euro area macroeconomic and financial indicators, we show that both approaches deliver similar forecasts and scenario assessments. In addition, conditional forecasts shed light on the stability of the dynamic relationships in the euro area during the recent episodes of financial turmoil and indicate that only a small number of sources drive the bulk of the fluctuations in the euro area economy.

C814: **Forecasting Chinese GDP growth with mixed frequency data: which indicators to look at?**
Presenter: **Heiner Mikosch**, ETH Zurich, Switzerland
Co-authors: Ying Zhang

Building on a mixed data sampling (MIDAS) model, the aim is to evaluate the predictive power of a variety of monthly macroeconomic indicators for forecasting quarterly Chinese GDP growth. We iterate the evaluation over forecast horizons from 370 days to 1 day prior to GDP release and track the release days of the indicators so as to only use information which is actually available at the respective day of forecast. This procedure allows us to detect how useful a specific indicator is at a specific forecast horizon relative to other indicators. Despite being published with an (additional) lag of one month the OECD leading indicator outperforms the leading indicators published by the Conference Board and by Goldman Sachs. Consumer price inflation is especially valuable at forecast horizons of 11 to 7 months. The reserve requirement ratio for small banks proves to be a robust predictor at forecast horizons of 9 to 5 months, whereas the big banks reserve requirement ratio and the prime lending rate have lost their leading properties since 2009. Industrial production can be quite valuable for now- or even forecasting, but only if it is released shortly after the end of a month.

CS94 Room I2 MACROECONOMETRICS Chair: Herman van Dijk

C071: **Forecasting in a data rich environment: Bayesian model averaging and principal components regression**
Presenter: **Rachida Ouysse**, University of New South Wales, Australia

The out-of-sample forecast performance of two alternative methods for dealing with dimensionality is studied, namely, Bayesian model Averaging (BMA) and principal components regression (PCR). We conduct a different out-of-sample investigation in which the predictors are chosen jointly for both output and in using Bayesian variable selection in each out-of-sample recursion using information available at the time of the forecast. This framework implies stochastic nonparametric time-varying reduced form which offers flexibility in capturing structural changes and instabilities of unknown forms. While the competing forecasts are highly correlated, PCR performed marginally better than BMA in terms of mean-squared forecast error. However, this marginal edge in the average global out-of-sample performance hides important changes in the local forecasting power. An analysis of the Theil index indicates that the loss of performance of PCR is due mainly to its exuberant biases in matching the mean of the two series especially the inflation series. BMA forecasts series matches the first and second moments of the GDP and inflation series very

well with practically zero biases and very low volatility. The fluctuation statistic that measures the relative local performance shows that BMA performed consistently better than PCR and the naive random walk benchmark over the period prior to 1985. Thereafter, the performance of both BMA and PCR was relatively modest compared to the naive benchmark.

C470: **Signals from the Government: policy uncertainty and the transmission of fiscal shocks**
Presenter: **Giovanni Callegari**, European Central Bank, United Kingdom
Co-authors: Giovanni Ricco, Jacopo Cimadomo

The influence of fiscal policy uncertainty in the propagation of government spending shocks in the US economy is investigated. We propose a new index to measure fiscal policy uncertainty which relies on the dispersion of government spending forecasts as presented in the Survey of Professional Forecasters (SPF). This new index is solely focused on the uncertainty surrounding federal spending and is immune from the influence of general macroeconomic uncertainty by as much as is possible. Our results indicate that, in times of elevated fiscal policy uncertainty, the output response to policy announcements about future government spending growth is muted. Instead, periods of low policy uncertainty are characterised by a positive and persistent output response to fiscal announcements. Our analysis also shows that the stronger effects of fiscal policy in less uncertain times is mainly the result of agents' tendency to increase investment decisions in these periods, in line with the prediction of the option value theory of Bernanke.

C675: **Dynamic predictive density combinations for large datasets**
Presenter: **Herman van Dijk**, Erasmus University Rotterdam, Netherlands
Co-authors: Roberto Casarin, Stefano Grassi, Francesco Ravazzolo

The purpose is to contribute to the literature on econometric methods for large datasets and introduce a Bayesian procedure for time-varying combinations of large sets of predictive densities that can deal with many data. The approach is based on clustering the set of predictive densities in mutually exclusive subsets and on a hierarchical specification of the combination weights. This modeling strategy reduces the dimension of the latent space and leads to parsimonious combinations. For efficient updating of the dynamic weights of the density combination use is made of parallel sequential Monte Carlo filters. We apply our proposed procedure to a large set of predictive densities using macroeconomic data and find substantial gains in point and density forecasting of US real GDP, GDP deflator inflation, Treasury Bill returns and employment over the last 25 years for all horizons from 1- quarter ahead to 5-quarter ahead. The highest accuracy is achieved when the four series are predicted simultaneously using our combination schemes with within and across cluster weights based on log score learning. Furthermore, we find that weights within clusters are very volatile, indicating that individual model performances are very unstable, strengthening the use of density combinations. Finally, we show that the GPU algorithm can reduce the computing time with respect to CPU version of several multiples of CPU time.

C890: **Output gap and inflation forecasts in a Bayesian dynamic factor model of the euro area**
Presenter: **Marek Jarocinski**, European Central Bank, Germany
Co-authors: Michele Lenza

A Bayesian dynamic factor model of the euro area macroeconomy is estimated. One of the equations of this model is a Phillips curve, that relates inflation to the common factor of real activity indicators (output gap) and to inflation expectations. We study how the choice of variables and prior specification affect the quality of inflation forecasts and the robustness of the output gap estimates in the presence of real-time data. We find that adding more real-activity variables to the bivariate model with just inflation and real output improves the inflation forecasts and the robustness of output gap estimates.

C1093: **Regime switching money market dynamics and its macroeconomic implications**
Presenter: **Stephan Fahr**, European Central Bank, Germany
Co-authors: Carlo Altavilla

The euro area money market during the financial crisis can be presented as a Markov-switching (MS) VAR with two regimes. In abundant liquidity regimes, the Euro OverNight Index Average (Eonia) rate is close to the rate of the deposit facility set by the ECB and remains unaffected by changes in liquidity. Instead, in low liquidity regimes, the Eonia rate is closer to the main policy rate and reacts negatively to increases in excess liquidity. The two regimes can be reproduced using a threshold BVAR using an economic stress index as threshold variable, built from individual financial and money market variables. The MS-VAR and T-BVAR offer a coherent framework of money market dynamics in the euro area and discriminate between shocks to money market rates, the quantity of liquidity and economic stress. These three shocks at the early stages of monetary policy transmission complement the more commonly used shocks to the monetary policy rate and have differing macroeconomic effects, such as inflation, industrial production and credit and compares these to shocks to the monetary policy rate. It can further illustrate how the introduction of OMT or forward guidance transmits to the money market.

CS97 Room C2 RISK MODELING AND APPLICATIONS Chair: Mario Maggi

C399: **Robust smooth transition threshold autoregressive models for electricity prices**
Presenter: **Luigi Grossi**, University of Verona, Italy
Co-authors: Fany Nan

Robust STAR (Smooth Transition Threshold AutoRegressive) processes are studied for modeling and forecasting electricity prices observed on deregulated markets. The main advantage of estimating STAR models is the possibility to capture nonlinearity produced by changes of regimes representing one very well-known stylized fact of electricity prices. However, there are few applications of classical (S)TAR-type models to electricity prices. Another stylized fact of electricity prices is the presence of isolated jumps as a consequence of sudden grid congestions which reflects immediately on prices because of lack of flexibility of the supply and demand curves. This feature must be considered carefully and robust techniques must be applied to avoid that few jumps could dramatically affect parameter estimates. Although many papers have applied quite sophisticated time series models to prices of electricity only few have considered the strong influence of jumps on estimates and the need to move to robust estimators. Among robust techniques for electricity prices, robust STAR models have never been estimated. Classical and robust estimators are applied to estimate parameters of STAR models on Italian electricity price data with and without external regressors, and a comparison of prediction accuracy among the methods is conducted.

C453: **Bayesian robust quantiles for risk management**
Presenter: **Marco Bottone**, Sapienza University of Rome, Italy
Co-authors: Mauro Bernardi, Lea Petrella

Traditional Bayesian quantile regression relies on the Asymmetric Laplace distribution mainly because of its tractability and the interesting property of having the τ-evel quantile as the natural location parameter. However, the Asymmetric Laplace distribution displays medium tails and it is not suitable for data characterized by strong deviations from the Gaussian hypothesis. We propose an extension of the Bayesian Conditional Autoregressive Value-at-Risk (CAViaR) models that accounts also for fat-tails using the Generalized Asymmetric Exponential Power (GAEP) distribution. The GAEP distribution is able to govern the decay of the upper and lower tails separately, making it attractive for robust modeling of conditional quantiles at different confidence levels. To estimate all the model parameters, we propose a new Gibbs sampling algorithm based on the representation of the GAEP distribution as scale mixture of Skew-Uniforms. We analyze some stocks belonging to the US S&P500 index and

compare our proposed model to various alternatives. The model selection method shows that the proposed CaViAR-EP model is strongly preferred over the simple CaViAR and asymmetric GARCH-type models.

C754: **Spatial regression models for UK SMEs**
Presenter: **Raffaella Calabrese**, University of Essex, United Kingdom
Co-authors: Galina Andreeva, Jake Ansell

In order to better understand UK SMEs performance during the financial crisis, an extension from standard scoring models into spatial econometrics techniques is proposed. We find that spatial effects alter the parameter estimates of risk determinants for both start-up SMEs and non-start-ups in different UK regions.

C347: **Credit rating announcements and bond liquidity**
Presenter: **Ana Escribano**, Universidad de Castilla-La Mancha, Spain
Co-authors: Pilar Abad, Antonio Diaz, M. Dolores Robles

The purpose is to investigate liquidity shocks on the United States corporate bond market induced by the information content of the credit rating change announcements and by regulatory constraints. Abnormal trading activity can be triggered by the release of information after any upgrade or downgrade but, even if the event conveys no new information to the market, changes on liquidity can be originated if there is a credit rating change from one main credit rating category to another involving implications on capital requirements for most institutional investors or on bond holding restrictions. We show that: (1) market anticipates rating changes since trading activity slows down days before the event, (2) there is a price pressure and high trading volume during two weeks after, but trading frequency is below normal values, (3) price converges to the fundamentals values and the level of trading activity clearly rises during the second fortnight, (4) among other characteristics of the rating change, the migration between investment- and speculative-grade categories involves further liquidity shocks.

C878: **Risk measurements in decision making with emotional arousal**
Presenter: **Mario Maggi**, University of Pavia, Italy
Co-authors: Caterina Lucarelli, Pierpaolo Uberti

Subjective perception and modeling of risk largely affect Decision Theory. We analyze subjective risk measures by exploiting an interdisciplinary research that merges neurophysiological evidence with standard micro-economic analyzes. We observe a large sample of individuals ($N = 690$) while taking risky choices in a laboratory setting, and measure their emotional arousal, via their change in Skin Conductance Response (SCR), before and after each decision. We assume individuals adapt their decision making based on both return/risk information and their emotional experience. Based on the anticipated SCR to choices, we compare individuals that seem to follow a Pareto-efficiency rule, against individuals that appear sensation-seekers. We test an array of alternative risk measures: 1. variance of the outcomes; 2. Value at Risk; 3. conditional Value at Risk or expected shortfall; 4. maximum drawdown; 5. semi-variance; 6. range of the outcomes; 7. entropic risk measure. We find that agents that optimize their portfolio choices seem to behave according to a variance of returns modeling of risk. Differences in subjective risk measures appear when comparing gender and financial professional experience.

CS98 Room A2 BAYESIAN ECONOMETRICS IN FINANCE Chair: Davide Pettenuzzo

C218: **On loss functions in Value-at-Risk estimation**
Presenter: **Georgios Tsiotas**, University of Crete, Greece

The Value at Risk (VaR) is a risk measure that is widely used by financial institutions to allocate risk. VaR forecast estimation involves the evaluation of conditional quantiles based on the currently available information. Recent advances in VaR evaluation incorporate conditional variance into the quantile estimation, which yields the Conditional Autoregressive VaR (CAViaR) models. Optimal VaR estimates are typically generated using the so-called 'check' loss function. However, issues like VaR's bias estimation and asymmetric financial decision making, based on the sign of the forecast error, can give grounds for the use of further asymmetric loss functions in addition to the check loss for forecasting VaR. We introduce a combination of loss functions and we investigate its effect on forecasting conditional VaR. We illustrate this method using simulated and daily financial return series.

C296: **On the long run variance of stocks**
Presenter: **Carlos Carvalho**, The University of Texas at Austin, United States

The question "what is the variance of long horizons portfolios?" is addressed. The work is directly motivated by "Are Stocks Really Less Volatile in the Long Run". In that work, authors use their "Predictive Systems" framework to challenge the conventional wisdom that stocks are less volatile over long horizons when compared to short horizons. Their conclusion is reached by the incorporation of parameter uncertainty and "imperfect" predictors. Preserving the economic motivation of their approach, we develop parsimonious alternatives to "Predictive Systems" and show that, when compared to the correct benchmark, stocks can still be appealing for long run portfolios. Central to our results is a careful assessment of the priors investors have about key quantities underlying market behavior. In particular we explore the impact of priors that enforce the correlation between expected and realized returns to be negative in a time-varying volatility context.

C433: **Bayesian semiparametric modeling of realized covariance matrices**
Presenter: **John Maheu**, McMaster University, Canada
Co-authors: Xin Jin

A semiparametric model of realized covariances (RCOV) based on an infinite mixture of Wishart type distributions is designed. This extends the work on dynamic specifications for RCOV to allow for very flexible distributions for positive definite matrices. We implement an MCMC method of estimation and provide a detailed comparison of models based on point forecasts and density forecasts. In addition, we show how the model can be incorporated with returns to provide a fully specified model of returns and RCOV. The new model contributes to large gains in density forecasts and improvements in return forecasts.

C767: **Reinforced urn processes for credit risk models**
Presenter: **Stefano Peluso**, Universita della Svizzera Italiana, Switzerland
Co-authors: Antonietta Mira, Pietro Muliere

A Bayesian nonparametric model is proposed to estimate rating migration matrices and default probabilities using the reinforced urn processes (RUP). The estimated default probability is our prior information in a parametric model for the prediction of bankruptcies, with the only assumption of exchangeability within rating classes. The Polya urn construction of the transition matrix justifies a Beta distributed de Finetti measure. Dependence among the processes is introduced through the dependence among the default probabilities, with the Bivariate Beta Distribution and its multivariate generalization.

C310: **Optimal portfolio choice under decision-based model combinations**
Presenter: **Davide Pettenuzzo**, Brandeis University, United States
Co-authors: Francesco Ravazzolo

In the context of stock return predictability and optimal portfolio allocation, model combination methods have been shown to often produce improvements upon the performance of even the best individual models entering the combination. We add to this literature and propose a novel

model combination method that works by combining the predictive densities of the individual models and where the model combination weights are estimated based on how the individual models fare relative to the underlying objective function of the investor. We label our new method "Decision-Based Density Combination", and allow the combination weights to vary over time following a learning mechanism based on the past economic performance of the individual models entering the combination. Empirically, we find that our decision-based density combination method improves both statistical and economic measures of out-of-sample performance, relative to the best individual models entering the combination as well as a variety of existing model combination techniques. Moreover, the benefits from using our novel model combination method appear to increase when we allow for model instabilities in the individual models entering the combination.

CS110 Room F2 CONTRIBUTIONS TO APPLIED ECONOMETRICS II Chair: Michael Smith

C1025: **Incorporating geospatial data in house price indexes: a hedonic imputation approach with splines**
Presenter: **Michael Scholz**, University of Graz, Austria
Co-authors: Robert Hill

The increasing availability of geospatial data (i.e., exact longitudes and latitudes for each house) has the potential to improve the quality of house price indexes. It is not clear though how best to use this information. We show how geospatial data can be included as a nonparametric spline surface in a hedonic model. The hedonic model itself is estimated separately for each period. Price indexes are then computed by inserting the imputed prices of houses obtained from the hedonic model into the Fisher price index formula. Using a data set consisting of 454507 observations for Sydney, Australia over the period 2001-2011 we demonstrate the superiority of a geospatial spline over postcode dummies as a way of controlling for locational effects. While the difference in the resulting price indexes is not that large – since the postcodes in Sydney are quite narrowly defined – we nevertheless find evidence of a slight bias in the postcode based indexes. This can be attributed to systematic changes over time within each postcode in the locational quality of houses sold.

C1167: **Separating the twins**
Presenter: **Makram El-Shagi**, Henan University, China
Co-authors: Puspa D. Amri

In the past decades, twin crises – banking crisis occurring in conjunction with currency turmoil or vice versa – have been increasingly common. However, the literature remains divided over the causal direction. Some argue that banking sector problems lead to currency instability, while others, contend the opposite: devaluation leads to banking crises, by weakening the balance sheet of banks with large foreign currency liabilities. We investigate common factors behind twin crises in a simultaneous-equation probit model that allows to "separate the twins". The work is innovative on two counts. First, we distinguish between "original" banking or currency crisis from those that are created by spillovers. Second, we take into account the possibility of observing a "joint-occurrences" of banking and currency crisis that are independent of each other. Our approach shows that the factors driving currency and banking crisis overlap far less than conventional approaches indicate. In many cases, factors driving currency crisis are mistaken as factors driving banking crisis (and vice versa) when spillovers are not accounted for. Moreover, the institutional factors driving the emergence of currency (banking) crisis, differ surprisingly strongly from the institutional factors that increase the probability of spillovers from banking (currency) to currency (banking) crises.

C277: **150 years of Italian CO_2 emissions and economic growth**
Presenter: **Emilio Zanetti Chini**, University of Rome Tor Vergata, Italy
Co-authors: Barbara Annicchiarico, Anna Rita Bennato

The aim is to examine the relationship between economic growth and carbon dioxide emissions in Italy considering the developments in a 150-year time span. Using several statistical techniques, we find that GDP growth and carbon dioxide emissions are strongly interrelated, with a dramatic change of the elasticity of pollutant emissions with respect to output. Our findings highlight lack of structural change in the reduction of the carbon dioxide, suggesting the difficulties for Italy to meet the emissions targets within the Europe 2020 strategy.

C254: **Higher order beliefs and the dynamics of exchange rates**
Presenter: **Davide Raggi**, University of Bologna, Italy
Co-authors: Francesca Pancotto, Giuseppe Pignataro

The aim is to investigate the role of higher order beliefs in the formation of exchange rates. Our model combines a standard macroeconomic dynamics for the exchange rates with a microeconomic specification of agents' heterogeneity and their interactions. The empirical analysis relies on a state space model estimated through Bayesian methods. We exploit data on macroeconomic fundamentals in a panel of subjective forecasts on the euro/dollar exchange rates. The equilibrium strategy on the optimization process of the predictors shows that higher order beliefs is the relevant factor in performing individual forecasting. Moreover public information, namely past exchange rates and fundamentals, plays a crucial role as a coordination device to generate expectations among agents on the basis of their forecasting abilities.

C1081: **The dynamics of the world cocoa price**
Presenter: **Christopher Gilbert**, University of Trento, Italy

A data series on annual real cocoa prices is constructed extending back to 1850. This series is used to analyse cocoa price dynamics using a small aggregative model based on a vintage production function reflecting the expected yields of cocoa tress of different ages. Cocoa prices are characterized by long (approximately 25 year) cycles which arise out of the combination of the long productive life (more than 40 years) of cocoa trees and the lack of mean or trend reversion in cocoa consumption. Investment, not storage, decisions are the main drivers of these long cycles. Although harvest shocks are the main factor contributing to year-to-year price volatility, shifts in the taste for chocolate and other cocoa products are more important in explaining long term price movements. Price shocks can be large but have low persistence, A forecasting exercise indicates that sustainable cocoa production will very probably involve higher prices than those observed over the recent past.

CS78 Room D2 CONTRIBUTIONS ON TIME SERIES ECONOMETRICS II Chair: Dietmar Bauer

C1204: **Semi-parametric particle filters**
Presenter: **Carles Breto**, Universidad Carlos III de Madrid, Spain

Standard particle filters rely on a parametric specification for the measurement noise. A residual-based approach is presented where the distribution of the measurement noise is estimated non-parametrically via a prediction error decomposition. Such a non-parametric estimate can be used in a second step to maximize the likelihood using plug-and-play approaches, like iterated filtering, which, instead of analytical results, only requires simulation from a numerical model. Such a two-step approach can be used to improve the fit of non-linear non-Gaussian state-space models.

C1097: **Density forecasts and the mixture autoregressive model**
Presenter: **Mary Akinyemi**, University of Manchester, United Kingdom
Co-authors: Georgi Boshnakov

Forecasts play a very significant role in economics, finance, and other fields of application of time series analysis. Density forecasts have become more popular as real life scenarios require not only a forecast estimate but also the uncertainty associated with such a forecast. A density forecast is an estimate of the future probability distribution of a random variable, conditional on the information available at the time of the forecast. Applications of density forecasts span across various industries, including macro economics and finance. A very important area is risk management. 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 density forecasting. The MAR models are able to capture many stylised properties of real data, such as multi-modality, asymmetry and heterogeneity. We evaluate the out-of-sample tail forecast density of some financial time series using MAR models with Gaussian and Student-t innovations. For comparison, we did the same using some popular GARCH models. The MAR(3;2,2,1) models with fat-tailed Student-t distributions delivered the best out-of-sample tail density forecasts. Our study 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 methods.

C1079: **Implications of stochastic singularity in linear multivariate rational expectations models**

Presenter: **Bernd Funovits**, Vienna University of Technology and Vienna Graduate School of Economics, Austria

In general, linear multivariate rational expectations models do not have a unique solution. The aim is to review some procedures for determining whether there exists a solution, whether it is unique, and infers on the dimension of indeterminacy and the number of free parameters in a parametrization thereof. A particular emphasis is given to stochastic singularity, i.e. the case in which the number of outputs is strictly larger than the number of (stochastic) inputs. First, it is shown that assuming stochastic singularity of the exogenous driving process has the same effects as (but is more natural than) assuming that some variables are predetermined, i.e have trivial one-step-ahead prediction error. Second, the dimension of the solution set is in general different from the one derived in the case where the number of outputs and inputs coincide. We derive this result in both the framework of Sims and Lubik/Schorfheide (who impose non-explosiveness conditions) and Broze/Gourieroux/Szafarz (who do not impose non-explosiveness conditions). In this context, previous results are corrected and extended. Last, we note that the latter framework can be adjusted to incorporate non-explosiveness conditions and lends itself to an identifiability analysis of DSGE models.

C1015: **Identifying the best technical trading rules: a .632 bootstrap approach**

Presenter: **Julien Hambuckers**, University of Liege, Belgium

Co-authors: Cedric Heuchenne

The interest is in estimating the out-of-sample predictive ability of a set of trading rules. Usually, this ability is estimated using a rolling-window sample-splitting scheme, true out-of-sample data being rarely available. We argue that this method makes a poor use of the available information and creates data mining possibilities. Instead, we propose an alternative bootstrap approach, based on the .632 bootstrap principle. This method enables us to compute a robust estimator of the out-of-sample predictive ability of a trading rule by generating both in-sample and out-of-sample bootstrap data. These data are later used to measure the performance over a large range of scenarios. We illustrate our methodology on IBM and Microsoft daily stock prices, where we compare 11 trading rules specifications, with a total of 11,668 parametrizations. Our results showed that the out-of-sample performance of all rules decreases strongly compared to the in-sample performance. For the datasets considered, two different filter rule specifications have the highest out-of-sample mean excess returns. However, all tested rules have mean excess returns close or below zero, suggesting that none of these rules can truly generate a profit, nor beat a simple buy-and-hold strategy.

C996: **Conditional heteroskedasticity in stochastic regression**

Presenter: **Tassos Magdalinos**, University of Southampton, United Kingdom

It is well known that, in the context of stochastic regression with stationary regressors, conditional heteroskedasticity in the innovation sequence affects the asymptotic variance of the least squares estimator and distorts the limit distribution of standard self-normalised test statistics (such as the t and Wald statistics): a White type of correction is required for these procedures to yield valid asymptotic inference. This issue does not arise in regression with highly persistent regressors, such as unit root processes. Some new limit theory is developed that clarifies the relationship between persistence and asymptotic contribution of GARCH effects. In particular, it shows that any amount of persistence (even of arbitrarily small order) in the regressor process alleviates asymptotically the effect of conditional heteroskedasticity and results to t and Wald statistics with standard limit distributions. A simple sufficient condition for asymptotic invariance of the Wald statistic to stationary GARCH innovations of arbitrary order is established in terms of the order of magnitude of the standard regression martingale transform. Applications are given in regression with near stationary processes, long memory processes, and in multivariate systems of predictive regressions.

Sunday 7.12.2014 10:55 - 13:00 Parallel Session J – ERCIM

ES04 Room O1 CLASSIFICATION AND DISCRIMINANT PROCEDURES FOR DEPENDENT DATA **Chair: Jose A. Vilar**

E098: **A new dissimilarity-based approach to cluster categorical time series**
Presenter: **Manuel Garcia-Magarinos**, Universidade da Coruna, Spain
Co-authors: Jose Antonio Vilar

A new framework for clustering categorical time series is proposed. In our approach, a dissimilarity-based partitioning method is considered. Measuring the dissimilarity between two categorical time series by assessing both closeness of raw categorical values and proximity between dynamic behaviours is suggested. For the latter, a particular index computing the temporal correlation for categorical-valued sequences is introduced. The dissimilarity measure is then used to perform clustering by considering a modified version of the k-modes algorithm specifically designed to provide us with a better characterization of the clusters. The problem of determining the number of clusters is analysed by comparing a range of procedures. Several graphical devices to interpret and visualize the temporal pattern of each cluster are also provided. The performance of this clustering methodology is studied on different simulated scenarios and its effectiveness is concluded by comparison with alternative approaches. The usefulness is illustrated by analysing navigation patterns of users visiting a specific news web site.

E207: **Long memory time series classification by the AR metric**
Presenter: **Marcella Corduas**, University of Naples Federico II, Italy

The model based approach to time series comparison moves within an inferential framework. It reduces the comparison of time series to the definition of a dissimilarity measure between the underlying generating processes which, after convenient transformations, may be generally assumed to be linear and Gaussian. In particular, the Autoregressive (AR) metric between two invertible ARIMA processes is defined as the Euclidean distance between the sequences of weights characterizing the corresponding pure AR representations. In some fields of applications, such as the hydrological regional classification, long memory components often affect the temporal structure of the observed streamflow series. For this reason, it is important to gain an insight into the use of the AR metric in the presence of long memory. The attention is focused on two aspects of the problem: a) the study of the statistical properties of the distance criterion between ARFIMA processes; b) the consequences that the use of proxy models to describe the generating long memory process produces on time series classification.

E226: **Clustering financial assets in presence of a dominant market**
Presenter: **Edoardo Otranto**, University of Messina, Italy
Co-authors: Romana Gargano

Clustering financial time series is a topic which creates an increasing interest in statistical literature, with several practical aspects largely used in portfolio management, the study of risk, the classification of volatility. The risk is generally linked to the volatility of the asset, but its level of predictability also plays a basic role in investment decisions; for example the movements in the U.S. stock indices could help to forecast the movements of the asset of interest. We propose to classify the markets in terms of unconditional volatility levels, spillover effects from a dominant market and similar dynamics of volatility in comparison to a dominant market. For this purpose we will develop these indicators in a model-based framework, deriving them from the estimated parameters of a recent class of models, the SAMEM (an extended version of the Multiplicative Error Model), and we will use a hierarchical clustering algorithm to classify the assets according to these indicators. This method is applied to a set of European assets, also performing a historical simulation experiment.

E228: **Clustering of time series via non-parametric tail dependence estimation**
Presenter: **Roberta Pappada**, Free University of Bozen-Bolzano, Italy
Co-authors: Fabrizio Durante, Nicola Torelli

Clustering of time series represents an important tool in finance and economics, since practitioners are often interested in identifying similarities in financial assets for portfolio optimization and risk management purposes. To this aim, different copula-based measures of tail dependence can be exploited with the intent to investigate the relationships between time series observations and check whether they cluster together in such a way that may be helpful to risk assessment problems. An approach to the clustering of financial assets according to their asymptotic tail behaviour is presented. The procedure is based on the non-parametric estimation of copula-based lower tail dependence coefficients, used to quantify the extent to which extremely negative events tend to occur simultaneously for pairs of financial assets. The main advantage of the proposed methodology is that it avoids to specify any parametric model assumption on the pairwise dependence structure of the involved time series. The performance of the presented procedure is checked via a simulation study, and a direct application to the analysis of MSCI Developed Market indices is given.

E583: **Sequential clustering procedure for functional data**
Presenter: **Ana Justel**, Universidad Autonoma de Madrid, Spain
Co-authors: Marcela Svarc

Functional data analysis is a very active research area mainly because it has become very easy to collect and store data in continuous time. We introduce a sequential top-down clustering procedure based on information from the function and its derivatives. The new cluster method should extract the relevant clustering information either from f, f' or f''. The algorithm simultaneously conforms and estimates the number of groups. In addition, it provides valuable information about the cluster structure that helps to understand which are the main characteristics to decide that two groups are different. The algorithm repeats two partition steps based on local and global criteria. Simulated and real data sets illustrate the new cluster method perfomance.

ES05 Room H1 ADVANCES IN SPATIO-TEMPORAL ANALYSIS **Chair: Tatiyana Apanasovich**

E272: **Spatiotemporal estimation of long-range dependence ocean surface temperature maps**
Presenter: **Maria Pilar Frias Bustamante**, University of Jaen, Spain
Co-authors: Maria Dolores Ruiz-Medina

Classical statistical Markovian spatiotemporal Gaussian models (e.g., Autoregressive Hilbertian processes) combined with physical models have recently been proposed for the estimation of ocean surface temperature curves. However, the Markovian dependence assumption implies short-range dependence in time and space. A flexible spatiotemporal semiparametric functional model, based on fractional - order pseudodifferential equations with Gaussian innovations, is proposed allowing strong dependence in time and/or space. Regression in the log-wavelet domain is applied for the estimation of the parameters characterizing the fractional order of the pseudodifferential equation. Nonparametric estimation of the probability distribution of the Gaussian driven process is performed in terms of a functional version of Nadaraya-Watson kernel estimator. Finally, ocean surface temperature maps are constructed by plug-in least-square estimation. Results will be illustrated using mean annual daily ocean surface temperature profiles in the area of Hawaii Ocean, latitude-longitude interval [22.7,22.8]x[-158.1,-157.94], during the period 2000-2007.

E482: **Spatial modelling with bivariate splines via SPDE for Gaussian fields**
Presenter: **Xiaoyu Liu**, University College London, United Kingdom
Co-authors: Serge Guillas, Ming-Jun Lai

Gaussian fields (GFs) are frequently used in spatial statistics for their versatility. The associated computational cost can be a bottleneck especially

in the realistic applications. It has been shown that computational efficiency can be gained by doing the computations using Gaussian Markov random fields (GMRFs) as the GFs can be seen as the weak solutions to the corresponding stochastic partial differential equations (SPDEs) using piecewise linear finite elements. We propose a new representation of GFs with bivariate splines in Bernstein form instead of finite elements. This allows an easier implementation of piecewise polynomial representations of various degrees. It leads to GMRFs that can be inferred efficiently and can be easily extended to non-stationary fields. The solutions approximated with higher order bivariate splines converge faster than lower degree representations. Intensive numerical simulations using both real and simulated data also demonstrate the effectiveness, efficiency and flexibility of the bivariate spline representations.

E794: **An asymptotic framework for multivariate tapering**
Presenter: **Reinhard Furrer**, University of Zurich, Switzerland

Parameter estimation for and smoothing or interpolation of spatially or spatio-temporally correlated random processes is used in many areas and often requires the solution of a large linear system based on the covariance matrix of the observations. In recent years the dataset sizes have steadily increased such that straightforward statistical tools are computationally too expensive to be used. In the univariate context, tapering, i.e., creating sparse approximate linear systems, has been shown to be an efficient tool in both the estimation and prediction setting. We present a short review of tapering in the context of temporal and spatial statistics. Key concepts in the framework of estimation and prediction for univariate spatial processes are discussed. A pragmatic asymptotic setting for the extension of tapering to multivariate spatial processes is given and illustrated. We conclude with open problems and challenges of tapering in the context of spatio-temporal modeling.

E131: **Bayesian modeling of skewed spatial distributions**
Presenter: **Anandamayee Majumdar**, Soochow University, United States

There is a need to model asymmetric or skewed spatial process with applications in finance, economics, hydrology and ecology. To this aim, we work in the context of spatially referenced data and use two different approaches. The first is the Double Zero Expectile Normal Processes proposed by Majumdar and Paul, 2014, and the second is a version of the "skewed normal process" previously defined, with closed skew normal multivariate marginals. Both models have useful properties in the sense that they are both ergodic and strongly stationary. We note that the latter process has been proposed, and a computationally intensive EM algorithm has been previously developed for the estimation, but the prediction procedure has not been discussed, and neither has the model been used for any simulated studies or data application. Thus the first attempt to develop a full-fledged Bayesian methodology for the estimation and prediction of this process is proposed. We use a hierarchical, alternative expression of this "skewed normal process", which utilizes a computationally less-intensive estimation and prediction procedure. Under a Bayesian paradigm, we compare the two approaches based on simulated examples, and real applications, to examine their performance, and compare their robustness. We also generalize the DZEXP normal model to multivariate and nonstationary cases.

ES09 Room E1 STATISTICAL INFERENCE FOR NONREGULAR MODELS Chair: Natalia Bochkina

E395: **Adaptive function estimation in nonparametric regression with one-sided errors**
Presenter: **Moritz Jirak**, Humboldt-Universitaet zu Berlin, Germany
Co-authors: Alexander Meister, Markus Reiss

The model of non-regular nonparametric regression is considered where smoothness constraints are imposed on the regression function f and the regression errors are assumed to decay with some sharpness level at their endpoints. The aim is to construct an adaptive estimator for the regression function f. In contrast to the standard model where local averaging is fruitful, the non-regular conditions require a substantial different treatment based on local extreme values. We study this model under the realistic setting in which both the smoothness degree $\beta > 0$ and the sharpness degree $a \in (0,\infty)$ are unknown in advance. We construct adaptation procedures applying a nested version of Lepski's method and the negative Hill estimator which show no loss in the convergence rates with respect to the general L_q-risk and a logarithmic loss with respect to the pointwise risk. Optimality of these rates is proved for $a \in (0,\infty)$. Some numerical simulations and an application to real data are provided.

E063: **Efficient estimation of functionals in one-sided nonparametric models**
Presenter: **Markus Reiss**, Humboldt University, Germany

It is well known that irregular (i.e., not Hellinger-differentiable) parametric models show unusual asymptotic behaviour (e.g., faster rates of convergence, non-Gaussian limit laws). In the case of nonparametric regression $Y_i = f(x_i) + \varepsilon_i$ with irregular errors, e.g. $\varepsilon_i \sim \Gamma(\alpha,\lambda)$, faster rates of convergence can be obtained for tail indices $\alpha < 2$ using a classical local parametric (e.g. polynomial) method. Surprisingly, for estimating linear functionals like $\int_0^1 f(x)dx$ the theory and methods are completely different from both, the parametric irregular and the nonparametric regular theory. In particular, plug-in methods are not rate-optimal and optimal estimators are unbiased even though they involve a bandwidth tuning parameter. In addition, a nonparametric MLE (over Hölder classes) is computationally feasible and reveals fascinating properties in connection with stochastic geometry. Nonparametric sufficiency and completeness arguments can be applied to show its non-asymptotic optimality in a fundamental Poisson point process model. The theory is motivated by volatility estimation based on best bid and ask prices in stock markets.

E332: **Efficiency in irregular models: semiparametric posterior limits**
Presenter: **Bartek Knapik**, VU University Amsterdam, Netherlands
Co-authors: Bas Kleijn

Maximum likelihood estimators and other estimation methods popular in regular models are inefficient in certain irregular situations. For instance, consider estimation of a point of discontinuity of a density from the family of exponential distributions with rate λ shifted by θ, based on n independent observations. The MLE for θ is equal to the minimum of the sample, and its quadratic risk equals $2/\lambda^2$. However, the minimum of the sample decreased by $1/(n\lambda)$ has the quadratic risk equal to $1/\lambda^2$ (the lower bound for the quadratic risk in this setting). Bayes point estimators are known to be asymptotically efficient in regular parametric models, and the proof of this result relies on the study of the limiting behavior of the posterior. We go beyond the parametric setting, and consider semiparametric models exhibiting certain type of irregular likelihood expansion, termed local asymptotic exponentiality. We show that under certain conditions on the model and the prior, the marginal posterior for the parameter of interest converges in total variation to a negative exponential distribution. The general theorem is applied to several examples, including the shifted exponential family example, where the mean of the limiting posterior attains the minimax risk.

E501: **Non-regular general semiparametric models under two-phase sampling**
Presenter: **Takumi Saegusa**, University of Washington, United States
Co-authors: Jon Wellner

Weighted likelihood estimation under two-phase sampling is studied. In this sampling design, a large sample is obtained from a population at the first phase, and partial information of variables are gathered for stratification. At the second phase, a subsample is collected from stratified sampling without replacement to obtain the rest of variables that are difficult or expensive to measure. Examples of this design include stratified case control and case cohort studies. A model of interest is a general semiparametric model where the infinite dimensional parameter can be estimated at a non-regular rate of convergence. Although various theoretical tools have been established for independent and identically distributed data, dependence due to sampling without replacement at the second phase prohibits their direct applications to our biased and dependent data. To address this issue, we extend empirical process theory to the inverse probability weighted empirical process, and establish consistency, asymptotic normality and rate of convergence of the weighted likelihood estimator. We further extend the exchangeably weighted bootstrap empirical process theory

to bootstrapping two-phase sampling. We discuss the important idea of separating randomness from different phases and strata for establishing asymptotic properties. Finite sample performance is illustrated in simulations and real data examples.

E487: **The Bernstein-von Mises theorem and misspecified nonregular models**
Presenter: **Natalia Bochkina**, University of Edinburgh, United Kingdom
Co-authors: Peter Green

The asymptotic behaviour of the posterior distribution is studied in a broad class of statistical models that can be misspecified, that is, the true distribution of the data does not belong to the considered parametric family of models. We focus on the case where the best parametric model that approximates the true distribution is non-regular, that is, where the parameter of the best parametric model is on the boundary of the parameter set. We show that in this case the posterior distribution has not only Gaussian components as in the case of regular misspecified models but also has Gamma distribution components. The form of these components depends on the behaviour of the prior distribution near the boundary, and the rate of convergence is faster than the parametric rate. This behaviour of the posterior distribution will be illustrated on a number of examples.

ES11 Room F1 MODELING DEPENDENCE FOR MULTIVARIATE TIME TO EVENTS DATA Chair: Roel Braekers

E052: **The liability threshold model for censored twin data**
Presenter: **Klaus Holst**, University of Copenhagen, Denmark

Family studies provide an important tool for understanding etiology of diseases, with the key aim of discovering evidence of family aggregation and to determine if such an aggregation can be attributed to genetic components. Heritability and concordance estimates are routinely calculated in twin studies of diseases, as a way of quantifying such a genetic contribution. The endpoint in these studies is typically defined as occurrence of a disease versus death without the disease. However, a large fraction of the subjects may still be alive at the time of follow-up without having experienced the disease thus still being at risk. Ignoring this right-censoring can lead to severely biased estimates. We propose to extend the classical liability threshold model with inverse probability of censoring weighting of complete observations. This leads to a flexible way of modelling twin concordance and obtaining consistent estimates of heritability. We apply the method in simulations and to data from the population based Danish twin cohort where we describe the dependence in prostate cancer occurrence in twins.

E053: **Modelling dependence for bivariate lifetimes**
Presenter: **Thomas Scheike**, University of Copenhagen, Denmark

Data from the Danish twin registry are considered with the aim of studying in detail how lifetimes for twin-pairs are correlated. We consider models where we specify the marginals using a regression structure, here Cox's regression model or the additive hazards model. The best known such model is the Clayton-Oakes model. This model can be extended in several directions. One extension is to allow the dependence to parameter to depend on covariates. Another extension is to allow the dependence to be piecewise constant cross-hazard ratio (CHR) models. We show how both these models can be implemented for large sample data, and suggest a computational solution for obtaining standard errors for such models for large registry data. We also suggest a way of assessing how and if the dependence is changing over time, by considering either truncated or right-censored versions of the data to measure late or early dependence. This can be used for formally testing if the dependence is constant, or decreasing/increasing. The proposed procedures are applied to Danish twin data to describe dependence in the lifetimes of the twins. Here we show that the early deaths are more correlated than the later deaths, and by comparing MZ and DZ associations we suggest that early deaths might be more driven by genetic factors. This conclusion requires models that are able to look at more local dependence measures. We further show that the dependence differs for MZ and DZ twins and appears to be the same for males and females, and that there are indications that the dependence increases over calendar time.

E198: **Additive gamma frailty models with applications to competing risks in related individuals**
Presenter: **Frank Eriksson**, University of Copenhagen, Denmark
Co-authors: Thomas Scheike

Epidemiological studies of related individuals are often complicated by the fact that follow-up on the event type of interest is incomplete due to the occurrence of other events. Within groups of related individuals there will often be information in the occurrence of causes of secondary interest about, e.g. environmental factors. An event from a competing cause can only be treated as independent censoring if the risks are independent and violation of the independent censoring mechanism may produce biased estimates. The aim is to present a class of frailty models with cause-specific proportional hazards for dependent competing events of related individuals and consider a set of frailty structures based on sums of gamma variables. The models offer closed form expressions for the cause specific observed intensities. An estimator is proposed based on the observed intensities and its asymptotic properties as well as a consistent covariance estimator are described. The procedure will be illustrated by a worked example on prostate cancer in twins.

E928: **On identification of a parametric competing risk model**
Presenter: **Petr Volf**, Institute of Information Theory and Automation AS CR, Czech Republic

The problem of modelling and estimation in the competing risk setting is studied. In general, true distributions of risks, i.e. marginal and simultaneous, are not identifiable without an additional information. A simple situation, in which identification is possible, is explored. A parametric model is considered, with assumed risks marginal distributions types, the same is assumed on the copula describing risks dependence, namely that its form is given and its parameter unknown. It is shown that if the likelihood based on the competeing risk data fulfils regularity conditions and therefore the parameters are estimable consistently, then underlying distributions are estimable, too. The result is demonstrated on both simulated data and on a real date case.

E1087: **Using nested Archimedean copula functions to model associations in hierarchically clustered survival data**
Presenter: **Roel Braekers**, Hasselt University, Belgium
Co-authors: Leen Prenen, Luc Duchateau

For Hierarchically clustered multivariate survival data, frailty models with multiple frailty terms are used to analyse and investigate the influence of covariates both on the individual event times as on the association between different event times. However, due to the construction of these models, it is often difficult to clearly interpret the proper influence of these covariates. We introduce several copula models as alternatives for these frailty models. This will allow us to study the influence of the covariates on the individual event times separately from their influence on the association between different event times. We use both Archimedean and nested Archimedean copula functions to model the associations between the event times. Using simulations, we compare the different copula models and investigate the influence of covariates on the individual event times and on the association between the clustered event times. We also illustrate these different models on a practical data set on udder infections in clustered milking cows.

ES15 Room L1 MODEL SPECIFICATION TESTS Chair: M. Dolores Jimenez-Gamero

E100: **Tests of randomness for time series**
Presenter: **Kilani Ghoudi**, United Arab Emirates University, United Arab Emirates
Co-authors: Bruno Remillard

The asymptotic behavior of multivariate empirical processes based on consecutive residuals of stochastic volatility models is first studied. These

processes are then used to develop tests of randomness for the innovations. Though the limiting distributions of the empirical processes depend on the unknown parameters, it is shown that the multipliers technique can be applied in order to estimate the P-values of test statistics. A simulation study is carried out to demonstrate the effectiveness of the proposed tests.

E239: **The Levy characterization and its application in testing for normal distributions**
Presenter: **Elizabeth Gonzalez-Estrada**, Colegio de Postgraduados, Mexico
Co-authors: Jose A. Villasenor Alva

Testing for normality is an important issue in statistical practice since many statistical methods rely on this assumption and it is known that departures from normality might lead to incorrect conclusions. There exist several methods for testing normality, which include tests considering a significance level, called formal statistical tests, as well as graphical tools. A powerful consistent formal test of fit for normal distributions is discussed, which is based on the Levy characterization of this family of distributions. The test statistic is the sample correlation coefficient of sums of pairs of observations from a random sample and normal quantiles. The null distribution of the test statistic is difficult to find analytically; however, critical values can be obtained by simulation since there is no need for estimating any parameters due to the fact that the test statistic is location-scale invariant. The results of a simulation power comparison study, including some existing directed tests for normality, show that the discussed test is comparable to Shapiro-Wilk test and it outperforms the popular Jarque-Bera test against a number of alternatives.

E240: **A non-bootstrap test for gamma distributions**
Presenter: **Jose A Villasenor Alva**, Colegio de Postgraduados, Mexico
Co-authors: Elizabeth Gonzalez-Estrada

The gamma distribution is a versatile probabilistic model having a number of applications in areas such as reliability and survival analysis, hydrology, economics, metagenomics, genetics, statistical analysis of cDNA microarray experiments, etc. Given the importance of the gamma distribution in statistical practice, several authors have addressed the problem of testing if the gamma distribution is the underlying distribution of a random sample; however, due to the fact that this family of distributions is defined in terms of a shape parameter, parametric bootstrap has been used for the implementation of tests. New moments estimators for the parameters of this distribution are proposed, as well as a new goodness of fit test based on the ratio of two variance estimators in which no parametric bootstrap is needed, since the asymptotic null distribution of the test statistic is obtained and the type I error probability can be maximized. A simulation power study indicates that the proposed test compares favorably against a test based on Hoeffding's measure of bivariate independence. An application is provided for testing the gamma distribution hypothesis.

E263: **Testing for validity of a transformation regression model**
Presenter: **Natalie Neumeyer**, University of Hamburg, Germany
Co-authors: Hohsuk Noh, Ingrid Van Keilegom

Before fitting a regression model it is very common in applications to transform the response variable. The purpose of the transformation is, e.g., to reduce skewness or heteroscedasticity, or to induce normality. Often a parametric class of transformations is considered, for instance the class of Box-Cox power transformations, from which an optimal' one should be selected (with a specific purpose in mind). The aim is to develop hypotheses tests to decide whether a suitable parametric class has been chosen. Specifically we consider transformations with the purpose of inducing independence between covariable and regression errors, and we test whether there exists a transformation within the considered class that fulfils this requirement. The test is based on an empirical independence process of covariables and residuals from a regression fit of the transformed response. To this end the transformation parameter is estimated by a profile-likelihood procedure. We consider homoscedastic as well as heteroscedastic nonparametric mean regression models.

E358: **Inference in poverty indices: An empirical processes approach**
Presenter: **Joseph Ngatchou-Wandji**, University of Lorraine, France
Co-authors: Tidiane Seck Cheikh, Lo Gane Samb

Making use of recent results on empirical processes indexed by functions, a functional central limit theorem for a class of poverty indices containing most of those proposed in the literature is established. From this result, we derive a Wald test for comparing the poverty index of a population to a reference poverty index. We study the test statistic under the null hypothesis, under fixed alternatives and under a sequence of local alternatives. A simulation experiment conducted shows that our test performs well for exponential and Pareto data. We finally apply the test to comparing the poverty indices of the ten regions of Senegal to the national poverty index of that country.

ES22 Room N1 DEPENDENCE MODELS AND COPULAS: THEORY III Chair: Fabrizio Durante

E136: **General extremal dependence concepts**
Presenter: **Giovanni Puccetti**, University of Firenze, Italy
Co-authors: Ruodu Wang

The mathematical modeling of the relationship between two or more random variables calls for a notion of dependence. Dependence modeling is a fundamental task in several applied sciences and requires a challenging statistical analysis. While the concept of perfect positive dependence is not ambiguously defined for random vectors of arbitrary dimensions, various notions of perfect negative dependence arise when more than two random variables are involved. We survey all the concepts of negative dependence given in the literature and introduce novel generalised notions of perfect negative dependence which include them all as particular cases. Even if much of the literature on dependence is actually focused on positive dependence, we show that negative dependence plays an equally important role in the solution of many optimisation problems.

E293: **Quantile regression from copula perspective**
Presenter: **Piotr Jaworski**, University of Warsaw, Poland

A regression curve describes a general relationship between an explanatory variable X and a response variable Y. Having observed X, the p-quantile of Y is given by the p-quantile regression function: $Y = m(X) + \varepsilon$, where m is the regression function and an observation error ε is subject to the following almost sure restrictions $\mathbb{P}(\varepsilon \geq 0|X) \geq p$ and $\mathbb{P}(\varepsilon \leq 0|X) \geq 1-p$, $0 < p < 1$. The aim is to characterize the regression function m in terms of the copula C of the joint distribution of X and Y. Furthermore basing on the characterization of copulas by ordinary differential equations a (computationally) simple method of determining value of m at a given point x is proposed. Also the possible links to the newly introduced systemic risk measure CoVaR are discussed.

E324: **Bernstein estimation for a copula derivative with application to conditional distribution and regression functionals**
Presenter: **Jan Swanepoel**, North West University Potchefstroom Campus, South Africa
Co-authors: Noel Veraverbeke, Paul Janssen

Bernstein estimators attracted considerable attention as smooth nonparametric estimators for distribution functions, densities, copulas and copula densities. An analogous result for the first order derivative of a copula function is derived, which leads to Bernstein estimators for a conditional distribution function and its important functionals such as the regression and quantile functions. Results of independent interest have been derived such as the almost sure oscillation behavior of the empirical copula process and a Bahadur type almost sure asymptotic representation for the Bernstein estimator of a regression quantile function.

E356: **Random times to assets' defaults and comparisons among exotic options**
Presenter: **Roy Cerqueti**, University of Macerata, Italy

Co-authors: Fabio Spizzichino

In a financial market, a basket of n different assets is considered and attention is focused on their times to defaults $X_1,\ldots,X_n$. Denote by $X_{(1)},\ldots,X_{(n)}$ the corresponding order statistics and let $\mathcal{P}$ be the family of all the $n!$ permutations of $1,...,n$. The progressive observation of the defaults gives rise to the random permutation $\Pi(\Pi(1),\ldots,\Pi(n))$, defined by $X_{(i)} = X_{\Pi(i)}$, $i \in \{1,\ldots,n\}$. We thus define a specific probability distribution P over $\mathcal{P}$. Different types of exotic options can be defined for the basket. We look at options as at coherent systems in the reliability field, so that comparing the risk profiles of options amounts to comparing systems. First, we point out the role of the distribution P in the problem. Furthermore we discuss relations between P and the form of stochastic dependence among $X_1,\ldots,X_n$. In the case when $X_1,\ldots,X_n$ are exchangeable, the risk profile of the option is strictly related to the concept of signature for the corresponding system. In particular, in the past literature, it has been shown that the signature has a relevant role in the comparison of systems with i.i.d. components. For the sake of our study, such arguments will be extended to the non-exchangeable case.

E377: **A conditional copula-based technique to impute complex dependent data**

Presenter: **F Marta L Di Lascio**, Free University of Bozen-Bolzano, Italy

Co-authors: Simone Giannerini, Alessandra Reale

An imputation method based on conditional copula function is presented. The method, called CoImp, allows us to impute multivariate missing data by accounting for both the (possibly complex) dependence structure and the shape of the margins. The basic idea is to derive the conditional density functions for each incomplete variable given the complete ones through the corresponding conditional copulas and, then, impute missing values by drawing observations from them. The conditional densities can be analytically derived once parametric models for the margins and the copula are specified. To overcome the difficulties arising from the analytical computations and gain in flexibility and applicability, the parametric version of the CoImp can be extended in a semiparametric fashion in that the margins are estimated non-parametrically through local likelihood methods.Hence, we evaluate the performance of the CoImp in both the two versions by assessing its ability to preserve the dependence structure as well as the microdata in different simulated scenarios. In the simulation study are also shown the advantages of the CoImp over classical imputation techniques. Moreover, the method has been implemented in the R software package CoImp.

ES24 Room D1 COMPONENT-BASED MULTILAYER PATH MODELS Chair: Cristian Preda

E230: **Three-mode analysis of multimode covariance matrices**

Presenter: **Pieter Kroonenberg**, Leiden University, Netherlands

Multimode covariance matrices contain the covariances of subject scores on variables for different occasions or conditions. A comparison of three-mode component analysis and three-mode factor analysis applied to such covariance matrices is done. The differences and similarities between the non-stochastic and stochastic approaches are demonstrated by an example. The empirical comparison is facilitated by deriving, as a heuristic device, a statistic based on the maximum likelihood function for three-mode factor analysis and its associated degrees of freedom for the three-mode component models. Furthermore, within the present context a case is made for interpreting the core array as second-order components.

E877: **NIPALS algorithms for Kronecker structured covariance matrices**

Presenter: **Laura Trinchera**, University of Naples Federico II, Italy

Co-authors: Pasquale Dolce, Vincenzo Esposito Vinzi, Alfred O. Hero

Three-way data refers to a set of variables observed on the same units under several occasions. Extension of the two-way component analysis to three-way component analysis has been well studied in the literature. Recently we proposed two alternative algorithms combining Kronecker Principal Component analysis (KPC) and the Non-linear Partial Least Squares (NIPALS) algorithm to provide scalable algorithms for large scale data. Both the algorithms apply directly to the multivariables-multioccasions covariance matrix that is supposed to have a Kronecker structure. The aim of both the algorithms is to find latent components in each of the two matrices defining the Kronecker product. The first algorithm looks for components in the variable and occasion spaces separately, while the second algorithm tries to account for the relationships between the components of the two spaces. We present a Monte Carlo simulation study whose aim is to compare the performance of the two algorithms in terms of convergence, model fit, bias and variability of the estimates. It is a preliminary study for the development of a more flexible approach for analyzing structured three-way data.

E650: **Regularized generalized canonical correlation analysis extended to three way data**

Presenter: **Arthur Tenenhaus**, Supelec, France

Co-authors: Laurent Le Brusquet

On the one hand, multiblock analysis concerns the analysis of data structured in blocks of variables. In that framework, a column partition $\mathbf{X} = [\mathbf{X}_1,\ldots,\mathbf{X}_l,\ldots,\mathbf{X}_J]$ is considered. In this case, each $I \times J_l$ data matrix $\mathbf{X}_l$ is called a block and represents a set of J_l variables observed on I individuals. The number and the nature of the variables usually differ from one block to another but the individuals must be the same across blocks. The main aim is to investigate the relationships between blocks. It appears that Regularized Generalized Canonical Correlation Analysis (RGCCA) is a general framework for multiblock analysis. On the other hand, suppose that measurements are avalaible from, for instance, I individuals on J variables measured at K occasions. Such data are called three-way data because per occasion measurements are available for the same group of individuals on the same set of variables. These three-way data can be collected in an $I \times J \times K$ three-way array $\underline{\mathbf{X}}$, with frontal slices containing the $I \times J$ data matrices for each of the K occasions. Three-way RGCCA (MGCCA) is proposed. MGCCA aims at studying the complex relationships between a set of three-way data table.

E245: **New procedure based on alternative least squares for estimating parameters of a recursive path analysis model**

Presenter: **Mohamed Hanafi**, ENITIAA, Nantes, Frace

Co-authors: Zouhair El Hadri

Given a recursive path analysis model connecting p exogenous variables and q endogenous variables, how to get an optimal estimate of parameter of this model when the well known ULS objective function is considered? Traditionally, the estimation was done by procedures like Fisher's scoring method, method of Davidson-Fletcher-Powell and Gauss-Newton algorithm. All these methods are based on the minimization of an objective function measuring the distance between the empirical correlation matrix and the correlation matrix implied by the model. This latter which is unknown and expressed in terms of the model parameters is not explicitly given and the derivative of the objective function is calculated approximately. A new procedure based on an alternating least squares strategy is proposed. Despite the high nonlinearity of ULS function, the new procedure can be built because new properties allowing separation of the parameters are established making the objective function quadratic in terms of each parameter. Finally, the new and existing procedures are compared thanks to a simulation study.

E882: **Quantile PLSPM versus classical PLSPM: methods, performances and interpretations**

Presenter: **Pasquale Dolce**, University of Naples Federico II, Italy

Co-authors: Cristina Davino, Vincenzo Esposito Vinzi

The aim is to compare one of the most widespread methods used to model relationships between concepts that cannot be directly measured, Partial Least Squares Path Modeling (PLSPM), with the recently introduced Quantile PLSPM. The former is in essence based on an iterative process based on simple and multiple ordinary least squares regressions. The latter has been proposed to enhance PLSPM potentialities exploiting the Quantile regression. Quantile PLSPM is able to distinguish regressor effects on the different parts of the dependent variable distribution. Moreover,

it allows us to handle heteroskedastic relationships among the dependent variables and the regressors, an event frequently occurring when the dependent variable is highly skewed. The potentialities of the Quantile PLSPM are demonstrated in a Monte Carlo study with artificial data where the relationships between explanatory and dependent variables differ across quantiles of the dependent variable distribution. A comparison between this new approach and the classical PLSPM is carried out in terms of performances and result interpretation with respect to different data structures.

ES25 Room B1 MULTIVARIATE EXTREMES Chair: Michael Falk

E331: **Modelling across extremal dependence classes**
Presenter: **Jennifer Wadsworth**, University of Cambridge, United Kingdom
Co-authors: Jonathan Tawn, Anthony Davison, Daniel Elton

A number of different dependence scenarios can arise in the theory of multivariate extremes, entailing careful selection of an appropriate class of models. In the simplest case of bivariate extremes, a dichotomy arises: pairs of variables are either asymptotically dependent or are asymptotically independent. Most available statistical models are suitable for either one case or the other, but not both. The consequence is a stage in the inference that is not accounted for, but which may have large impact upon the subsequent extrapolation. Previous modelling strategies that address this problem are either applicable only on restricted parts of the domain, or appeal to multiple limit theories. We present a unified representation for bivariate extremes that encompasses a wide variety of dependence scenarios, and is applicable when at least one variable is large. The representation motivates a parametric statistical model that is able to capture either dependence class, and model structure therein. We implement a simple version of this model, and show that it offers good estimation capability over a variety of dependence structures.

E350: **Concurrence probabilities for spatial extremes**
Presenter: **Clement Dombry**, Universite de Franche Comte, France
Co-authors: Mathieu Ribatet, Stilian Stoev

While most of the time extreme value analysis focus on the magnitude of extreme events, i.e. how large extreme events are, little interest has been paid to their genesis. In the framework of multivariate or spatial extremes, we define a notion of concurrency of extremes: have these locations been impacted by the same extremal event or was it a consequence of several different ones? Formally, given a sequence $Z_1,\dots,Z_n$ of independent copies of a stochastic process Z, we say that extremes are concurrent at locations $x_1,\dots,x_k$ if there is only one stochastic process Z_ℓ that contributes to the pointwise maxima at those locations, i.e. $\max_{1\le j\le n} Z_i(x_j) = Z_\ell(x_j), \quad j = 1,\dots,k$. The associated concurrence probability $p_n(x_1,\dots,x_k)$ is an interesting object that contains a lot of information about the dependence structure. We investigate its properties such as its limit for large samples ($n\to\infty$), closed formulas for explicit models, connections with other dependance measures such as the extremal coefficient function.

E557: **Markov particle systems and Levy driven Brown-Resnick processes**
Presenter: **Sebastian Engelke**, University of Lausanne and Ecole Polytechnique Federale de Lausanne, Switzerland
Co-authors: Jevgenijs Ivanovs, Zakhar Kabluchko, Martin Schlather

The Brown-Resnick process $\xi(t)$, $t\in\mathbb{R}$, is defined as $\xi(t) = \max_{i\in\mathbb{N}} U_i + B_i(t) - |t|/2$, where $\{U_i : i\in\mathbb{N}\}$ is a Poisson point process on $\mathbb{R}$ with intensity $e^{-u}du$ and B_i, $i\in\mathbb{N}$, are independent copies of a two-sided standard Brownian motion. The process ξ is the pointwise maximum of the stationary system of particles U_i moving along the Brownian trajectories and, as such, it is max-stable and stationary. More general particle systems are considered where the particles follow Markovian trajectories. In particular, when the Markov process is a Levy process, the pointwise maxima yield a new class of stationary, max-stable processes that can be used in extreme value theory to model temporally dependent extreme events. This is important if the data exhibits jumps or other non-Gaussian features. The probabilistic properties, simulation methods and statistical inference for these Levy driven Brown-Resnick processes are investigated.

E708: **On generalized max-linear models in max-stable random fields**
Presenter: **Maximilian Zott**, University of Wuerzburg, Germany
Co-authors: Michael Falk

In practice, it is not possible to observe a whole max-stable random field. Therefore, a way how to generate a max-stable random field in $C^k[0,1]$ from a max-stable random vector in $\mathbb{R}^d$ by establishing so-called *generalized max-linear models* is proposed. This allows the approximation of a max-stable random field by the observation of some finite dimensional marginal distribution. Promising results have been established in the case $k = 1$ in a previous paper. However, the extension to higher dimensions is not straightforward since the natural order of the index space is not available for $k \ge 2$.

E740: **Computational methods for multivariate extreme value distributions**
Presenter: **John Nolan**, American University, United States
Co-authors: Anne-Laure Fougeres, Cecile Mercadier

Multivariate extreme value distributions are characterized by an angular measure on the unit simplex. We examine the case where the angular measure is a piecewise polynomial. Computational techniques are described that allow one to work with this class of models in dimension greater than two.

ES82 Room A1 ROBUSTNESS IN SURVEY SAMPLING Chair: Anne Ruiz-Gazen

E278: **Robust estimation of a mean electricity curve by sampling**
Presenter: **Anne de Moliner**, EDF RD, France
Co-authors: Herve Cardot, Camelia Goga

Many studies carried out in the French electricity company EDF are based on the mean electricity consumption curve of groups of customers. For example, these aggregated load curves are used in order to manage the electricity distribution network or in marketing studies to analyse the impact of equipment or customer behaviors on the electricity consumption. These load curves are estimated using samples of thousands of curves measured at a half hourly time step and collected according to a sampling design. Due to the skewness of the distribution of electricity consumption, these samples often contain outliers which can have a huge impact on the estimation especially as far as small areas are concerned. Three robust estimators based on the concept of conditional bias are proposed to address this problem: the first two methods consist in using functional outliers detection methods (based on Modified Band Depth or on functional Principal Component Analysis) to identify the outliers in order to reduce their influence and the last one consists in using robust estimators developed in the context of survey sampling independently for each instant. These three methods are compared to each other on real datasets.

E503: **On the external consistency property for domain estimation in the presence of influential units**
Presenter: **David Haziza**, University of Montreal, Canada
Co-authors: Cyril Favre-Martinoz, Jean-Francois Beaumont

Influential units are those which make classical estimators (e.g., the Horvitz-Thompson estimator or calibration estimators) very unstable. The problem of influential units is particularly important in business surveys, which collect economic variables, whose distribution is highly skewed (heavy right tail). In order to reduce the impact of influential units, a number of methods can be used, including winsorization. In practice, estimates are required not only at the overall population level but also for subgroups called domain. In this context, an important property is that of external

consistency: the sum of domain estimates must be equal to the estimate at the overall level. Methods for ensuring this property in the presence of influential units will be discussed. The results of an empirical study will be shown.

E611: **Robust inference for GLM and GLMM in finite population**
Presenter: **Cyril Favre-Martinoz**, Crest-Ensai-Irmar, France
Co-authors: David Haziza, Nikos Tzavidis
Influential units occur frequently in surveys, especially in the context of business surveys that collect economic variables whose distributions are highly skewed. A unit is said to be influential when its inclusion or exclusion from the sample has an important impact on the magnitude of survey statistics. In a model based framework, a robust version of the Best Linear Unbiased Predictor (BLUP) has already been proposed. In pratice, it is not unusual to deal with some categorical variables or count data, which require generalized linear model. That's why we constructed a robust estimation of the BLUP in GLM using conditional biais as a mesure of influence. These results will serve as a starting point for small area estimation and comparaison with robustified maximum likelihood estimators and Bias-corrected estimators proposed in the literature. The generalisation for a generalized linear mixed model will be also discused.

E613: **Robust small area estimation under semi-parametric models**
Presenter: **Laura Dumitrescu**, ENSAI, France
In small area estimation auxiliary information is used through linking models and the classical approach is to use linear models and to obtain the empirical best linear unbiased predictors (EBLUPs) of small area means. However, the efficiency of the estimators depends on the correct specification of the linking models so the performance of the EBLUPs can be severely affected by the presence of representative outliers in the data or departures from the normal distribution assumption of the random effects. In recent years several robust techniques have been developed for unit-level linear mixed models. Furthermore, the assumption of linearity can be restrictive in practice and extensions of the EBLUPs of small area means under semi-parametric models have been introduced. We discuss recent developments of robust methods which are adequate for these more general models and present robust empirical best linear unbiased predictors of the small area means which are shown to be more efficient than the corresponding EBLUPs under several types of contamination.

E631: **Bias-corrected outlier robust small domain predictors under spatial correlation**
Presenter: **Timo Schmid**, Freie Universitaet Berlin, Germany
Co-authors: Nikos Tzavidis, Ralf Munnich, Ray Chambers
The presence of outliers in conjunction with small domain-specific sample sizes can cause problems when using model-based estimation of domain finite population parameters. In those cases, outlier robust estimation methods offer protection against the influence of outliers. Conventionally, model-based domain predictors are built by using random effects models that assume independent domain random effects. With real data, however, spatial dependencies in the data often occur. Outlier robust domain prediction in the presence of spatial correlation was recently considered in literature. Conventionally, outlier robust predictors are plug-in leading in some cases to bias. Recent literature in outlier robust domain estimation has introduced bias-corrected predictors that employ both local and global bias-correction terms. We propose local and global bias-corrected small domain predictors under spatial correlation. In the case of a global bias-adjustment the correction term includes the potential spatial impact from other domains on the domain of interest. Inference is performed by using parametric bootstrap. Simulations based on outlier-contaminated data indicate that the proposed predictors can lead to more efficient results. We conclude with an application using business survey data for estimating average labor costs in Italian provinces.

ES83 Room M1 DATA ON MANIFOLDS AND MANIFOLD DATA Chair: Laura Sangalli

E554: **Statistical methods for projective shape analysis**
Presenter: **John Kent**, University of Leeds, United Kingdom
Projective shape is the information in a geometric object that is invariant under projective transformations. The main application is to camera images, where a projective transformation corresponds to a change in the camera view. The simplest example is the cross ratio for a set of four collinear points. There are two standard ways to represent the projective shape of an object in m-dimensional space: (a) frames, in which $m+2$ landmarks are mapped to fixed positions, and (b) Tyler standardization which does not depend on the ordering of the landmarks. The strengths and weaknesses of each approach will be illustrated on a variety of examples.

E717: **Representations and analysis of manifold data**
Presenter: **Adrian Bowman**, University of Glasgow, UK
Co-authors: Liberty Vittert, Stanislav Katina
Imaging systems based on stereophotogrammetry or laser scanning produce representations of surfaces (two-dimensional manifolds in three-dimensional space) through triangulations of a large number of estimated surface points. Traditional forms of analysis have been based on point locations (landmarks) which are defined anatomically or mathematically. A much richer representation is available through curves which track the ridges and valleys of the surface and subsequently by the relatively smooth surface patches which lie between these curves. Methods for identifying these curves and the resulting surface representation, based on shape index, curvature, smoothing techniques and bending energy, will be described. These ideas will be applied to surface images of the human face. Issues of sexual dimorphism (differences in shapes between males and females) and change in shape with age will be explored. The information contained in landmarks, curves and surfaces will be compared. Data on a variety of patient groups, including those who have undergone facial surgery will also be examined. The assessment of asymmetry is often of particular interest. The issues involved in comparing facial shapes in general, at both the individual and the group level, will also be considered.

E837: **Geodesic Gaussian processes for the reconstruction of 3D noisy surfaces**
Presenter: **Bianca Maria Colosimo**, Politecnico di Milano, Italy
Co-authors: Enrique del Castillo
Reconstruction of a surface starting from 3D noisy data points is a frequent problem in reverse engineering and quality inspection and monitoring. To this aim, a model describing the surface height as a function of the other two coordinates, i.e., $z = z(x,y)$ is frequently assumed in the literature. In this framework, Euclidean Gaussian Processes is considered as viable modeling tool when the surface has no clear parametric pattern. The aim is to discuss possible drawbacks of this approach and present a different method -called "Geodesic Gaussian Process" (GGP) - which consists of firstly re-parameterizing the (x,y,z) Cartesian coordinates as a function of a different set of (u,v) coordinates using algorithms from the manifold learning and computer graphics literature. These new coordinates are then used as locations in a spatial Gaussian process model that considers correlations between two points on the surface a function of their geodesic distance on the surface, rather than a function of their Euclidean distances over the *xy* plane. The effectiveness of the proposed procedure is shown with reference to simulated and real surfaces.

E933: **Principal nested spheres analysis of molecular dynamics data**
Presenter: **Kwang-Rae Kim**, University of Nottingham, United Kingdom
Co-authors: Ian Dryden, Huiling Le
Molecular dynamics simulations produce huge datasets of temporal sequences of molecules. It is of interest to summarize the shape evolution of the molecules in a succinct, low-dimensional representation. However, Euclidean techniques such as principal components analysis (PCA) can be problematic as the data may lie far from a flat manifold. Principal nested spheres can lead to striking insights which may be missed when

using PCA. We provide some fast fitting algorithms and apply the methodology to a large set of 100 runs of 3D protein simulations, investigating biochemical functions in applications in Pharmaceutical Sciences.

E965: **Metric-based registration and analysis of objects (curves, surfaces, and images)**
Presenter: **Anuj Srivastava**, Florida State University, United States

The problem of statistical analysis and modeling of objects using image data is encountered in many applications. While past works in shape analysis mostly relied on discrete, landmark-based representations of objects, the recent trend has been to study parametrized objects - curves and surfaces. The key novelty is the one that performs registration of objects, i.e. optimal matching of points across objects, while comparing their shapes. This framework, termed elastic shape analysis, is enabled by: (1) a proper choice of metric that satisfies certain isometry properties, and (2) a square-root based transform that maps this elastic metric into the common Euclidean metric. This framework has been developed for shape analysis of curves in Euclidean spaces and surfaces, and has been demonstrated in many applications with real data. This success has motivated renewed interest in registration problems in signal and image analysis, where current methods have certain important shortcomings including the lack of "inverse consistency" - the registration of A to B is not consistent with the registration of B to A. I will describe the use of shape-type metrics that can not only help register images, with a lot more desired properties than the current methods, but also be used for further statistical analyses such as PCA or regression modeling. The biggest advantage comes from the fact that registration and comparison of images are all performed in a unified way, under the same metric. I will demonstrate these ideas using examples of functional data analysis (biosignals, mass-spectrometry, etc) and image analysis (MRI slices, handwritten characters, etc).

ES87 Room I1 DATA SCIENCE AND THEORY Chair: M. Brigida Ferraro

E841: **Fuzzy probabilistic-distance clustering of time and numerical series modeled by penalized spline**
Presenter: **Antonio D'Ambrosio**, University of Naples Federico II, Italy
Co-authors: Carmela Iorio, Gianluca Frasso, Robarta Siciliano

Clustering of numerical series (time series, longitudinal data, ...) has application in various domains. We propose a fuzzy method for clustering data series modeled by weighted penalized spline. Raw data are simultaneously analyzed through weighted penalized splines allowing for an efficient separation of the signal from the measurement noise. The probabilistic clustering framework makes it possible to jointly update the weighting system involved in the definition of the within cluster trends and the probabilistic allocation of the observed series to the clusters. We evaluate the performances of our procedure dealing with simulated and real data using different distance measures. We discuss the applicability of the proposed clustering method to analyze general classes of time series emphasizing its fuzzy nature.

E891: **A semi-parametric model for clustering hospitals by similarity in patients' outcome: a study of cesarean sections rates in Sardinia**
Presenter: **Massimo Cannas**, University of Cagliari, Italy
Co-authors: Claudio Conversano, Francesco Mola, Emiliano Sironi

The differences in rates of cesarean deliveries across hospitals cause concern and debate about the appropriateness of many interventions. This problem is particularly relevant in Italy, which has one of the highest intervention rates in Europe. Using data from hospital abstracts on deliveries that occurred in Sardinia over a two-year period, we fit a semiparametric logistic regression model with a Dirichlet process prior for the random effects. The model is useful to assess whether the observed differences in cesarean rates across hospitals can be justified by case-mix differences across hospitals. Moreover, the discrete nature of the random effects is exploited in order to obtain an optimal clustering of the hospitals affecting decisions on cesarean section in a similar way.

E1138: **Data fusion of heterogeneous data by means of simultaneous matrix/array factorization: overview and challenges**
Presenter: **Tom Frans Wilderjans**, KU Leuven, Belgium
Co-authors: Eva Ceulemans, Iven Van Mechelen

Challenging research questions often imply the simultaneous analysis (i.e., data fusion) of heterogeneous data sets that contain information regarding the same research objects or the same variables, with this information possibly coming from different sources. Heterogeneous data can be presented as a set of coupled (two-way) matrices and/or (three-way or higher) arrays in which each data block (i.e., matrix/array) shares at least one mode with at least one other data block. Examples can be found in, amongst others, bio-informatics (e.g., fusion of Fluorescence Spectroscopy, Nuclear Magnetic Resonance and Liquid Chromatography-Mass Spectrometry data) and psychology (e.g., simultaneous analysis of individuals' behaviour-reaction patterns and their dispositions). To find the joint structure underlying such data, some way of data fusion is needed. To perform this task, we propose the simultaneous matrix/array factorization approach. In this approach, each data block is factorized using a suited decomposition method (e.g., PCA, PARAFAC, HICLAS). For the common mode, the restriction is imposed that the quantification (e.g., component matrix, bundle pattern) of this mode is equal for all data blocks this mode belongs to. The model parameters are estimated by means of alternating least squares algorithms in which different sets of parameters are alternatingly updated conditionally upon the other model parameters. The goal is to introduce the simultaneous matrix/array factorization framework by highlighting its main principles and giving some illustrative examples. To round off, some important issues in data fusion are discussed (e.g., how to weight the different data blocks in the analysis and model selection).

E993: **The statistical swarming method and its validation**
Presenter: **Alain Morineau**, Deenov, France
Co-authors: Thi Minh Thao Huynh, Roland Marion-Gallois, Yann De Rycke

Statistical swarming is creating pseudo-samples called swarms, from an observed initial sample. That simulation method can expand the initial sample to a higher size. For any swarm, this procedure retains the statistical properties of the sample including frequencies, dispersion and connection amongst variables. Any swarm can play the role of a new sample. Swarming is performed in several steps. In the case of qualitative variables, a multiple correspondence analysis is carried out, and new points are simulated according to the distribution of coordinates on principal axes. Relations between variables are transferred by a double weighted average of nearest neighbors: in order to respect the marginal frequencies, and to dilate points beyond the convex envelope of observations. Quality of swarming is estimated using 9108 subjects and 23 nominal variables. The distribution of errors between the global frequencies and any 100-sized sample is estimated by Monte Carlo simulation. For several thousands of these 100-sized samples, we create 250, 500, 700, 1000 and 9000-sized swarms. The errors on frequencies decrease when the size increases (the swarm correctly fills the space around the sample). The errors are around 5% and factorial analyses of swarms are similar to that of the population.

E848: **Binary sentiment classification of documents**
Presenter: **Rahul Biswas**, Indian Statistical Institute, India

Sentiment analysis of documents holds great importance in today's world, when numerous information is stored in databases and in the world wide web. An efficient algorithm to elicit such information, would be beneficial for social, economic as well as medical purposes. We propose a novel binary sentiment classification algorithm, which classifies a document into positive or negative by sentiment. It is important to note that, in the algorithm we have not used the independence assumption, which is used by many procedures like the Naive Bayes. This makes the algorithm more general in scope. We have focused on individual occurences of words and have taken care of misleading occurences, thus making the algorithm more accurate. Moreover, considering the sparsity and high dimensionality of such datasets, we did not use the empirical distribution for estimation, but developed a novel estimation method, by finding degree of close clustering of the data points. We applied our algorithm on a movie review dataset from the Internet Movie Database (IMDb) and obtained satisfactory results.

ES95 Room Q1 BAYESIAN INFERENCE FOR BIG DATA **Chair: Mattias Villani**

E900: **Speeding up MCMC by efficient data subsampling**
Presenter: **Matias Quiroz**, Stockholm University and Sveriges Riksbank, Sweden
Co-authors: Mattias Villani, Robert Kohn

The computing time for Markov Chain Monte Carlo (MCMC) algorithms can be prohibitively large for datasets with many observations, especially when the data density for each observation is costly to evaluate. We propose a framework based on a Pseudo-marginal MCMC where the likelihood function is unbiasedly estimated from a random subset of the data, resulting in substantially fewer density evaluations. The subsets are selected using efficient sampling schemes, such as Probability Proportional-to-Size (PPS) sampling where the inclusion probability of an observation is proportional to an approximation of its contribution to the likelihood function. We illustrate the method on a large dataset of Swedish firms containing half a million observations.

E887: **Yin Yang sampling**
Presenter: **Alexandra Posekany**, WU University of Economics and Business, Austria
Co-authors: Sylvia Fruhwirth-Schnatter

Only recently, methods for splitting big data into subsets, performing inference independently in parallel and merging these outputs have become a topic of interest in Bayesian computing. Such approaches are of vast importance in fields like social network and search engine data, as well as medical statistics and econometrics, as data frequently become too large for a single analysis due to the computational burden. Combining the independently obtained results, i.e. posterior distributions, in order to obtain a common result which recovers the joint posterior distribution and resulting posterior estimators or decisions is far from trivial. To resolve this issue, we propose yin-yang sampling which provides a mathematically sound way of merging two samples from posterior distributions based on different subsets of the data by correcting for applying the same prior for each subset instead of only once for the full data set. Applying yin-yang steps sequentially or in a pairwise way allows for recovering the full sample's posterior from samples from posteriors of any given number of reasonably large subsets. For demonstrating our approach, we analyse simulated linear regression data and real economic data.

E1028: **Delayed acceptance with prefetching**
Presenter: **Clara Grazian**, Universite Paris-Dauphine, France
Co-authors: Christian P. Robert, Marco Banterle

MCMC algorithms such as Metropolis-Hastings algorithms are slowed down by the computation of complex target distributions as exemplified by huge datasets. We offer an approach to reduce the computational costs of such algorithms by a simple and universal divide-and-conquer strategy. The idea behind the generic acceleration is to divide the acceptance step into several parts, aiming at a major reduction in computing time that outranks the corresponding reduction in acceptance probability. The division decomposes the "prior x likelihood" term into a product such that some of its components are much cheaper to compute than others. Each of the components can be sequentially compared with a uniform variate, the first rejection signalling that the proposed value is considered no further. The approach can in turn be accelerated as part of a prefetching algorithm taking advantage of the parallel abilities of the computer at hand. We illustrate those accelerating features on a series of toy and realistic examples.

E1295: **Efficient variational Bayes inference for generalized linear mixed models with large datasets**
Presenter: **Robert Kohn**, University of New South Wales, Australia

A hybrid Variational Bayes algorithm is developed that combines the mean-field and fixed-form Variational Bayes methods. The new estimation algorithm can be used to approximate any posterior without relying on conjugate priors. We propose a divide and recombine strategy for the analysis of large datasets, which partitions a large dataset into smaller pieces and then combines the variational distributions that have been learnt in parallel on each separate piece using the hybrid Variational Bayes algorithm. We also describe an efficient model selection strategy using cross validation, which is straightforward to implement as a by-product of the parallel run. The proposed method is applied to fitting generalized linear mixed models. The computational efficiency of the parallel and hybrid Variational Bayes algorithm is demonstrated on several simulated and real datasets.

ES115 Room C1 ELECTRICITY LOAD FORECASTING **Chair: Jean-Michel Poggi**

E180: **Statistical models for electricity load forecasting at different scales**
Presenter: **Yannig Goude**, EDF, France
Co-authors: Pierre Gaillard

Short-term electricity forecasting has been studied for years at EDF and different forecasting models were developed from various fields of statistics or machine learning (functional data analysis, time series, non-parametric regression, boosting, bagging). We are interested in the forecasting of energy data at different scales (national electricity load, substations, heat curves ...) based on these different approaches. We investigate in the empirical study how to use them to improve prediction accuracy. First, we show how combining members of the original set of forecasts can lead to a significant improvement. Second, we explore how to build various and heterogeneous forecasts from these models and analyze how we can aggregate them to get even better predictions.

E114: **Control forecasting performance over time: An application of control charts to the energy sector**
Presenter: **Daniele Amberti**, iC Analytics, Italy

In the energy consumption forecast practice, the problem of choosing a good forecasting model is as important as controlling its performance over time, on a cluster of time series. We focus on this latter topic and discuss monitoring processes through sequential schemes on prediction errors. Our aim is to evaluate the capability to detect process shifts and changes in the explanatory variables, with techniques such as Shewhart, CUSUM, and EWMA and their potential in driving decision making in the forecasting process and, at the end, in reducing forecasting errors.

E138: **Sparse functional regression for intra day load curve forecasting**
Presenter: **Mathilde Mougeot**, Universite Paris Diderot, France
Co-authors: Dominique Picard, Vincent Lefieux, Laurence Teyssier-Maillard

Managing and developing the electricity transport network is essential to provide quality electricity on a continuous basis to all consumers. We investigate here sparse functional regression models to forecast electricity consumption. The consumption time series is analyzed through intra-day load curves of 24 hours sampled each 30mn. Using a non-parametric model, we first show that each curve can be approximated by a sparse linear combination of functions of a dictionary composed of both specific well elaborated endogenous functions and exogenous functions provided by weather conditions. The forecasting strategy begins with an information retrieval task. Several sparse prediction models are provided by different 'experts'. Each expert computes a model based on a dedicated strategy for choosing the most accurate selection of dictionary variables and estimation of the linear combination. The final forecast is computed using an aggregation of these different forecasters, with exponential weights. We elaborate and test this method in the setting of predicting the national French intra-day load curve, over a period of time of 7 years on a large data basis including daily French electrical consumptions as well as many meteorological inputs, calendar statements and functional dictionaries. The results on the national French intra-day load curves strongly show the benefits of using a sparse functional model to forecast the electricity consumption.

E139: **A prediction interval for a function-valued forecast model. Application to electricity load curves**
Presenter: **Jairo Cugliari**, Universite Lyon, France
Co-authors: Anestis Antoniadis, Xavier Brossat, Jean-Michel Poggi

The motivation comes from electricity consumption forecasting and uses a function-valued time series representation of the discrete electricity records. The data are seen as a sequence of functions $Z_1(t),\ldots,Z_n(t)$ with $t \in T$ representing daily load curves and typically the aim is to predict the function $Z_{n+1}(t)$ that corresponds to the next day load curve. If Z is stationary, then it has been previously proposed the KWF (Kernel + Wavelet + Functional) predictor based on a non-linear autoregressive model. The general principle of the forecasting model is to find in the past similar situations to the present and linearly combine their futures to build the forecast. Thus, using an appropriate similarity measure one can obtain the functional predictor in terms of weights defined through a kernel and a dissimilarity measure between curves based on wavelets. If the functional time series Z is non-stationary, the KWF predictor fails to correctly predict and several strategies were previously proposed to take into account the various sources of non-stationarity allowing to handle situations such that the mean level of the series changes over time or if there exist groups in the data that can be modeled as classes of stationarity. We study the construction of a confidence interval for the prediction. The original prediction method, assuming that Z is stationary, uses a bootstrap re-sampling scheme to construct the confidence interval. We adapt some of the non stationarity corrections proposed for the pointwise prediction to construct the confidence interval for the prediction on the non-stationary case.

E167: **Construction of an informative hierarchical prior for a small sample with the help of historical data and application**
Presenter: **Anne Philippe**, Nantes, France

The interest is in the estimation and prediction of a parametric model on a short dataset upon which it is expected to overfit and perform badly. To overcome the lack of data (relatively to the dimension of the model) we propose the construction of an informative hierarchical Bayesian prior based upon another longer dataset which is assumed to share some similarities with the original, short dataset. We illustrate the performance of our prior on simulated datasets from two standard models. We then apply the methodology to a working model for the electricity load forecasting on real datasets, where it leads to a substantial improvement of the quality of the predictions.

ES123 Room G1 SPARSITY AND NETWORK Chair: Gareth Peters

E089: **Fused community detection**
Presenter: **Yi Yu**, University of Cambridge, United Kingdom
Co-authors: Yang Feng, Richard Samworth

Community detection is one of the most widely studied problems in network research. In an undirected graph, communities are regarded as tightly-knit groups of nodes with comparatively few connections between them. Popular existing techniques, such as spectral clustering and variants thereof, rely heavily on the edges being sufficiently dense and the community structure being relatively obvious. These are often not satisfactory assumptions for large-scale real-world datasets. We therefore propose a new community detection method, called fused community detection (fcd), which is designed particularly for sparse networks and situations where the community structure may be opaque. The spirit of fcd is to take advantage of the edge information, which we exploit by borrowing sparse recovery techniques from regression problems. Our method is supported by both theoretical results and numerical evidence. The algorithms are implemented in the R package fcd, which is available on cran.

E144: **Sparse high-dimensional networks**
Presenter: **Ernst Wit**, University of Groningen, Netherlands
Co-authors: Fentaw Abegaz

A graph is one possible way to simplify complex relationships between many actors, such as people, genes or cognitive features. When data is obtained from noisy measurements on the nodes in the graph, possibly through time, then (chain) graphical models present an appealing and insightful way to describe graph-based dependencies between the random variables. Although potentially still interesting, the main aim of inference is not the precise estimation of the parameters in the graphical model, but the underlying structure of the graph. Graphical lasso and related methods opened up the field of sparse graphical model inference in high-dimensions. We show how extensions of such methods in more structured settings can improve interpretation. Moreover, we show how novel model selection criteria can deal with the determination of the underlying graph in an efficient way.

E213: **Bayesian model averaging of stochastic block models to estimate the graphon function**
Presenter: **Stephane Robin**, INRA - AgroParisTech, France
Co-authors: Pierre Latouche

W-graph refers to a general class of random graph models that can be seen as a random graph limit. It is characterized by both its graphon function and its motif frequencies. The stochastic block model is a special case of W-graph where the graphon function is block-wise constant. We propose a variational Bayes approach to estimate the W-graph as an average of stochastic block models with increasing number of blocks. We derive a variational Bayes algorithm and the corresponding variational weights for model averaging. In the same framework, we derive the variational posterior frequency of any motif. A simulation study and an illustration on a social network complete our work.

E413: **Community detection in sparse random networks**
Presenter: **Nicolas Verzelen**, INRA Montpellier, France
Co-authors: Ery Arias-Castro

The problem of detecting a tight community in a sparse random network is considered. This is formalized as testing for the existence of a dense random subgraph in a random graph. Under the null hypothesis, the graph is a realization of an Erdos-Renyi graph on N vertices and with connection probability p_0 ; under the alternative, there is an unknown subgraph on n vertices where the connection probability is $p_1 > p_0$. We derive a detection lower bound for detecting such a subgraph in terms of N, n, p_0 , p_1 and exhibit a test that achieves that lower bound. We do this both when p_0 is known and unknown. We also consider the problem of testing in polynomial-time.

E1218: **Bootstrap method for networks and its properties in random graphs**
Presenter: **Haruhisa Nagata**, Osaka University, Japan
Co-authors: Hidetoshi Shimodaira

Bootstrap is a popular method to estimate properties of estimators for sampled data, but not for network data. Applying bootstrap for network data is preferable since it can be used without any parametric assumptions and also is suitable computationally. However, it is not obvious how to apply bootstrap for networks, since a network data is a pair of a vertex set and an edge set, and we can not simply resample them. To overcome this issue, we propose a resampling method for edge set based on poisson distribution. We show this resampling method has consistency for some estimators in random graphs.

Sunday 7.12.2014 14:45 - 16:25 Parallel Session L – CFE

CS05 Room Q2 MEASURING SYSTEMIC RISK **Chair: Massimiliano Caporin**

C418: **Systemic and systematic risk**
Presenter: **Roberto Panzica**, Goethe University House of finance, Italy
Co-authors: Monica Billio, Massimiliano Caporin, Loriana Pelizzon

The need for understanding the propagation mechanisms behind the recent financial crises lead the increased interest for works associated with systemic risks. In this framework, network-based methods have been used to infer from data the linkages between institutions (or companies). Part of the literature postulates that systemic risk is strictly related (if not equal to) systematic risk. We elaborate on this hypothesis and introduce a modeling framework where systemic and systematic risks co-exist. The model is a variation of the traditional CAPM where networks are used to infer the exogenous and contemporaneous links across assets. The systemic risk component acts in an additive way on both the systematic and idiosyncratic risk components. Our proposed methodology is verified both on simulations as well as on real data.

C521: **Modeling financial sector joint tail risk, with an application to the euro area**
Presenter: **Bernd Schwaab**, European Central Bank, Germany

A novel high-dimensional non-Gaussian modeling framework is developed to infer conditional and joint risk measures for financial sector firms. The model is based on a dynamic Generalized Hyperbolic Skewed-t block-equicorrelation copula with time-varying volatility and dependence parameters that naturally accommodates asymmetries, heavy tails, as well as non-linear and time-varying default dependence. We demonstrate how to apply a conditional law of large numbers in this setting to define reliable risk measures that can be evaluated within seconds. We apply the modeling framework to assess the joint risk from multiple financial firm defaults in the euro area during the financial and sovereign debt crisis, and document unprecedented tail risks during 2011-12, as well as their steep decline after subsequent policy reactions. Interestingly, whereas joint default probabilities appear to decline after the policy announcements, conditional probabilities remain high, indicating that perceived systemic clustering is still almost as high as during the sovereign debt crisis.

C559: **Information flow and entropy in networks of financial markets**
Presenter: **Harald Schmidbauer**, Istanbul Bilgi University, Turkey
Co-authors: Angi Roesch

Financial markets can be seen as nodes in a directed network with sophisticated edge weights quantifying news-to-volatility spillovers in terms of forecast error variance decompositions (fevds) of daily returns on their equity prices. Several market connectedness measures which turned out to be closely related to concepts of network connectedness have been proposed, including an average connectedness index, to provide an assessment of systemic risk. Transformation of the adjacency matrix into the transition matrix of a Markov process running forward in time enables us to monitor information-theoretic aspects of the network's dynamics. A day-to-day perspective leads us to measure the amount of information injected into the network of markets on a daily basis, which we refer to as the relative market entropy. From a within-day perspective, the rate at which information is generated and digested can be quantified, which translates into a timely assessment of network stability. Comparison of this dynamical entropy per unit time to its time reversed analogue provides a measure of the information flow's irreversibility and hence of the information asymmetry among markets. Analyzing a network of equity markets, we demonstrate that increasing trends in connectedness as well as in speed of information digestion are an empirical fact but no logical necessity.

C1267: **Systemic risk spillovers in the European banking and sovereign network**
Presenter: **Tuomas Peltonen**, European Central Bank, Germany
Co-authors: Frank Betz, Nikolaus Hautsch, Melanie Schienle

A framework is proposed for estimating network-driven time-varying systemic risk contributions that is applicable to a high-dimensional financial system. Tail risk dependencies and contributions are estimated based on a penalized two-stage fixed effects quantile approach, which explicitly links bank interconnectedness to systemic risk contributions. The framework is applied to a system of 51 large European banks and 17 sovereigns through the period 2006 to 2013, utilizing both equity and CDS prices. We provide new evidence on how banking sector fragmentation and sovereign-bank linkages evolved over the European sovereign debt crisis and how it is reflected in network statistics and systemic risk measures. Illustrating the usefulness of the framework as a monitoring tool, we provide indication for the fragmentation of the European financial system having peaked and that recovery has started.

CS15 Room B2 MULTIVARIATE MODELLING OF ECONOMIC AND FINANCIAL TIME SERIES **Chair: Gianluca Cubadda**

C233: **Non-causal MIDAS models for nowcasting GDP**
Presenter: **Alain Hecq**, Maastricht University, Netherlands
Co-authors: Thomas Gotz

Time series data are available at mixed frequencies and a popular model to account for this is the MI(xed) DA(ta) S(ampling) model introduced by Ghysels. The main virtue of the MIDAS is its parsimonious specification, which allows researchers to make use of potentially all high-frequency observations with a relatively small number of parameters to estimate. However, due to the publication delay, preliminary estimates (the first release) of important low frequency data such as the GDP are available 25 days after the corresponding quarter in the US and about 45 days for the EU member states; it takes about 60 days to get the second release. This means that for the first quarter of 2014 for instance, we have, say, the second evaluation at the end of May 2014 while most high frequency data, and particular financial variables (stock prices, interest rates) are already published. Authors call this observation the "x before y" case. We introduce a new non-causal MIDAS model specification to deal with this important situation, extending both the usual MIDAS causal model a la Ghysels and non-causal ARMA models whose behavior has recently been reevaluated in forecasting exercises by Lanne and co-authors. After having identified the lag structure of the MIDAS, we use a Student-t distribution to estimate and to nowcast the German GDP.

C437: **Cointegration between processes integrated at different frequencies**
Presenter: **Tomas del Barrio Castro**, University of the Balearic Islands, Spain
Co-authors: Giaunluca Cubadda, Denise Osborn

The purpose is to show that it is possible to establish cointegration relationship between integrated processes associated to different frequencies. In particular we show that the only cointegration possibility between two processes integrated two different frequencies is full polynomial periodic integration. Furthermore we show the connection between the demodulator operator introduced in the Granger and Hatanaka book with full polynomial cointegration. We also propose three alternative ways to test if processes integrated at different frequencies could be cointegrated. In a Monte Carlo experiment we show the behavior on the three methods. Finally, we illustrate our proposal with empirical quarterly and monthly data.

C552: **Forecasting private consumption by consumer surveys and canonical correlations**
Presenter: **Marco Centoni**, LUMSA, Italy
Co-authors: Giancarlo Bruno, Claudio Lupi

In recent years there has been a growing interest in using consumer surveys as leading indicators of real economic activity, especially to assess the

future path of private consumption. However, consumer surveys data amount to large data sets, often condensed in rather ad hoc measures of the so-called consumer sentiment. We are interested in evaluating the pros and cons deriving from using consumer surveys and canonical correlations to forecast private consumption components in real time. In the present framework, there are different reasons to consider canonical correlations. (1) Canonical correlations lead to optimal forecasting linear combinations; (2) the solution offered by canonical correlations should be more flexible than that offered by, for example, vector autoregressions (VARs), given that the information set has not to be uniformly updated to time $t-1$: this is especially relevant for nowcasting; (3) it should be possible to use a much larger set of potentially important predictors than allowed by VAR analysis; (4) beside producing forecasts of private consumption growth, using canonical correlations we are able to offer an alternative measure of the consumer sentiment. The forecasting performance of the proposed methods is compared to that of some widely used benchmarks using Italian data.

C658: **Analysis of cointegrated models with measurement errors**
Presenter: **Sung Ahn**, Washington State University, United States
Co-authors: Hanwoom Hong, Sinsup Cho

The asymptotic properties of the reduced-rank estimator of error correction models of vector processes observed with measurement errors are studied. Although it is well known that there is no asymptotic measurement error bias when predictor variables are integrated processes in regression models, we systematically investigate the effects of the measurement errors (in the dependent variables as well as in the predictor variables) on the estimation of not only cointegrating vectors but also the speed of adjustment matrix. Furthermore, we present the asymptotic properties of the estimators. We also obtain the asymptotic distribution of the likelihood ratio test for the cointegrating ranks and investigate the effects of the measurement errors on the test through a Monte Carlo simulation study.

CS19 Room H2 BANKS, INTEREST RATES AND PROFITABILITY AFTER THE FINANCIAL CRISIS Chair: Paul Mizen

C1007: **A factor model of interest rate pass through for four large Euro area countries**
Presenter: **Anindya Banerjee**, University of Birmingham, United Kingdom
Co-authors: Paul Mizen, Victor Bystrov

The results of estimation of a factor-augmented vector auto-regressive model are described that studies the markup on mortgage corporate in the four largest economies of the Euro Area (Germany, France, Italy, and Spain). The markup is defined as the difference between a retail rate and a synthetic cost-of-fund measure, constructed as a weighted average of rates on alternative sources of financing for Monetary Financial Institutions (MFIs). The factors, extracted from a large set of macroeconomic and financial indicators, summarize information that can be used by MFIs in the forecasting and evaluation of risk. Impulse response analysis is employed to evaluate the dynamic effect of the extracted factors onto the markup. Loadings of factors onto subsets of variables are used to obtain information about the relative importance of macroeconomic and financial variables in the determination of the markup.

C718: **Funding costs, retail rates and the relative demand for bank loans**
Presenter: **Arvid Raknerud**, Statistics Norway, Norway
Co-authors: Bjorn Helge Vatne

The aim is to examine the relation between banks' funding costs, retail rates and the relative demand for loans from different banks. The data set consists of quarterly accounts data on Norwegian banks for the period 2002Q1-2011Q3. We identify demand elasticities by using a model of monopolistic competition and derive exclusion restrictions, i.e., variables that affect retail rates but not the demand for bank loans directly. Key exogenous variables are the three-month money market rate (NIBOR) and corresponding forward rates. Our estimates show that a 10 basis points increase in the NIBOR leads to approximately 85 and 95 basis points increases in retail rates for household mortgages and corporate loans, respectively. We identify a significant positive relation between forward rates and retail rates for corporate loans, with forward rates accounting for about 45 percent of the marginal funding costs for loans to businesses, compared to 20 percent for households. Our estimated demand elasticities indicate a high degree of market power among major Norwegian banks, especially for corporate loans. We estimate that, on average, if the ratio of the loan rate of an individual bank (or bank group) to the market loan rate increases by one percent, its relative loan volume (compared to a market quantity index decreases) by, respectively, 1.44 percent for household mortgages and 0.65 percent for corporate loans.

C1021: **Bank profitability and monetary policy**
Presenter: **Boris Hofmann**, BIS, Switzerland
Co-authors: Claudio Borio, Leonardo Gambacorta

The aim is to investigate how monetary policy affects bank profitability. We use data for 109 large international banks headquartered in 14 major advanced economies for the period 1995-2012. We find that the relationship between the net interest income and the interest rate structure is positive but concave, because deposit rates cannot fall below zero. By contrast, the impact of interest rates on other non-interest income is negative and convex, due to derivatives hedging and valuation effects of securities. We also find a non-linear relationship between provisions and the interest rate structure. It is positive at the beginning because higher rates increase the service of debt for borrowers and their default probability, and negative after a critical threshold when high rates depress the overall quantity of supplied loans. Overall, we find a positive and concave relationship between the interest rate structure and bank profitability.

C982: **Did European banks set interest rates too high after the global financial crisis?**
Presenter: **Paul Mizen**, University of Nottingham, United Kingdom
Co-authors: Marco Lombardi, Anamaria Illes

The global finance crisis prompted central banks in many countries to cut short-term policy rates to near zero levels, but in spite of this lending rates set by banks did not fall as much as policy rates suggest they should have done. Were banks taking advantage by failing to pass on these rates to loans when policy rates were at historically low values? We argue that reference to the policy rate is misleading. Banks do not obtain funds at policy rates, and their own cost of funds rose after the crisis. Comparisons between lending rates and the true weighted average cost of funds suggest that banks have not widened their margins in the low interest rate environment. Interest rate pass-through relationships across eleven countries in Europe are stable when compared to weighted cost of funds, but unstable when compared to policy rates.

CS31 Room O2 DYNAMIC CONDITIONAL SCORE MODELS Chair: Andrew Harvey

C585: **Modeling the interactions between volatility and returns**
Presenter: **Andrew Harvey**, University of Cambridge, United Kingdom
Co-authors: Rutger-Jan Lange

Volatility of a stock may incur a risk premium, leading to a positive correlation between volatility and returns. On the other hand the leverage effect, whereby negative returns increase volatility, acts in the opposite direction. We propose a two component ARCH in Mean model to separate the two effects; such a model also picks up the long memory features of the data. An exponential formulation, with the dynamics driven by the score of the conditional distribution, is shown to be theoretically tractable as well as practically useful. In particular it enables us to write down the asymptotic distribution of the maximum likelihood estimator, something that has not proved possible for standard formulations of ARCH in Mean.

Our EGARCH-M model in which the returns have a conditional skewed generalized-t distribution is shown to give a good fit to daily S&P500 excess returns over a 60-year period (4 Jan 1954 - 30 Dec 2013).

C710: **Shrinkage for large time-varying parameter models: a penalized score driven approach**
Presenter: **Davide Delle Monache**, Queen Mary - University of London, United Kingdom
Co-authors: Christian Brownlees, Ivan Petrella

Allowing for a large a number of parameters in time-varying parameter models may be detrimental to predictive accuracy. In order to counter the curse of dimensionality we propose a class of penalized score driven models based on Ridge and LASSO penalization. The modeling approach consists of updating the time-varying parameter on the basis of the regularized score of the log-likelihood function. The Ridge regularization scheme shrinks parameter updates to zero while the Lasso regularization scheme produces sparse parameter updates by truncating components of the parameter vector to zero. A macro time-series forecasting application using a large number of predictors is used to highlight the usefulness of the methodology.

C954: **High dimensional dependence modelling with heavy tailed, asymmetric factor models**
Presenter: **Steve Thiele**, University of Cambridge, United Kingdom

A new approach for modelling the dependence between a large cross section of financial assets is proposed. It brings together traditional factor analysis with recent advances in the modelling of time varying parametric distributions using dynamic conditional score (DCS) models. Asymptotic properties of the model are derived in a large N,T setting. An application to a large cross section of US equity returns is considered, yielding out-of-sample risk management and portfolio construction results that are favourable against a leading alternative from the literature.

C659: **The distribution of the error term in dynamic conditional score models**
Presenter: **Alessandra Luati**, University of Bologna, Italy

Dynamic conditional score (DCS) models have been recently introduced to account for data that may be generated from heavy tail distributions. Due to their flexibility and the fact that a comprehensive asymptotic theory based on maximum likelihood can be developed in closed form, these models have immediately received a great deal of attention. The key feature is that the dynamics of the parameters that specify the distribution of the data depend on the score of the likelihood function. When the data come from a Student-t distribution, then the score is a function of a beta random variable. This function is the identity when the scale of the distribution is the parameter of interest. In the case when the location parameter is dynamic, then the distribution of the error term is unknown. The distribution of the score in a DCS-t location model based on integral transforms is derived. We then develop goodness of fit test and related inferential procedures for this class of models.

CS45 Room F2 DEPENDENCE MODELING AND COPULAS Chair: Hans Manner

C158: **Efficient iterative maximum likelihood estimation of high-parameterized time series models**
Presenter: **Ostap Okhrin**, Humboldt-University Berlin, Germany
Co-authors: Alexander Ristig, Nikolaus Hautsch

The aim is to propose an iterative procedure to efficiently estimate models with complex log-likelihood functions and the number of parameters relative to the observations being potentially high. Given consistent but inefficient estimates of sub-vectors of the parameter vector, the procedure yields computationally tractable, consistent and asymptotic efficient estimates of all parameters. We show the asymptotic normality and derive the estimator's asymptotic covariance in dependence of the number of iteration steps. To mitigate the curse of dimensionality in high-parameterized models, we combine the procedure with a penalization approach yielding sparsity and reducing model complexity. Small sample properties of the estimator are illustrated for two time series models in a simulation study. In an empirical application, we use the proposed method to estimate the connectedness between companies by extending a previous approach to a high-dimensional non-Gaussian setting.

C508: **Construction and sampling of Archimedean and nested Archimedean Levy copulas**
Presenter: **Oliver Grothe**, University of Cologne, Germany
Co-authors: Marius Hofert

The class of Archimedean Levy copulas is considered with focus on the construction and sampling of Levy processes with these underlying Levy copulas. Furthermore, the class of nested Archimedean Levy copulas is introduced. This class allows one to model hierarchical dependencies between Levy processes. It also overcomes the inherent symmetry linked to Archimedean Levy copulas. Finally, a new sampling algorithm for multivariate Levy processes with dependence structure specified by either Archimedean or nested Archimedean Levy copulas is derived from a Marshall-Olkin-type algorithm. In contrast to the widely used conditional sampling method, this algorithm does not require (inverses of) conditional Levy copulas to be known. It also does not suffer from an asymmetric bias introduced by the conditional sampling method.

C197: **Investigating financial contagion with copulae**
Presenter: **Alexander Ristig**, Humboldt-University, Germany
Co-authors: Ostap Okhrin, Jeffrey Sheen, Stefan Trueck

A unified framework for investigating contagion in financial systems is proposed. The copula function plays a central role in this analysis, as it describes the dependence properties of the entire financial system and provides a natural link between the modeled system and the proposed risk measures. These quantities share the same features as elasticities, so that conclusions about the vulnerability of system parts with respect to an increase in the distress in other parts can be easily drawn. For representing a financial system with a few parameters in practice, we estimate the parameters and aggregate the structure of hierarchical Archimedean copulae simultaneously by imposing a non-concave penalty on the structure. The asymptotic properties of this sparse estimator are derived and the small sample properties are illustrated in a simulation study. In the empirical study we investigate the fragility of Australia's financial market within its relevant financial system.

C238: **Detecting financial contagion in a multivariate system**
Presenter: **Hans Manner**, University of Cologne, Germany
Co-authors: Dominik Blatt, Bertrand Candelon

An original three-part sequential testing procedure (STP) is proposed, with which to test for contagion using a multivariate model. First, it identifies structural breaks in the volatility of a given set of countries. Then a structural break test is applied to the correlation matrix to identify and date the potential contagion mechanism. As a third element, the STP tests for the distinctiveness of the break dates previously found. Compared to traditional contagion tests in a bivariate set-up, the STP has high testing power and is able to locate the dates of contagion more precisely. Monte Carlo simulations underline the mportance of separating variance and correlation break testing, the endogenous dating of the breakpoints and the usage of multi-dimensional data. The procedure is applied for the 1997 Asian Financial Crisis, revealing the chronological order of the crisis events.

CS67 Room P2 FINANCIAL AND MACROECONOMIC FORECASTING Chair: Francesco Ravazzolo

C357: **Probabilistic forecasting based on MCMC output**
Presenter: **Fabian Krueger**, Heidelberg Instititute for Theoretical Studies, Germany
Co-authors: Sebastian Lerch, Thordis Thorarinsdottir, Tilmann Gneiting

A systematic analysis of how to make and evaluate probabilistic forecast distributions based on Markov Chain Monte Carlo (MCMC) output is

conducted. We first survey a variety of approaches proposed in the interdisciplinary, rapidly growing literature. We then use analytical arguments, simulation evidence and empirical analysis in order to evaluate these proposals. In particular, we compare different methods of estimating the stationary distribution underlying the MCMC output. Given such an estimate and a realizing outcome, we consider forecast evaluation based on proper scoring rules. We find that an attractive option is to use the Continuous Ranked Probability Score.

C624: **Bayesian nonparametric calibration and combination of predictive distributions**
Presenter: **Roberto Casarin**, University Ca Foscari of Venice, Italy
Co-authors: Francesco Ravazzolo, Tilmann Gneiting

The aim is to introduce a Bayesian approach to predictive density calibration and combination that accounts for parameter uncertainty and model set completeness through the use of random calibration functionals and random combination weights. Building on previous works, we use infinite beta mixtures for the calibration. The proposed nonparametric approach takes advantage of the flexibility of infinite mixtures to achieve any continuous deformation of linearly combined predictive distributions. Estimation is based on Dirichlet process priors and accounts for uncertainty in the number of mixture components, mixture weights, and calibration parameters. We study the methodology in simulation examples with fat tails and multimodal densities, and apply it to predict daily stock returns and daily maximum wind speed.

C509: **Modeling, decomposing, and forecasting interest rate movements in Brazil and Mexico**
Presenter: **Michiel De Pooter**, Board of Governors of the Federal Reserve System, United States

The term structure of interest rates for Brazil and Mexico is analyzed. We first construct historical time-series of zero-coupon yield curve estimates from prices on Brazilian and Mexican nominal bonds, with our curves extending back to July 2006 and January 2003 for Brazil and Mexico, respectively. We then use our zero-coupon curves to estimate affine term structure models for each country using a state space model approach and incorporating survey values for short-term interest rates from Consensus Economics to help guide the estimation of the short-term interest rate path. Using these models we decompose interest rate movements into their average expected rate and term premium components and we find a strong co-movement between these components for Brazilian and Mexican yields and their U.S. yield counterparts, a testament of how much Brazilian and Mexican bond markets have developed over the past decade. We then use a regression approach to establish the main drivers of term premiums in Brazil and Mexico and we hone in on the effects that the Fed's QE programs and in particular its recent "tapering" of asset purchases has had on Brazilian and Mexican yields. Finally, we examine the predictability of yields in each country.

C683: **Credit conditions and external source of financing**
Presenter: **Giulio Nicoletti**, European Central Bank, Germany
Co-authors: Carlo Altavilla, Matthieu Darraq

Credit conditions of non-financial corporations (NFCs) in the euro area largely depend on banks' loans. However, we recently observed significant substitution between banks' intermediated credit and non-financial corporations' debt issuance. The aim is to analyse the macroeconomic implications of such loans-into-bond substitution using a Bayesian VAR model on a set of macroeconomic and credit variables. In particular, we disentangle three main drivers of substitution: credit supply; exogenous firms' preferences and monetary policy. Structural identification is achieved using external instruments following previous works. Results show that a tightening in credit supply triggers a contraction in both economic activity and inflation. The resulting contraction in loans and surge in firms' financing costs enhances private sector's incentives to substitute bank loans with bonds. In our historical decomposition, depressed credit supply explains about half of the loan-into-bond substitution observed in the aftermath of the Lehman event, much less however for the period following the euro area sovereign crisis. In this latter period the loan-into-bond substitution is mostly explained by an autonomous firms' preference shock. Concerning the role of monetary policy, non-standard measures in the euro area seem to stimulate creation of loans more strongly than bond issuance.

CS77 Room N2 MODELLING UNIVERSITY OUTCOMES THROUGH SURVIVAL ANALYSIS Chair: Mark Steel

C204: **Bayesian survival modelling of university outcomes**
Presenter: **Catalina Vallejos**, MRC Biostatistics Unit and EMBL European Bioinformatics Institute, United Kingdom
Co-authors: Mark Steel

The aim is to model the length of registration at university and its associated academic outcome for undergraduate students at the Pontificia Universidad Catolica de Chile. Survival time is defined as the time until the end of the enrollment period, which can relate to different reasons - graduation or two types of dropout - that are driven by different processes. Hence, a competing risks model is employed for the analysis. The issue of separation of the outcomes (which precludes maximum likelihood estimation) is handled through the use of Bayesian inference with an appropriately chosen prior. We are interested in identifying important determinants of university outcomes and the associated model uncertainty is formally addressed through Bayesian model averaging. The methodology introduced for modelling university outcomes is applied to three selected degree programmes, which are particularly affected by dropout and late graduation.

C281: **A multilevel competing risks model for analysis of university students' careers: evidence from Italy**
Presenter: **Anna Giraldo**, University of Padova, Italy
Co-authors: Silvia Meggiolaro, Renata Clerici

Individual and institutional characteristics which may influence the outcomes of university students' careers are examined. In particular, the outcomes considered are the withdrawals, course changes, delays or graduations of students enrolled in first-cycle degree courses in a large public university in North-East Italy. Individual longitudinal data from administrative archives were used, taking into account both the temporal dimension in a discrete-time competing risks approach, and the organisational and structural characteristics of the degree courses in which students were enrolled. In order to examine the hierarchical nature of the data properly, analyses were carried out within a multilevel framework. At individual level, results indicate that the profile of a successful student is defined by both socio-demographic factors and pre-university educational experience. At course level, some characteristics such as restricted access to some courses, study fields, and course size were important for students' university careers, although the effects were not always in the expected direction.

C407: **Predictors of dropout and degree completion in the Belgian French community's higher education system**
Presenter: **Catherine Dehon**, Universite libre de Bruxelles, Belgium
Co-authors: Elena Arias

The aim is to study the factors that influence both dropout and degree completion throughout the university by applying the set of discrete-time methods for competing risks in event history analysis. In the French-speaking Belgian community, participation rates are very high given that higher education is largely financed through public funds, but at the same time, the system performs very poorly in terms of degree completion. In this particular context, we explore the following question: to what extent is socioeconomic background still a determinant of success for academic careers in a system that, by construction, aims to eliminate economic barriers to higher education? In a second section, the determinants of 'time to dropout' from doctoral studies and 'time to Ph.D. completion' are studied using similar tools for a sample of 3.092 doctoral candidates from the ULB. Not surprisingly, results show that students supported with research fellowships have much higher Ph.D. completion hazards than teaching assistants or unfinanced students.

C822: **The effect of a tuition fee reform on the timing of drop out from Higher Education in the UK**
Presenter: **Giuseppe Migali**, Lancaster University Management School, United Kingdom

Co-authors: Steve Bradley

Tuition fees were increased to 3,000 GBP per student per annum as a result of a policy reform that was implemented in 2006. This policy applied to English and Northern Irish students, and in 2007 was extended to Welsh students - Scottish students were exempt. Furthermore, if students dropped out before 1st December in the year in which they started their undergraduate education then the accumulated debt was 'written off', so providing an incentive for students to drop out early. Student level data obtained from HESA for the period 2003-2010 are used to: a) evaluate the effects of the policy reform on the risk of drop out; in addition we investigate b) whether students are increasingly 'debt averse', reflected by an increased risk of drop out prior to December in the post-reform period. Our focus is upon 1st year students and we model male and female behaviour separately. We use survival analysis with and without a control for unobserved heterogeneity. Our (very) preliminary analysis suggests that there is an effect of the policy reform insofar as it increases the risk of drop out.

CS79 Room I2 MODELLING AND COMPUTATION IN MACRO-ECONOMETRICS Chair: Rodney Strachan

C026: **Modelling inflation volatility**

Presenter: **Rodney Strachan**, The University of Queensland, Australia

Co-authors: Eric Eisenstat

Estimation of US inflation volatility is discussed using time varying parameter models, in particular whether it should be modelled as a stationary or random walk stochastic process. Specifying inflation volatility as an unbounded process, as implied by the random walk, conflicts with prior beliefs, yet a stationary process cannot capture the low frequency behaviour commonly observed in estimates of volatility. We therefore propose an alternative model with a change-point process in the volatility that allows for switches between stationary models to capture changes in the level and dynamics over the past forty years. To accommodate the stationarity restriction, we develop a new representation that is equivalent to our model but is computationally more efficient. All models produce effectively identical estimates of volatility, but the change-point model provides more information on the level and persistence of volatility and the probabilities of changes. For example, we find a few well defined switches in the volatility process and, interestingly, these switches line up somewhat with economic slowdowns or changes of the Federal Reserve Chair.

C774: **Efficient Bayesian inference in inverted Wishart stochastic volatility models**

Presenter: **Roberto Leon**, GRIPS, Japan

An inverted Wishart Process for stochastic volatility is considered and a simple and efficient MCMC algorithm is developed. By conditioning on some auxiliary variables, all volatility matrices can be sampled jointly from the joint conditional posterior. Because of its simplicity and efficiency, the MCMC algorithm can be used to tackle relatively large dimensions.

C474: **The stochastic volatility in mean model with time-varying parameters: an application to inflation modeling**

Presenter: **Joshua Chan**, Australian National University, Australia

The popular stochastic volatility in mean model is generalized to allow for time-varying parameters in the conditional mean. The estimation of this extension is nontrival since the volatility appears in both the conditional mean and the conditional variance, and its coefficient in the former is time-varying. We develop an efficient Markov chain Monte Carlo algorithm based on band and sparse matrix algorithms instead of the Kalman filter to estimate this more general variant. We illustrate the methodology with an application that involves US inflation. The estimation results show substantial time-variation in the coefficient associated with the volatility, highlighting the empirical relevance of the proposed extension. Moreover, in a pseudo out-of-sample forecasting exercise, the proposed variant also forecasts better than various standard benchmarks.

C1148: **Comparing computational methods for predictive scores**

Presenter: **Craig Thamotheram**, University of Warwick, United Kingdom

Co-authors: Anthony Garratt, Leif Thorsrud, Shaun Vahey

Computational methods for the predictive scores of forecast densities are compared. This comparison is relevant because when the predictive performances for a continuous variable produced by two (or more) Bayesian forecasting models are to be compared, the researcher wishes to calculate predictive scores quickly and accurately. We simulate a number of time series models and implement a battery of commonly-deployed solution methods to calculate the logarithmic score and the continuous ranked probability score. The data generating processes comprise both a linear autoregressive model and a threshold-switching model. In each case, we consider variants with Gaussian errors and non-Gaussian errors. The computational methods comprise the analytical approach, and both kernel-estimation and discretisation of the forecast densities. We document and discuss the characteristics of the methods in terms of computing time, complexity, and accuracy.

CS87 Room E2 STATISTICAL ANALYSIS OF FINANCIAL RETURNS Chair: Toshi Watanabe

C408: **Modeling and estimation of stock returns with skew t-copula**

Presenter: **Toshinao Yoshiba**, Bank of Japan, Japan

In financial portfolio risk management, Student's t-copula is frequently used to capture the tail dependence of risk factors. However, the copula is restrictive because of its symmetric dependence at both the upper and lower tails. We, therefore, construct a copula from a well-known multivariate skew t-distribution to capture the asymmetric dependence of risk factors. We indicate two problems in estimating the parameters of the skew t-copula. The first problem is that the log-likelihood function includes univariate skew t-quantile functions, which makes calculating a log-likelihood extremely time consuming. The second problem is that the extended correlation matrix in the parameters should be positive semi-definite and that the each diagonal element should be one. Solving the two problems, we estimate the parameters of the skew t-copula for the daily returns of three major stock indices: the Nikkei225, SP500, and DAX. We investigate the asymmetric tail dependence both for the unfiltered returns and for the filtered returns by some GARCH models. The marginal risk factors are separately estimated by a univariate skew t-distribution to capture the fat-tail property and the asymmetry of the marginal distribution.

C464: **Robust early warning signals of abrupt switches in stock markets**

Presenter: **Teruko Takada**, Osaka City University, Japan

Financial markets have repeatedly experienced abrupt switches from bull to bear phases, and predicting these changes is a challenging problem. The difficulty increases if the size of the trend reversal is on the scale of a financial bubble. The aim is to propose a set of robust and efficient early warning signals of sudden, large trend reversals, demonstrating that three statistical indicators effectively warn of the risks of a critical transition in major stock markets, outperforming conventional indicators based on moment statistics. Moreover, the degree of instability of investors' expectations is visualized by the non-parametrically estimated free energy of the price movement.

C498: **Measurement of causality change between returns of financial assets**

Presenter: **Kosuke Oya**, Osaka University, Japan

Co-authors: Ryo Kinoshita

It is indisputable that the financial turmoil such as the Lehman shock caused some structural change in the economic system. The standard way of detection of structural change is to find out the significant change in the parameters of the model for the period prior to and after the shock. In the case of multiple time series model, the causality between the time series is also change when there is a structural change. However the magnitude of change in causality is not clear even if we detect the significant structural change. We explore the measure of causality change between the time series and propose the test statistic whether there is any significant change in the causal relationship using frequency domain causality measure

given in the previous studies. The properties of the measure and test statistic are examined through the Monte Carlo simulation. As an example of application, the change in causality between United States and Japanese stock indexes is tested.

C571: **Open-end fund characteristics and the effects on financial stability by Investors' herding redemption in Taiwan**
Presenter: **Nan-Yu Wang**, Ta Hwa University of Science and Technology, Taiwan
Co-authors: Chih-Jen Huang, Ying-Lin Hsu

The aim is to examine whether funds with illiquid assets exhibit stronger sensitivity of redemption outflows to bad past performance than funds with liquid assets. An important aspect of our study is whether large outflows should damage future fund performance in illiquid funds more than in liquid funds. When redeemed on a large scale, the liquidity risk of open-end funds will increase, which in turn leads to a vicious circle between fund redemption and the net asset value decline. Accordingly, stepwise superior predict ability test and a manipulation-proof performance measure method are taken into account of the financial stability problem and to control for the data-snooping bias. Based on the sample of underperformed mutual funds in Taiwan, the empirical results show that (1) bad past performance in liquid funds is more sensitive on flow-performance relations; (2) the evidence in (1) exists only for institutional-oriented funds, but not for retail-oriented funds; and (3) Illiquid funds damage from a large number of redemptions with significant return persistence. The above findings provide valuable references for fund managers to make the plan of their investments.

CS93 Room A2 TIME SERIES ANALYSIS IN ECONOMICS Chair: Raffaella Giacomini

C035: **Optimality tests in presence of instabilities**
Presenter: **Barbara Rossi**, Universitat Pompeu Fabra and ICREA, Spain
Co-authors: Tatevik Sekhposyan

Forecast optimality tests that can be used in unstable environments are proposed. They include tests for forecast unbiasedness, efficiency, encompassing, serial uncorrelation, and, in general, regression-based tests of forecasting ability. The proposed tests are applied to evaluate the rationality of the Federal Reserve Greenbook forecasts as well as a variety of survey-based private forecasts. Our robust tests suggest more empirical evidence against forecast rationality than previously found but confirm that the Federal Reserve has additional information about current and future states of the economy relative to market participants.

C043: **Bond market and macroeconomic news**
Presenter: **Michele Modugno**, Federal Reserve Board, United States
Co-authors: Carlo Altavilla, Domenico Giannone

The reaction of the U.S. Treasury bond market to innovations in macroeconomic fundamentals is investigated. We identify these innovations with macroeconomic news defined as differences between the actual releases and their market expectations. We show that macroeconomic news explain about 35% of the quarterly fluctuations in long-term bond yields. This percentage decreases to 8% when focusing on daily movements. This result suggests that although non-fundamental factors substantially influence the day-to-day movements in bond yields their effects are mean-reverting and short-lived. The effects of macro news, instead, are longer-lasting: the surprise component of macroeconomic releases persistently influences the yield curve. When analyzing the robustness of our findings over different sub-samples we find that the interaction between macro news and yields did not break down after the zero lower bound was attained. Our evidence suggests that the Federal Reserve non-standard monetary policies, have been successful in keeping bond yields anchored to macroeconomic news in a period of economic uncertainty.

C069: **Forecasting with large time-varying parameters vars**
Presenter: **Michele Lenza**, European Central Bank, Germany
Co-authors: Domenico Giannone, Gianni Amisano

Time-varying parameters VARs (TVARs) are very useful exploratory devices in applied macro both for forecasting purposes and for estimating the effects of structural shocks. However, due to over-parameterization, these models are particularly affected by the curse of dimensionality. For this reason, the previous literature was confined to TVARs including only a handful of variables. We propose a methodology that handles the estimation of TVARs of a large dimension, without having to abandon the use of MCMC techniques as in previous papers. Our approach rests on postulating a convenient Kronecker structure for the covariance matrix of the errors, which greatly reduces the space of the parameters to estimate. In our empirical application, we estimate a TVAR with seven US macroeconomic variables and four lags and we show that our method (i) allows to easily draw from the posterior of the parameters by means of conventional MCMC methods and (ii) provides accurate out-of-sample forecasts when compared to the usual naive benchmark models considered in the forecasting literature.

C078: **Robust Bayes inference for non-identified SVARs**
Presenter: **Raffaella Giacomini**, UCL, United Kingdom
Co-authors: Toru Kitagawa

A procedure is proposed for conducting robust Bayesian inference in structural vector autoregressions (SVARs) whose parameters are not point identified. Partial identification could arise because the number of equality restrictions one can credibly impose is not sufficient for point identification and/or the restrictions take the form of sign restrictions on the model's parameters or on the impulse response functions. The presence of partial identification makes Bayesian inference sensitive to the choice of prior: unlike for the point-identified case, the effect of the prior does not vanish asymptotically. We show how to conduct inference that is robust to the choice of prior. For example, for impulse response analysis, we propose considering the class of all possible priors (ambiguous beliefs) for the non-identified aspects of the model and reporting the range of posterior means and posterior probabilities for the impulse response function as the prior varies over this class. The posterior bounds we construct asymptotically converge to the true identified set, which is the object of interest of frequentist inference in partially-identified models. In terms of implementation, the posterior bound analysis is computationally simpler and can accommodate a larger class of zero and sign restrictions than the frequentist confidence intervals.

CS44 Room M2 CONTRIBUTIONS IN FINANCIAL ECONOMETRICS I Chair: Henri Nyberg

C1076: **Inflation, deflation, and uncertainty: what drives euro area option-implied inflation expectations?**
Presenter: **Jelena Stapf**, Deutsche Bundesbank, Germany
Co-authors: Michael Scharnagl

Distributions of long-term inflation expectations for the euro area are derived using inflation options on the harmonised consumer price index. Albeit a decreasing mean the uncertainty about expected future realisations of inflation rates soared especially since the intensification of the sovereign debt crisis in mid-2011. Around the same time the influence of monetary policy announcements diminished. Both developments can be explained by a growing disagreement on the influence of monetary policy towards future inflation outcomes among market participants. Tail events such as deflation although still contained became more probable. The impact of macroeconomic news to explain inflation probabilities overall decreased and shifted towards countries more affected by the crisis. For judging the anchoring of inflation expectations the new data set on inflation options provides a twofold result. The mean and low reactivity to actual news speak for firmly anchored inflation expectations whereas the growing uncertainty reveals market participants concerns about possible extreme inflation outcomes in the future.

C1174: **Pairs trading and relative liquidity in the European stock market**
Presenter: **Isabel Figuerola-Ferretti**, Universidad Carlos III de Madrid, Spain
Co-authors: Tao Tang, Ioannis Paraskevopoulos

An equilibrium demand is used and framework is supplied to describe the no-arbitrage relationship between two distinct but cointegrated assets, showing how adjustment dynamics between them can be exploited through pairs trading strategies to gain arbitrage profits out of temporary mispricings. In this framework, the two close substitutes are cointegrated meaning that they measure a common non-stationary factor. Price discovery is determined by the relative number of participants or relative liquidity in both markets. This theoretical model, which builds on FG is applied to all the EURO STOXX 50 traded equities and pairs equity portfolios, to explore the risk-return characteristics of pairs trading strategies that arise from cointegrated assets. Empirical results demonstrate that cointegration-based pairs strategies generate positive abnormal profits, and deliver superior Sharpe ratios relative to the correlation-based pair strategies proposed by GGR and to the used benchmark market portfolios. These findings are also robust out of sample and after accounting for transaction costs.

C105: **Predictive systems under economic constraints**
Presenter: **Maxime Bonelli**, Inria Sophia Antipolis research centre, France
Co-authors: Daniel Mantilla-Garcia

A variation of a predictive system is proposed, incorporating the economically motivated assumption that equity expected returns are unlikely to be negative. This belief is introduced through a different model for the unobservable expected returns process. The modification has implications on the way past realized returns impact expected returns, on the variance and persistence of expected returns and on the conditional variance of returns produced by the system. The implications of the modified system are consistent with well established empirical facts and intuition. Empirical tests show that the out-of-sample performance of the system can improve compared to a system allowing negative expected returns and relative to the historical return mean.

C609: **Research of the determinants of the systemic importance of global banks**
Presenter: **Maiya Anokhina**, HSE, Russia
Co-authors: Henry Penikas

The increased role of financial institutions in the economy leads to a need to determine those that are systemically important. The bankruptcy of such institutions creates negative effects for the economy on the global scale. The aim is to identify important financial coefficients that can be used in the methodology of identification of G-SIB. Models of binary choice and models of ordered choice are used in this article, several models are highly predictive. Besides this paper has revealed several financial coefficients, that helped to find the probabilities of G-SIB for Russian banks. The list of systemically important banks compare with the one suggested by the Central Bank.

Sunday 7.12.2014 14:45 - 16:25 Parallel Session L – ERCIM

ESI04 Room Sala Convegni STATISTICAL ANALYSIS OF EVENT TIMES Chair: Thomas Scheike

E867: **Analysis of stratified mark-specific proportional hazards models with missing marks, with an application**
Presenter: **Yanqing Sun**, University of North Carolina at Charlotte, United States
Co-authors: Peter Gilbert

An objective of randomized placebo-controlled preventive HIV vaccine efficacy trials is to assess the relationship between the vaccine effect to prevent infection and the genetic distance (marks) of the exposing HIV to the HIV strain(s) represented in the vaccine construct. A high percentage of genetic marks of interest may be missing due to rapid evolution of transmitted HIV strains. The research investigates the stratified mark-specific proportional hazards model with missing marks where the baseline functions may vary with strata. Two-stage efficient estimation and hypothesis testing procedures are developed. The asymptotic properties and finite-sample performances of the procedures are investigated, demonstrating double-robustness and effectiveness of the predictive auxiliaries to recover efficiency. The methods are applied to the RV144 vaccine trial.

E1033: **Mechanistic survival regression models in the presence of long-term survival**
Presenter: **Francisco Louzada**, University of Sao Paulo, Brazil
Co-authors: Francisco Louzada

Some mechanistic survival models are discussed in presence of a proportion of individuals which are not susceptible to the occurrence of the event of interest. Our approach enables different underlying activation mechanisms, which lead to the event of interest, and it capable of accommodating physic and structural characteristics observed in practice, as well as characteristics of the treatment, such as the number of doses, the time interval between doses and the efficiency of each dose. It assumes that the number of causes of the occurrence of the event of interest is latent, in the sense that it is unobserved. As an advantage, our approach scans all underlying activation mechanisms from the first to last one based on order statistics. The usefulness of our modeling is illustrated on real Brazilian data on breast cancer and recurrent malaria episodes.

E1304: **Variable selection and deselection in Cox models, with confidence**
Presenter: **Jelle Goeman**, Leiden University Medical Center, The Netherlands

In regression models, variable selection methods are typically pragmatic, selecting a subset of covariates that are relevant for a response variable of interest. Covariates that are not selected are discarded, and the selected subset is often treated as if it was given a priori. It is a statisticians' task, however, to quantify uncertainty, and variable selection is not an error-free process. The question then arises: what is the number of truly relevant covariates present in the selected subset? We propose confidence statements for this quantity, using the closed testing procedure. The confidence statements we obtain are simultaneous for all possible subsets, so that they are not affected by the selection that is inherent in variable selection procedures. We apply our method to well-known examples in regression, focusing on the Cox proportional hazard model. An interesting, and perhaps surprising conclusion is that subsets of covariates selected by variable selection methods are typically not optimal from a confidence perspective.

ES01 Room N1 OODA-TREE STRUCTURED DATA OBJECTS Chair: Ana M. Aguilera

E593: **Principal components analysis in tree-space**
Presenter: **Megan Owen**, Lehman College CUNY, United States
Co-authors: Sean Cleary, Aasa Feragen, Daniel Vargas

Data generated in such areas as medical imaging and evolutionary biology are frequently tree-shaped, and thus non-Euclidean in nature. As a result, standard techniques for analyzing data in Euclidean spaces become inappropriate, and new methods must be used. One such framework is the space of metric trees constructed by Billera, Holmes, and Vogtmann. This space is non-positively curved (hyperbolic), so there is a unique geodesic path (shortest path) between any two trees and a well-defined notion of a mean tree for a given set of trees. Algorithms for finding a first principal component for a set of trees in this space have also been developed, but they cannot be used in an iterative fashion. We present the first method for computing multiple principal components, and apply it to a variety of datasets.

E422: **Brownian motion, bridges and inference in phylogenetic tree space**
Presenter: **Tom Nye**, Newcastle University, United Kingdom

Data sets consisting of samples of evolutionary trees, or *phylogenies*, for some fixed set of species arise in many different contexts in evolutionary biology. Analysing such data sets is challenging, since the space of all phylogenetic trees is highly non-Euclidean, although it has a natural geodesic metric. In particular, it is very difficult to construct tractable distributions on tree space, so many analyses have relied on minimising sums of squared geodesic distances. We propose an alternative approach by constructing probability distributions via random walks in tree space and their continuous-time limits. We consider particles undergoing Brownian motion in tree space from some fixed starting point, and their distribution at subsequent times. In order to fit such distributions to data, it is necessary to simulate random walks in tree space with fixed start and end points. A fast algorithm for simulating these walks, or *bridges*, is described. We then describe a Bayesian method for fitting distributions to samples of trees. This uses a Markov chain Monte Carlo scheme which relies heavily on the bridge simulation algorithm. These methods open up the possibility of fitting a variety of different models based on Brownian motion to samples of trees.

E507: **Learning on graphs**
Presenter: **Brijnesh Jain**, TU Berlin, Germany

Patterns are typically represented by feature vectors living in a Euclidean or Banach space, because those spaces provide powerful analytical techniques for data analysis, which are usually not available for other representations such as point patterns, trees, and graphs. A standard technique to solve a learning problem in vector spaces consists in setting up a smooth error function, which is then minimized by using local gradient information. We suggest a representation of attributed graphs as points of some quotient space, called graph orbifold. The graph orbifold framework enables us to transfer geometrical concepts such as length and angle as well as analytical concepts such as the derivative from vector spaces to the graph domain. We show, for example, that the gradient of a function on graphs is a well-defined graph pointing in direction of steepest ascent. Exploiting the analytical and geometrical properties of graph orbifolds, it will turn out that the principle of empirical risk minimization with differentiable loss function amounts to an optimization problem of locally Lipschitz risk functions on graphs that can be minimized with subgradient-like methods. We present and discuss some results in supervised and unsupervised learning on graphs showing that the orbifold framework complements existing learning methods on graphs.

E785: **kdetrees: nonparametric estimation of phylogenetic tree distributions**
Presenter: **Ruriko Yoshida**, University of Kentucky, United States

While the majority of gene histories found in a clade of organisms are expected to be generated by a common process (e.g. the coalescent process), it is well-known that numerous other coexisting processes (e.g. horizontal gene transfers, gene duplication and subsequent neofunctionalization) will cause some genes to exhibit a history quite distinct from those of the majority of genes. Such "outlying" gene trees are considered to be biologically interesting and identifying these genes has become an important problem in phylogenetics. We propose and implement kdetrees, a nonparametric method of estimating distributions of phylogenetic trees, with the goal of identifying trees which are significantly different from

the rest of the trees in the sample. Our method compares favorably with a similar recently-published method, featuring an improvement of one polynomial order of computational complexity (to quadratic in the number of trees analyzed), with simulation studies suggesting only a small penalty to classification accuracy. Application of kdetrees to a set of Apicomplexa genes identified several unreliable sequence alignments which had escaped previous detection, as well as a gene independently reported as a possible case of horizontal gene transfer. We also analyze a set of *Epichloë* genes, fungi symbiotic with grasses, successfully identifying a contrived instance of paralogy.

ES08 Room I1 STATISTICS FOR IMPRECISE-VALUED DATA Chair: Angela Blanco-Fernandez

E1276: **Statistical tests for inclusion/exclusion of Aumann means**
Presenter: **Raffaello Seri**, University of Insubria, Italy

A method was recently proposed to obtain a confidence set for the Aumann mean of a random closed set (RACS). We adapt the technique to provide statistical tests for inclusion/exclusion of the Aumann mean with respect to a given set C. This is easily done exploiting the duality between statistical tests and confidence sets. However, in this context it is often the case that the set C depends itself on the Aumann mean of the RACS. The most prominent example of this kind arises when C is a sphere whose radius depends on the volume of the Aumann mean. We extend the statistical test suggested by duality by considering this possibility. This allows one to test hypotheses associated with rotational invariance of the Aumann mean.

E1283: **Mediterranean diet: an interval-valued approach**
Presenter: **Marta Garcia-Barzana**, University of Oviedo, Spain
Co-authors: Ana Colubi, Pablo Camblor, Jose Ramon Quiros

In the literature, many epidemiological studies have investigated the protective effect of the Mediterranean Diet (MD) on mortality rates and risk of developing other health problems (cardiovascular diseases, diabetes and cancer among others). However, the definition of the adherence to the MD is still under discussion, as there is a high number of scores trying to express numerically this adherence according to the researchers goals. The proposal is to work with an interval containing most of the scores, so that no information is missing, and to adjust a Cox regression model to deal with this new interval framework. A real-case example showing the effect that the MD has over patient with heart attacks is presented.

E1285: **Multi-sample test for the equality of real-valued distributions based on the fuzzy representation**
Presenter: **Angela Blanco-Fernandez**, University of Oviedo, Spain
Co-authors: Ana B. Ramos-Guajardo

The general problem of testing the equality of distributions of several real-valued random variables is addressed from a new perspective. When the condition of normality for the distributions is violated, classical ANOVA techniques are not applicable and non-parametric hypothesis testing procedures are to be considered. Recent statistical techniques for random fuzzy sets have been developed to characterize the distribution of a real-valued random variable through its associated fuzzy representation. The aim is to profit from this characterization to test the equality of real-valued distributions by means of the corresponding fuzzy representations, and to apply powerful statistical techniques for random fuzzy elements to solve the inferences. Besides the theoretical validity of the method, some application examples and simulation studies are carried out to show the empirical performance of the procedures as well as to compare the results with classical approaches.

E1155: **SMART fuzzy weighted averages of fuzzy numbers**
Presenter: **Andrea Capotorti**, Universita degli Studi di Perugia, Italy
Co-authors: Gianna Figa-Talamanca

A proposal of two different weighted fuzzy averages between fuzzy numbers is given. Operators profit from α-cuts and LR representations of fuzzy numbers. One operator is intended to generalize, through specific deformations of standard fuzzy means, the disjunction and the other to generalize the conjunction. Generalizations emphasize agreement or not between different sources of information. Such conflicts, as well as agreements, are endogenously embedded inside the average weights of the two new operators by measuring distances or superimposition between α-cuts. No exogenous elements are added, except for the choice of the parameters of the deformation that emphasizes conflicts. The proposal is motivated by the practical problem of assessing the fuzzy volatility parameter in the Black and Scholes environment via both the historical volatility and the VIX estimators. Emphasis is posed on the consequences of the new operators on the fuzzy option pricing both in uni-periodal binary and in Black and Scholes. Crisp bid-ask price intervals are compared with fuzzy prices obtained through both new operators and standard fuzzy mean. Such comparisons are based on proper similarity indexes. The methodology evidences its meaning when a full process of elicitation of parameters that encompasses data and expert evaluation is adopted.

ES28 Room C1 RECENT ADVANCES IN STATISTICAL GENETICS Chair: Florian Frommlet

E188: **Multiscale DNA partitioning: Statistical evidence for segments**
Presenter: **Andreas Futschik**, JKU Linz, Austria
Co-authors: Thomas Hotz, Axel Munk, Hannes Sieling

DNA segmentation, i.e. the partitioning of DNA in compositionally homogeneous segments, is a basic task in bioinformatics. Different algorithms have been proposed for various partitioning criteria such as GC content, local ancestry in population genetics, or copy number variation. A critical component of any such method is the choice of an appropriate number of segments. Some methods use model selection criteria, and do not provide a suitable error control. Other methods that are based on simulating a statistic under a null model provide suitable error control only if the correct null model is chosen. We focus on partitioning with respect to GC content and propose a new approach that provides statistical error control: it guarantees with a user specified probability that the number of identified segments does not exceed the number of actually present segments. The method is based on a statistical multiscale criterion, rendering this as segmentation method which searches segments of any length (on all scales), simultaneously. It is also very accurate in localizing segments: under bench-mark scenarios, our approach leads to a segmentation that is more accurate than the approaches discussed in a previous comparative review. In our real data examples, we find segments that often correspond well to the available genome annotation.

E445: **Model selection approach for genome wide association studies in admixed populations**
Presenter: **Malgorzata Bogdan**, Wroclaw University of Technology, Poland
Co-authors: Florian Frommlet, Piotr Szulc, Hua Tang

Large number of markers used in Genome Wide Association Studies (GWAS) requires substantial adjustments for multiple testing and results in a relatively low power of detection of influential genes. In admixed populations, one can locate influential genes by using admixture mapping, where the information on the genotypes of genetic markers is replaced with the information on the ancestry of a given region of the genome. Due to the strong correlation between ancestry states in the neighboring loci, the multiple testing correction for the admixture mapping can be substantially less stringent than in case of GWAS. This advantage is however counterbalanced by the non-perfect correlation between the genotype and the ancestry state. To utilize the strength of both approaches, we propose an extension of the modified Bayesian Information Criterion, which works with the design matrix including the dummy variables both for the genotypes and the ancestry. Our simulation studies show that the proposed method allows us to control the false discovery rate at a reasonable level. We also show that including the ancestry variables helps to detect influential genes in the regions of a low linkage disequilibrium without compromising the power of detection of other genes.

E566: **Using mixed model analysis to describe and predict complex traits**
Presenter: **Doug Speed**, UCL, United Kingdom
Co-authors: Michael Johnson, David Balding
In 2010 it was shown how it was possible, by applying mixed model analysis to genome-wide SNP data for unrelated individuals, to calculate the total proportion of phenotypic variance explained by common variants. The results demonstrated that common SNPs explain a large proportion of heritability for many complex traits, but typically that the traits are highly polygenic, with many causal variants of small effect size. We have developed extensions of mixed model analysis to provide methods to further investigate the genetic architecture of complex traits; these include estimating the number of causal variants, examining overlap between traits and dividing heterogeneous traits into subtypes. We also present our new prediction method MultiBLUP, which offers improvements over existing methods both in terms of prediction accuracy and computational performance.

E776: **Unified gene-gene interaction analysis of multiple phenotypes**
Presenter: **Taesung Park**, Seoul National University, Korea, South
Co-authors: Ik-Soo Huh, Min-Seok Kwon, Wenbao Yu
Despite of many successful results from genome-wide association studies (GWAS), only a small number of genetic variants tend to be identified and replicated given a stringent genome-wide significance criterion. Furthermore, in many GWAS, one binary trait is commonly used which is derived from multiple quantitative traits. We consider a multivariate gene-gene interaction approach which uses all information about multiple phenotypes and identifies focus gene-gene interaction effects on the multiple phenotypes. Generalized multifactor dimensional reduction (GMDR) method has been commonly used in identifying gene-gene interactions. We propose a unfied framework based on the multivariate GMDR approach in order to identify gene-gene interaction for the multiple phenotypes. Our proposed multivariate GMDR method uses multiple phenotypes simultaneously. We apply the multivariate GMDR approach to a GWA dataset of a Korean population.

ES40 Room H1 NEW APPROACHES IN SPATIAL STATISTICS Chair: Tatiyana Apanasovich

E736: **An ANOVA-type procedure for replicated spatio-temporal point patterns**
Presenter: **Jorge Mateu**, University Jaume I, Spain
Co-authors: Jonatan Gonzalez, Ute Hahn
Several methods to analyse structural differences between groups of replicated spatio-temporal point patterns are presented. We calculate a number of functional descriptors of each spatio-temporal pattern to investigate departures from completely random patterns, both among subjects and groups. The statistical distributions of our functional descriptors and of our proposed tests are unknown, and thus we use bootstrap and permutation procedures to estimate the null distribution of our statistical test. A simulation study provides evidence of the validity and power of our procedures. An application to forestry data is finally presented.

E788: **Flexible modelling of turbulent in-flows using radial basis functions**
Presenter: **Ioanna Manolopoulou**, University College london, United Kingdom
Co-authors: Serge Guillas, Peter Moonen
Turbulent flows are ubiquitous, but they are also unsteady, irregular, seemingly random and chaotic. Accurate knowledge of the flow field is essential in order to advance our understanding of turbulent flows, but direct solution of the Navier-Stokes equations, i.e. the equations which govern turbulent flows, is not possible for applications with practical relevance due to the prohibitively high computational cost. Instead, numerical methods are often preferred. Large Eddy Simulation (LES) is a powerful technique based on filtering of the Navier-Stokes equations, but its widespread application is hampered by the difficulty to generate appropriate initial and boundary conditions which accounts for 50-80% of the computational time (using existing methods such as driver-domain simulations, vortex-spot methods and spectral synthesizers). This presents great potential for improvement. We propose modelling the vortices as spatial structures evolving through time according to the laws of physics (mostly the decay of energy from Kolmogorov's law and the shear stress breaking large and weak vortices). The vorticity field is modelled through a non-parametric radial basis function approach using Normal bivariate forms in space. The resulting LES has the potential to tackle a greater diversity of situations in much reduced computational time.

E820: **Computational methods for spatially Markovian stochastic PDEs**
Presenter: **Finn Lindgren**, University of Bath, United Kingdom
Large scale space-time environmental data present many modelling and computational challenges. Spatial non-stationarity and temporal non-separability of global temperatures can be modelled with stochastic PDEs driven by random noise processes. By converting to a Markov random field representation and using the multi-scale model structure to split the computations into manageable sub-problems, realistic reconstruction and uncertainty estimates can be obtained.

E1284: **Modeling space anisotropy and non-stationarity via Partial Differential Equations**
Presenter: **Laura Sangalli**, Politecnico di Milano, Italy
A novel approach to spatial statistics is proposed consisting in regression with regularizations based on Partial Differential Equations (PDE). The PDE is used to model the space variation of the phenomenon, using problem-specific information. The PDE modeling allows for important modeling flexibility in this respect, accounting also for space anisotropy and non-stationarity in a straightforward way, as well as unidirectional smoothing effects. The method is illustrated in various applied contexts.

ES42 Room P1 BAYESIAN ANALYSIS OF GRAPHICAL MODELS Chair: Michele Guindani

E134: **Genotype-environment effects analysis using Bayesian networks**
Presenter: **Marco Scutari**, University College London, United Kingdom
Co-authors: Alison Bentley, Ian Mackay
Models for genome-wide prediction and association studies usually target a single phenotypic trait. However, in animal and plant genetics it is common to record information on multiple phenotypes for each genotyped individual across several trials subject to different environmental effects. Disregarding one or more of these factors ignores potential associations due to pleiotropy, shared biological basis and genotype-environment interactions, thus providing only a partial, confounded view of interactions. Bayesian networks provide a convenient and interpretable framework for modelling the interplay between all these factors, and versatile tools for subsequent inference.

E312: **Zodiac: A comprehensive depiction of genetic interactions in cancer by integrating TCGA data**
Presenter: **Yuan Ji**, NorthShore University HealthSystem University of Chicago, United States
The Cancer Genomes Atlas (TCGA) data are unique in that multimodal measurements across genomics features, such as copy number, DNA methylation, and gene expression, are obtained on matched tumor samples. The multimodality provides an unprecedented opportunity to investigate the interplay of these features. Graphical models are powerful tools for this task that address the interaction of any two features in the presence of others, while traditional correlation- or regression-based models cannot. We introduce Zodiac, an online resource consisting of two main components, 1) a large database containing nearly 200 million interaction networks of multiple genomics features produced by applying novel Bayesian graphical models on TCGA data through massively parallel computation, and 2) analytics tools that perform high-quality inference on

data-enhanced networks. Setting a new way of integrating TCGA data, Zodiac, publically available at www.compgenome.org/ZODIAC, is expected to facilitate the generation of new knowledge and hypotheses by the community.

E538: **Bayesian regularized graphical model estimation in high dimensions**
Presenter: **Suprateek Kundu**, Emory University, United States
Co-authors: Veerabhadran Baladandayuthapani, Bani Mallick

There has been an intense development of Bayesian graphical model estimation approaches over the past decade - however, most of the existing methods are restricted to moderate dimensions. We propose a novel approach suitable for high dimensional settings, by decoupling model fitting and covariance selection. First, a full model based on a complete graph is fit under novel class of continuous shrinkage priors on the precision matrix elements, which induces shrinkage under an equivalence with Cholesky-based regularization while enabling conjugate updates of entire precision matrices. Subsequently, we propose a post-fitting graphical model estimation step which proceeds using penalized joint credible regions to perform neighborhood selection sequentially for each node. The posterior computation proceeds using straightforward fully Gibbs sampling, and the approach is scalable to high dimensions. The proposed approach is shown to be asymptotically consistent in estimating the graph structure for fixed p when the truth is a Gaussian graphical model. Simulations show that our approach compares favorably with Bayesian competitors both in terms of graphical model estimation and computational efficiency. We apply our methods to high dimensional gene expression and microRNA datasets in cancer genomics.

E1266: **miRNA-target gene regulatory networks: a Bayesian approach to biomarker selection with application to kidney cancer**
Presenter: **Kim-Anh Do**, University of Texas MD Anderson Cancer Center, United States
Co-authors: Francesco Stingo, Thiery Chekouo, James Doecke

The availability of cross-platform, large-scale genomic data has enabled the investigation of complex biological relationships for many cancers. Identification of reliable cancer-related biomarkers requires the characterization of multiple interactions across complex genetic networks. MicroRNAs are small non-coding RNAs that regulate gene expression; however, the direct relationship between a microRNA and its target gene is difficult to measure. We propose a novel Bayesian model to identify microRNAs and their target genes that are associated with survival time by incorporating the microRNA regulatory network through prior distributions. We assume that biomarkers involved in regulatory networks are likely to be associated with survival time. We employ non-local prior distributions and a stochastic search method for the selection of biomarkers associated with the survival outcome. We use KEGG pathway information to incorporate correlated gene effects within regulatory networks. Using simulation studies, we assess the performance of our method, and apply it to experimental data of kidney renal cell carcinoma (KIRC) obtained from The Cancer Genome Atlas. Our novel method validates previously identified cancer biomarkers and identifies biomarkers specific to KIRC progression that were not previously discovered. Using the KIRC data, we confirm that biomarkers involved in regulatory networks are more likely to be associated with survival time, showing connections in one regulatory network for five out of six such genes that we identified.

ES44 Room L1 MODEL-BASED CLUSTERING FOR CATEGORICAL AND MIXED-TYPE DATA Chair: Antonio Punzo

E096: **Model-based clustering of multivariate ordinal data relying on a stochastic binary search algorithm**
Presenter: **Julien Jacques**, University Lille I and INRIA, France
Co-authors: Christophe Biernacki

A new probability distribution for univariate ordinal data is introduced and used to define a model-based clustering algorithm for multivariate ordinal data. This distribution, governed by two position and dispersion parameters, is obtained by modeling the data generating process, assumed to be a stochastic binary search algorithm in a sorted table. The main interest of this approach is to give a specific model for ordinal data, without any reference to numerical or nominal data, as it is often the case for competitors. The resulting distribution is easily estimated by an EM algorithm. Using then the classical latent class assumption, a model-based clustering algorithm for multivariate ordinal data is proposed as a direct extension of the previous univariate ordinal model. Again, its parameters can be estimated by an EM algorithm. Simulated and real data sets illustrate potentiality of the new model.

E177: **A multivariate latent variable model for analysing longitudinal mixed data**
Presenter: **Cinzia Viroli**, University of Bologna, Italy
Co-authors: Silvia Cagnone

A latent variable model for the analysis of multivariate mixed longitudinal data is proposed. The model is based on the introduction of two hidden variables: a continuous latent variable for modelling the association among the observed variables at each time point and a latent discrete variable that follows a first-order Markov chain with the aim of taking into account the unobserved heterogeneity. The aim of the proposed model is twofold: it allows us to perform dimension reduction when data are of mixed type and it performs model based clustering in the latent space. We derive an EM algorithm for the maximum likelihood estimation of the model parameters. The method is illustrated by an application to a longitudinal dataset on aging and health.

E770: **A partial membership clustering of categorical data**
Presenter: **Yang Tang**, University of Guelph, Canada
Co-authors: Ryan P Browne, Paul D McNicholas

A framework for modelling partial membership of categorical data to clusters is presented. Unlike a standard mixture model which assigns each observation to one and only one mixture component, a partial membership model assumes that each observation belongs to more than one cluster simultaneously. We use a latent trait model for each component, and a product of these components with weighted parameters to model each observation. A variational approximation to the likelihood is exploited to derive a fast algorithm for determining the model parameters. Our approach is demonstrated on real and simulated data and their performance is discussed.

E786: **Mixture models for mixed-type data through a pairwise likelihood approach**
Presenter: **Monia Ranalli**, Sapienza University of Rome, Italy
Co-authors: Roberto Rocci

A mixture model is proposed to classify continuous and/or ordinal variables. The latter are considered as a discretization of an underlying finite mixture of Gaussians, whose parameters depend on the model specified for the continuous data. Indeed the likelihood can be decomposed into two factors: the first corresponds to the observed Gaussian mixture for the continuous variables, while the second one to the likelihood for the ordinal variables given the continuous ones. Since the last likelihood function involves multidimensional integrals, whose evaluation is computationally demanding as the number of observed variables increases, the proposal is to replace this cumbersome likelihood with a surrogate objective function that is easier to maximize. A pairwise likelihood is used and the estimation of the model parameters is carried out within an EM-like algorithm. In order to classify the objects, the joint posterior probabilities are approximated combining the a posteriori probabilities for the continuous data with the pairwise posterior probabilities of the surrogate function by running an Iterative Proportional Fitting algorithm. The effectiveness of the proposal is investigated through a simulation study. Finally, an application to real data is illustrated.

ES51 Room Q1 BAYESIAN SEMI- AND NONPARAMETRIC MODELLING III Chair: Li Ma

E775: **Multivariate regression with additive Gaussian processes and its applications**
Presenter: **Surya Tokdar**, Duke University, United States

Numerous statistical applications involve accurately forecasting an outcome, identifying its important predictors and quantifying how it would react to an intervention. Achieving these interrelated goals simultaneously is extremely challenging when the number of predictors is large. Popular approaches either impose linearity/additivity conditions which diminish accuracy or employ ensemble learners with a poorer quantification of predictor importance and prediction uncertainty; many are prone to introducing bias when quantifying intervention effect in non-experimental studies. We seek solutions under a promising new regression model where total predictor effect equals the sum of smaller effects each involving a limited number of interacting predictors. A Bayesian implementation is pursued by using additive Gaussian processes to create a novel multivariate regression tool that efficiently captures a wide range of naturally occurring predictor-response relations, identifies important predictors, and their interactions, enables intervention effect quantification in nonrandomized studies and, most importantly, achieves all these when the number of predictors is large or extra large but the sample size remains limited. We demonstrate the new tool's high impact applications to QTL fine mapping and public policy research.

E216: **Recent advances for profile regression models in the R package PReMiuM**
Presenter: **Silvia Liverani**, Brunel University, United Kingdom

Profile regression is a Dirichlet process mixture model which co-clusters covariates and responses as an alternative to parametric regression. The clusters are determined by both the covariates and the response allowing for implicit handling of potentially high dimensional interactions which would be very difficult to capture in traditional regression. The approach also allows for 'fixed effects' which have a global (ie. non-cluster specific) effect on the response. This method has been recently implemented in the R package PreMiuM. The package can be used for a large range of datasets thanks to the flexibility of the model and its implementation: it can accommodate continuous, count and discrete data, and it also includes variable selection. We will discuss some of the most recent developments in profile regression. We have extended the model to allow for survival responses, modelled with a Weibull distribution and allowing for type-I censored observations. To account for possible spatial correlation within the residuals we have included a spatial conditional autoregressive term in the response. Moreover, we will discuss the difficulties of achieving good mixing in MCMC samplers of this nature and investigate sensitivity to initialisation. We introduce a new label switching move that helps to surmount these difficulties. We will illustrate the methods with simulated data and real epidemiological data.

E225: **A Bayesian nonparametric approach to reconstruction of dynamic and clinical PET data**
Presenter: **Mame Diarra Fall**, University of Orleans, France

A Bayesian nonparametric (BNP) method for reconstructing dynamic Positron Emission Tomography (PET) images from clinical data is proposed. PET is a medical imaging modality providing in vivo images of functional processes. The goal is to extend our preliminary BNP model for PET reconstruction based on simulated data. We have proposed to substitute the finite and fixed basis of voxels and temporal functions usually used with an infinite dimensional functional reconstruction basis suited to the available data sets. The nonparametric approach allows us to directly infer on the activity distribution of interest without any ad hoc space or time discretization. Being fully Bayesian, we have access to the entire distribution and not only to a point estimate. As a consequence, any functional (e.g., variance, credible intervals) on a given region of interest can be estimated. However, reconstruction of real data requires to deal with all physical artefacts and then weighting the statistical modeling. Furthermore, the specific scanner geometry must be taken into account. Results on cardiac data highlight that the proposed reconstruction scheme, in a context of low dose (small samples implementation) outperforms the conventional reconstruction used in clinical platforms and under standard dose.

E208: **Multiscale Bernstein polynomials for densities**
Presenter: **Antonio Canale**, University of Turin, Italy
Co-authors: David Dunson

The focus is on constructing a multiscale nonparametric prior for densities. The Bayes density estimation literature is dominated by single scale methods, with the exception of Polya trees, which favor overly-spiky densities even when the truth is smooth. We propose a multiscale Bernstein polynomial family of priors, which produce smooth realizations that do not rely on hard partitioning of the support. At each level in an infinitely-deep binary tree, we place a beta dictionary density; within a scale the densities are equivalent to Bernstein polynomials. Using a stick-breaking characterization, stochastically decreasing weights are allocated to the finer scale dictionary elements. A slice sampler is used for posterior computation, and properties are described. The method characterizes densities with locally-varying smoothness, and can produce a sequence of coarse to fine density estimates. An extension for Bayesian testing of group differences is introduced and applied to DNA methylation array data.

ES79 Room D1 COUNTING PROCESSES Chair: Paula R. Bouzas

E146: **Multivariate survival models with common baseline risk under event dependence and unknown number of previous episodes**
Presenter: **David Morina Soler**, Centre for Research in Environmental Epidemiology, Spain
Co-authors: Georgina Casanovas, Albert Navarro

Recurrent event data refers to situations where the subject can experience repeated episodes of the same type of event, such as asthma attacks. If these phenomena are studied via a cohort, the application of survival methods would seem appropriate. Recurrent events present two problems which cannot be handled using the standard methods. The first is individual heterogeneity, usually tackled using frailty models. The second problem refers to the within-subject correlation. Event dependence is tackled through the application of models employing baseline hazards specific for the episode to which the individual is at risk. Among these models, the most widely used is that proposed by Prentice, Williams and Peterson, an extension of the Cox proportional hazards model. On the basis of the obtained results, the use of survival models with common baseline hazard appears only to be appropriate in situations of low levels of event dependence, and when the effect to be estimated is small. In general these models cannot be used in the present context as their levels of coverage of the true parameter by the 95% CI, and levels of compliance with the assumption of proportional hazards, are extremely low and considerably overestimate the effect of exposure.

E270: **Modelling and generating dependent nonhomogenous point processes**
Presenter: **Ana C Cebrian**, University of Zaragoza, Spain

The modelling of dependent point processes, and in particular Poisson processes, is a usual problem in insurance and environmental studies. A frequent approach in operational risk analysis is based on multivariate point processes in R^d with dependent marginal processes. However, this approach does not cover all the possible types of dependence between d point processes in R. Hence, several approaches for generating nonhomogenous processes with different dependence structures are suggested. The first generates nonhomogenous Poisson processes based on tandems of $M \setminus M \setminus 1$ queues, with exponential services. The second approach generates dependent point processes marking a Poisson process with a Markov chain of d states. The third one generates dependent nonhomogenous Neyman-Scott processes with the same process of cluster centers, which is useful in situations where the same shock triggers events on different processes. Although there is no general way of measuring dependence between nonhomogeneous point processes, some alternatives for quantifying it are suggested. These construction methods provide a broad palette for modelling dependent point processes.

E400: **Self-exciting process with phase-type intensity: simulation and estimation**
Presenter: **Nuria Ruiz-Fuentes**, University of Jaen, Spain

Co-authors: Paula R. Bouzas, Carmen Montes-Gijon

Over the last two decades, counting processes with stochastic intensities have been more extensively studied and applied. The intensity of these processes is usually assumed to follow a given stochastic model. The work deals with a type of counting process, the self-exciting process, which generalizes many others. Some of the counting statistics pertaining to this process are estimated merely from sample paths observed in discrete time points using a point estimator of the count-conditional intensity. As phase-type distributions generalize many non-negative distributions (every discrete one with finite support, Erlang, linear combinations of exponentials, etc.), it is reasonable to think of them as models of intensity. In order to work with a general self-exciting process whose intensity is stochastic depending on the past and the time, the intensity proposed has phase-type marginals. Both the past and time condition these marginals. This model is simulated to illustrate the estimations of the count-conditional intensity, the probability mass function, the mean and the mean of the intensity process. In addition, the estimation errors are studied and conclusions drawn.

E528: **Forecasting compound Cox processes**

Presenter: **Paula R. Bouzas**, University of Granada, Spain

Co-authors: Nuria Ruiz-Fuentes, Juan Eloy Ruiz-Castro

Compound Cox processes are highly flexible marked point processes due to the stochastic nature of their intensity. Some examples of these include Cox processes with random deletions or with simultaneous events and multichanel or time-space Cox processes. Assuming no stochastic model of the intensity, the aim is to state the expressions of counting and time statistics of the compound Cox process with specific marks in terms of the intensity and also of the mean process. These expressions are completed by virtue of the representation theorems and functional data analysis. This is an important fact because it allows for derivation of a method to forecast those statistics using principal components prediction models from the raw data of several discretely observed sample paths. The forecasting of some time statistics such as the survival conditional function, reaches intractable equations which are solved by imposing weak restrictions to the principal components of the mean process. Simulations of compound Cox processes illustrate the forecasting methodology.

ES90 Room A1 ROBUST AND NONPARAMETRIC METHODS Chair: Stefan Van Aelst

E287: **Robust exploratory data analysis of compositional tables**

Presenter: **Kamila Facevicova**, Palacky University Olomouc, Czech Republic

Co-authors: Karel Hron, Matthias Templ, Valentin Todorov

Compositional tables represent a special case of compositional data as positive vectors and continuous counterpart of contingency tables, which carry relative information about relationships between row and column factors, thus for statistical analysis of compositional tables only ratios between cells are informative. The standard Euclidean geometry is not appropriate for this purpose and should be replaced by the Aitchison geometry on the simplex that fully accounts for their specific features. Moreover, the Aitchison geometry enables us to decompose the compositional table into its independent and interactive parts; in particular, the latter one is crucial for further investigation of the relationship between row and column factors. Furthermore, by assigning orthonormal coordinates to decomposition of compositional tables, standard methods for their statistical processing can be applied. In practice, the classical statistical analysis of a sample of compositional tables can be substantially affected by the presence of outlying observations, thus a robust alternative is needed. The contribution deals with the problem of how robust explorative analysis like outlier detection and PCA can be applied to a sample of compositional tables, with a particular focus on proper choice of orthonormal coordinates for this purpose and necessary assumptions for the corresponding robust estimators.

E295: **Posterior inference in Bayesian quantile regression with asymmetric Laplace likelihood**

Presenter: **Yunwen Yang**, Drexel University, United States

Co-authors: Huixia Wang, Xuming He

Quantile regression is gradually developing into a systematic tool for analyzing the conditional quantile functions of a response variable given covariates. In recent years, Bayesian quantile regression has attracted more and more attention. A typical quantile regression model does not specify a parametric likelihood, thus a working likelihood is needed for Bayesian quantile regression. A popular option of the working likelihood is the asymmetric Laplace (AL) distribution. We discuss the asymptotic validity of posterior inference of Bayesian quantile regression methods with complete or censored data, when an AL likelihood is used. The Bayesian AL quantile regression methods utilize an efficient Markov Chain Monte Carlo algorithm for estimation. However, our asymptotic results suggest that the posterior chain from the Bayesian AL quantile regression does not lead to valid posterior inference. We propose a simple correction to the covariance matrix of the posterior chain to enable asymptotically valid posterior inference.

E454: **Robust estimators in a generalized partly linear regression model under mononicity constraints**

Presenter: **Graciela Boente**, Universidad de Buenos Aires and CONICET, Argentina

Co-authors: Daniela Rodriguez, Pablo Vena

The situation in which the observations follow an isotonic generalized partly linear model is considered. Under this model, the mean of the responses is modeled, through a link function, linearly on some covariates and nonparametrically on an univariate regressor in such a way that the nonparametric component is assumed to be a monotone function. A class of robust estimates for the monotone nonparametric component, and for the regression parameter, related to the linear one, is defined. The robust estimators are based on a spline approach combined with a score function which bounds large values of the deviance or Pearson residuals. The empirical influence function allows us to study the sensitivity of the estimators to anomalous observations. We investigate the performance of the estimators through a Monte Carlo study for the Binomial and Gamma families of distributions.

E1018: **Robust estimation of multivariate scatter with fixed center**

Presenter: **Wannes Van den Bossche**, KULeuven, Belgium

Co-authors: Mia Hubert, Peter Rousseeuw

There exist many methods for robust estimation of multivariate location and scatter. But some situations call for a fixed center. In principle this is a simpler problem since no location needs to be estimated, but currently no software is readily available for this task. An algorithm is constructed to compute the Minimum Covariance Determinant (MCD) estimator with fixed center. Special care is taken to handle the possibility that the majority of the data lie in a lower-dimensional subspace. One application is extending the ROBPCA algorithm for robust principal components to the case with a fixed center. A second application is to the estimation of a common scatter matrix in robust discriminant analysis, and other applications are underway. The code will be made available in the Matlab toolbox LIBRA. Some real data examples will be shown.

ES105 Room G1 FLEXIBLE MODELS FOR DIRECTIONAL DATA Chair: Shogo Kato

E336: **Families of unimodal distributions on the circle**

Presenter: **Chris Jones**, The Open University, United Kingdom

The work starts on the real line, briefly describing some families of unimodal distributions with three or four parameters, controlling location, scale, skewness and perhaps some aspect(s) of tailweight. It will then address the question: "Can this technology be transferred to the case of distributions on the circle?" The answer is a qualified yes. In particular, I will describe three families of four-parameter unimodal circular distributions which arise, mostly from the relatively obscure linear notion of "transformation of scale" distributions. The first two will be "direct" and "inverse" versions of a type of distribution that can be found in Batschelet's book on circular statistics. The third, on which I shall principally focus, appears

to prove best of all, possessing numerous attractive properties; this family of circular distributions has been developed in collaboration with Shogo Kato (ISM, Tokyo).

E630: **Some properties of a family of distributions on the sphere related to the Mobius transformation**
Presenter: **Shogo Kato**, Institute of Statistical Mathematics, Japan
Co-authors: Peter McCullagh

Some properties of a family of distributions on the sphere are discussed. The family is a generalization of the wrapped Cauchy or circular Cauchy family on the circle. Its connection with the Mobius transformation is explored. It is shown that the family is closed under the Mobius transformation on the sphere of the sample space and that there is a similar induced transformation on the parameter space. The family is related to the conformal Cauchy distribution on the Euclidean space via the stereographic projection. Maximum likelihood estimation is studied in some detail. The unimodality of the likelihood function is shown for samples of size $n \geq 3$. Closed-form expression for the maximum likelihood estimator is available for $n \leq 3$. An extension of the family is derived using a generalized map of the Mobius transformation.

E748: **Skew-rotationally symmetric distributions on hyperspheres and optimal goodness-of-fit tests**
Presenter: **Thomas Verdebout**, Universite Libre de Bruxelles, Belgium
Co-authors: Christophe Ley

A new skew-rotationally-symmetric family of distributions on the unit hypersphere is provided. Then, we discuss some examples, and investigate the corresponding Fisher information matrix. The Le Cam approach is used to establish the uniform local asymptotic normality (ULAN) property of our general family of distributions. We deduce from it the forms of optimal parametric goodness-of-fit tests against skew-rotationally-symmetric local alternatives. The finite-sample properties of our test statistics are investigated via a Monte Carlo simulation experiment and a real data set is analyzed.

E345: **Linear-circular models and their properties**
Presenter: **Toshihiro Abe**, Tokyo University of Science, Japan
Co-authors: Christophe Ley, Angelo Efoevi Koudou

Cylindrical distributions are considered obtained by combining the sine-skewed von Mises distribution (circular part) with Generalized inverse Gaussian distributions (linear part). An immediate advantage of this construction is that the normalizing constant of the density can be expressed in terms of known functions. The properties of sub-models will be investigated, and parameter estimation based on the method of moments and maximum likelihood techniques will be discussed.

ES117 Room B1 HIGH-DIMENSIONAL BAYESIAN MODELING Chair: Jaeyong Lee

E403: **Using a subclass of log-linear models to search for dependencies among bacterial genomes with metagenomic NGS data**
Presenter: **Bertrand Clarke**, University of Nebraska at Lincoln, United States
Co-authors: Adrian Dobra, Jennifer Clarke, Camilo Valdes

Metagenomics is the study of genetic material recovered directly from a physical sample containing two or more genomes. When the genetic material is DNA, next generation sequencing (NGS) techniques can be applied to generate data typically consisting of many thousands of 'short reads'. In a bacterial context, the short reads may come from any of a known collection of bacterial genomes or an unrecognized bacterial genome. Since the short reads may be used to generate a large sparse contingency table, we search for statistical dependence among bacterial genomes using a subclass of log-linear models that have generating classes that are decomposable, have void separators, and components that are maximal. We call these clique log-linear models; their interaction terms encapsulate the possible dependencies among genomes. To find the dependency classes at the level of genera, species, and strains we explore the posterior on clique log-linear models using a simple MCMC approach. This approach seems to scale up to meaningful collections of genera, strains, and species and generates connectivity graphs to visualize the dependence. It can also be used to generate posterior probabilities of existence for individual members of genera, species and strains. Our procedure is illustrated using NGS data from the Human Microbiome Project.

E829: **Prediction and model selection for multi-task learning**
Presenter: **Feng Liang**, University of Illinois at Urbana-Champaign, United States
Co-authors: Bin Li

In multi-task learning one simultaneously fits multiple regression models. We are interested in inference problems like model selection and prediction when there are a large number of tasks. A simple version of such models is a one-way ANOVA model where the number of replicates is fixed but the number of groups goes to infinity. We examine the consistency of Bayesian procedures using Zellner's g-prior and its variants (such as mixed g-priors and Empirical Bayes), and compare their prediction accuracy with other procedures, such as the ones based AIC/BIC and group Lasso. Our results indicate that the Empirical Bayes procedure (with some modification for the large p small n setting) can achieve model selection consistency, and also have better estimation accuracy than other procedures being considered. We focus on the analysis on the one-way ANOVA model, but also give a summary on our findings for multi-tasking learning involving a more general regression setting.

E211: **Penalized regression, standard errors, and Bayesian lassos**
Presenter: **Minjung Kyung**, Duksung Womens University, Korea, South
Co-authors: Jeff Gill, Malay Ghosh, George Casella

Penalized regression methods for simultaneous variable selection and coefficient estimation, especially those based on the lasso of Tibshirani, have received a great deal of attention in recent years, mostly through frequentist models. Properties such as consistency have been studied, and are achieved by different lasso variations. We look at a fully Bayesian formulation of the problem, which is flexible enough to encompass most versions of the lasso that have been previously considered. The advantages of the hierarchical Bayesian formulations are many. In addition to the usual ease-of-interpretation of hierarchical models, the Bayesian formulation produces valid standard errors (which can be problematic for the frequentist lasso), and is based on a geometrically ergodic Markov chain. We compare the performance of the Bayesian lassos to their frequentist counterparts using simulations, data sets that previous lasso papers have used, and a difficult modeling problem for predicting the collapse of governments around the world. In terms of prediction mean squared error, the Bayesian lasso performance is similar to and, in some cases, better than, the frequentist lasso.

E866: **False discovery rate smoothing**
Presenter: **James Scott**, University of Texas at Austin, United States
Co-authors: Wesley Tansey

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 test statistics have a natural spatial lattice structure (voxels in the brain, distance along the chromosome), and where a combined analysis can lead to poorly calibrated error rates. To address this issue, we introduce an approach called false-discovery-rate (FDR) smoothing. FDR smoothing can, at modest computational cost, identify localized spatial regions of the underyling lattice structure where the frequency of signals is enriched versus the background. We introduce the method, describe a model-fitting algorithm based on an augmented-Lagrangian formulation of the problem, and show an application to fMRI data analysis in functional neuro-imaging.

ES119 Room M1 HIGH-DIMENSIONAL DATA ANALYSIS Chair: Juan Romo

E634: **Robust statistics for multivariate functional data: depth measures in cardiovascular disease prediction**
Presenter: **Anna Maria Paganoni**, MOX-Politecnico di Milano, Italy
Co-authors: Francesca Ieva, Nicholas Tarabelloni

The aim is to develop statistical methods to compare two independent samples of multivariate functional data that differ in terms of covariance operators. In particular, we generalize the concept of depth measure to this kind of data, exploiting the role of the covariance operators in weighting the components that define the depth. Two simulation studies are carried out to validate the robustness of the proposed methods. We present an application to Electrocardiographic (ECG) signals aimed at comparing physiological subjects and patients affected by Left Bundle Branch Block. The proposed depth measures computed on data are then used to perform a nonparametric comparison test among these two populations. They are also introduced into a generalized regression model aimed at classifying the ECG signals and predict the subject specific probabilities of being affected by Left Bundle Branch Block.

E656: **Classification of multivariate functional data based on depth**
Presenter: **Peter Rousseeuw**, KU Leuven, Belgium
Co-authors: Mia Hubert, Pieter Segaert

Recently, the notion of multivariate functional halfspace depth has been proposed. It considers a set of multivariate curves on the same time interval, and defines the depth of a curve as the (possibly weighted) integral of its halfspace depth at each time point. This depth function is applied to supervised classification. A new distance measure is based on it, from which a classifier is derived. Comparisons are made with the functional k-nearest neighbor method, the maximum depth classifier, and a method based on the depth-depth plot. Their behavior is also studied in the presence of outlying curves.

E726: **Efficiency in functional nonparametric models with autoregressive errors**
Presenter: **Sophie Dabo**, University-Lille, France
Co-authors: Serge Guillas, Camille Ternynck

A kernel-based procedure of estimation for a nonlinear functional regression is introduced in the case of a functional predictor and a scalar response. More precisely, the explanatory variable takes values in some abstract function space and the residual process is stationary and autocorrelated. The procedure consists in a pre-whitening transformation of the dependent variable. The main idea is to transform the original regression model, so that this transformed regression has a residual term that is uncorrelated. The asymptotic distribution of the proposed estimator is established considering that the explanatory variable is an $\alpha-$mixing process, the most general case of weakly dependent variables. Although, for the kernel methods proposed in the literature, it is generally better to ignore the correlation structure entirely ("working independence estimator"), it is shown that the autocorrelation function of the error process has useful information for improving estimators of the regression function. The skills of the methods are illustrated on simulations where the relative efficiency of the proposed efficient estimator over the conventional estimator is given for different values of the auto-regressive parameters.

E1060: **On the cosine of von Mises-Fisher distributions with statistical applications**
Presenter: **Mario Romanazzi**, Ca Foscari University of Venice, Italy

The von Mises - Fisher (vMF) distribution is a classical model for data constrained to lie on the surface of a hypersphere. The random vectors belonging to vMF family are indexed by the center $\mu = (\mu_1, ..., \mu_p)^T$, $\| \mu \| = 1$, and the concentration parameter $\kappa \geq 0$. Typical statistical applications deal with circular ($p = 2$, clock and compass data) or spherical data ($p = 3$, meteorological and astronomical data) but higher dimensions are increasingly encountered, e. g., compositional data via squared-root transformation, and text categorization and gene expression analysis via transformation to unit norm. From these research fields a demand is emerging for methods able to deal with vMF distributions in full generality. We extend previous work in discriminant analysis with vMF distributions to general dimension, allowing computation of misclassification probabilities and ROC curves. The key result is the probability distribution of the cosine transformation of a vMF distribution, that is, the random variable $U_a = a^T X$, where $X = (X_1, ..., X_p)^T$ is a random direction of $\mathcal{S}_p$ with vMF distribution and $a = (a_1, ..., a_p)^T$ is a fixed non-random direction of $\mathcal{S}_p$. This transformation is of general importance in multivariate analysis, in particular it underlies discriminant analysis both in two-group and multiple group problems.

ES132 Room O1 COMPUTATIONAL STATISTICS I Chair: Sonja Kuhnt

E999: **Run-time model for a class of parallel probabilistic statistical computations on graphical processing units**
Presenter: **Masanari Iida**, Tokyo University of Science, Japan
Co-authors: Naoto Niki

A class of statistical computation, including bootstrap and Monte Carlo approximation to sampling distributions of statistics for unknown and known populations, respectively, involves a large and stochastically varying number of repetitions of a fixed procedure. In spite of clear merit, parallelization on GPU (Graphic Processing Unit) for the class has been avoided by workers since, under SIMD (Single Instruction Multiple Data) parallel architecture build in GPU, (1) conditional branching has considerable overhead of time, i.e., every thread in a SIMD group executes both if and else blocks, either in vain, in case of mixed truth values; and (2) the maximum of random repeat counts observed in the threads dictates the execution time, i.e., the slowest thread keeps the others waiting and being unemployed. A run-time model is proposed in order to explain the empirical execution time distributions observed in our experiments and from which coarse grain threading is shown effective in parallel GPU computation of the class, unless if and else blocks are not heavily time-consuming, by revealing asymptotic properties of the model in addition to numerical results.

E1019: **Sparse SEM: computational identification of inter-factor relations in confirmatory factor analysis**
Presenter: **Kohei Adachi**, Osaka University, Japan
Co-authors: Nickolay Trendafilov

Confirmatory factor analysis (CFA) is a general technique to model causal relationships among latent factors typically expressed as a family of structural equation models (SEM). The causal links among factors are usually specified by the users in the ordinary SEM. We propose a new procedure called Sparse SEM (SpaSEM) in which the causal relationships are automatically identified. The key point of SpaSEM is to computationally obtain a sparse inter-factor coefficient matrix whose nonzero elements indicate which factors are mutually linked. Such a sparse matrix is found by solving a penalized version of the ordinary SEM log likelihood, where the factors are classified into explanatory and dependent ones. The penalized maximization locates the zero elements in the coefficient matrix by joint estimation of all involved unknown parameters. An EM algorithm for SpaSEM is presented and it is illustrated with real data examples.

E1147: **Parallelized EGO optimization of a sheet metal forming process via block additive Sobol index decomposition**
Presenter: **Momchil Ivanov**, TU Dortmund University, Germany
Co-authors: Sonja Kuhnt

The optimization of expensive black-box functions is often performed with metamodel-based sequential strategies. A popular choice is the efficient global optimization algorithm (EGO), known for its effectiveness. Unfortunately a big limitation of EGO is the fact that it allows only one

simulation at a time. Making use of several simulators simultaneously, and thus parallelizing the optimization, is a very appealing idea. An elegant parallel optimization procedure based on a sensitivity analysis technique - the FANOVA graph- is presented. Estimated by Sobol indices, the FANOVA graph method is able to detect the additive structure of a function. Given such an additive structure, often observed in practical applications, the additive blocks can be simultaneously optimized independently. This allows the experimenter to use several simulators and to perform a parallel optimization. The aims are to present the implementation of the parallelization technique and to discuss some of its issues. Finally the procedure is applied to a sheet metal forming simulation. Sheet metal forming is used for the production of automobile body parts and is unfortunately susceptible to defects such as tearing, wrinkling or spring back. It is shown that the new procedure allows for an efficient process optimization, leading to reduced sheet tearing.

E1170: **Solving augmented systems using a block-updating generalized QR decomposition**
Presenter: **Panagiotis Paoullis**, University of Oviedo, Spain
Co-authors: Erricos Kontoghiorghes

A computationally efficient serial algorithm for solving the augmented system, when the Hessian matrix is symmetric and positive semi-definite is proposed. The augmented system is an equivalent formulation of the equality constrained quadratic programming problem and derives the BLUE of General Linear Models with singular non-spherical dispersion matrix. The proposed algorithm is based on orthogonal transformations and exploits the triangular structure of the Cholesky factor of the Hessian matrix. Specifically, it computes the solution of an augmented system by solving recursively a series of smaller in dimension augmented systems. The theoretical computational complexity of the algorithm is derived and analyzed. The new algorithm is found to outperform significantly the corresponding LAPACK routine. Experimental results are presented which confirm the theoretical analysis. Furthermore, an adaptation of the algorithm for recomputing the solution of an augmented system after a *rank-k* updating of the Hessian matrix is considered.

ES37 Room F1 CHANGE-POINTS IN TIME SERIES II Chair: Daniel Vogel

E080: **Nonparametric method for change point detection in functional time series**
Presenter: **Martin Wendler**, Ruhr-University Bochum, Germany
Co-authors: Olimjon Sharipov, Johannes Tewes

The focus will be on limit theorems and nonparametric statistical methods for functional data modeled as random variables taking their values in a Hilbert space. The random variables are assumed to be weakly dependent in the sense of near epoch dependence, where the underlying process fulfils some mixing conditions. We will prove a sequential central limit theorem and the consistency on the nonoverlapping block bootstrap. These results are used to establish a test for detecting a change in the marginal distribution of a time series.

E416: **Sequential change-point procedures based on estimating functions**
Presenter: **Silke Weber**, Karlsruhe Insitute of Technology, Germany
Co-authors: Claudia Kirch

A general construction principle for detecting a change point if the data is observed sequentially is proposed. This includes examples as diverse as mean change, linear, non-linear, autoregressive and binary models. The key assumption is the existence of a training period without a change, which is used to get an initial estimator for the unknown parameters with an estimating function G. The detector is a score-type statistic which allows to check whether the initial estimate still describes the current data well enough. More precisely, we use three different types of detector statistics based on partial sums of a monitoring function H, which can be equal to the function G. They only differ in the number of observations included in the partial sum. We develop the asymptotics under the null hypothesis and alternatives under mild conditions for each test statistic. While the standard cumulative-sum-detector has recently been considered in this general setup, our schemes have only been studied in mean change or linear models. Our general conditions unify the theory for those but also many new examples such as the ones mentioned above. In a simulation study we compare these test procedures in terms of their size, power and run length.

E764: **Change-point test and estimator based on estimation functions**
Presenter: **Stefanie Schwaar**, University of Kaiserslautern, Germany

For an AMOC-model where the processes before and after the change-point are assumed to be nonlinear autoregressive processes, one approach uses neural network functions. The best fitting neural network parameter for this change-point model is estimated via least squares. By the assumptions to the generating functions of the neural network, the first derivative w.r.t. the neural network parameter is used to determine the neural network estimator. The derivative of the neural network is then an estimation function. The corresponding test statistic only uses the first component of the derivative. We generalize this set-up in the way that we use estimation functions for the parameter estimation as well as for the test-statistic.

E1151: **Dating multiple change points in the correlation matrix**
Presenter: **Pedro Galeano**, Universidad Carlos III de Madrid, Spain
Co-authors: Dominik Wied

A nonparametric procedure for detecting and dating multiple change points in the correlation matrix of a sequence of random variables is proposed. The procedure is based on a recently proposed test for changes in correlation matrices at an unknown point in time. Although the procedure requires constant expectations and variances, only mild assumptions on the serial dependence structure are assumed. The convergence rate of the change point estimators is derived and the asymptotic validity of the procedure is proved. Moreover, the performance of the proposed algorithm in finite samples is illustrated by means of a simulation study and the analysis of a real data example with financial returns. These examples show that the algorithm has large power in finite samples.

ES130 Room E1 NON-PARAMETRIC AND SEMI-PARAMETRIC METHODS Chair: Leopold Simar

E288: **A nonparametric approach to dynamic discrete choice models**
Presenter: **Valentin Zelenyuk**, University of Queensland, Australia
Co-authors: Byeong Uk Park, Leopold Simar

Dynamic discrete choice models are very important in applied research, frequently used in parametric contexts, and we develop a non-parametric approach to such models. The main contribution of the work is generalization of the non-parametric quasi-likelihood method to the context that allows for time series models, and in particular with lags of the (discrete) dependent variable appearing among regressors. We show consistency and asymptotic normality of the estimator for such models and illustrate it with simulated and real data examples.

E294: **Semiparametric model selection in panel data models with deterministic trending and cross-sectional dependence**
Presenter: **Jia Chen**, University of York, United Kingdom
Co-authors: Jiti Gao

A model selection issue in semiparametric panel data models with fixed effects is considered. The modelling framework under investigation can accommodate both nonlinear deterministic trends and cross-sectional dependence. And we consider the so-called “large panels" where both the time series and cross sectional sizes are very large. A penalised profile least squares method with first-stage local linear smoothing is developed to select the significant covariates and estimate the regression coefficients simultaneously. The convergence rate and the oracle property of the

resulting semiparametric estimator are established by the joint limit approach. The developed semiparametric model selection methodology is illustrated by two Monte-Carlo simulation studies, where we compare the performance in model selection and estimation of three penalties, i.e., the least absolute shrinkage and selection operator (LASSO), the smoothly clipped absolute deviation (SCAD), and the minimax concave penalty (MCP).

E502: **Nonparametric threshold regression: Estimation and inference**
Presenter: **Christopher Parmeter**, University of Miami, United States
Co-authors: Daniel Henderson, Liangjun Su

A simple approach is describe to estimate the location of a threshold/change point in a nonparametric regression. This model has connections both to the time-series and regression discontinuity literature. A simple estimator is proposed for the regression, threshold location and the size of the jump. Large sample properties are developed which extend existing results on semiparametric M-estimators. Optimal bandwidth selection and a suite of testing facilities which leverage the smooth bootstrap are also presented. Several empirical examples are provided to illustrate the implementation of the methods discussed here.

E569: **Generalized nonparametric smoothing with mixed discrete and continuous data**
Presenter: **Leopold Simar**, Universite Catholique de Louvain, Belgium
Co-authors: Degui Li, Valentin Zelenyuk

The nonparametric smoothing technique with mixed discrete and continuous regressors is considered. 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 a numerical study, it is found that in many cases, the performance of the resulting nonparametric regression estimates may deteriorate if the discrete variables are smoothed in the way previously addressed, and that a fully separate estimation without any smoothing of the discrete variables may provide significantly better results both for bias and variance. As a solution, it is suggested a simple generalization of the nonparametric smoothing technique with both discrete and continuous data to address this problem and to provide estimates with more robust performance. The asymptotic theory for the new nonparametric smoothing method is developed and the finite sample behavior of the proposed generalized approach is studied through extensive Monte Carlo experiments as well an empirical illustration.

EP01 Room First floor Hall POSTER SESSION II Chair: Cristian Gatu

E990: **Generalized p values in regression analysis**
Presenter: **Berna Yazici**, Anadolu, Turkey
Co-authors: Seray Mankir

In regression analysis, in case of comparing two regression models and coefficients where the distribution of variables is not known, generalized p values may be used. The generalized p value is an extended version of the classical p value which provides only approximate solutions. Use of approximate methods, generalized p value, have better results, performance with small samples. The generalized p value – which may be used alternatively when different assumptions aren't fulfilled - is researched theoretically; the suitable simulation program is written and an application in regression analysis is given.

E1040: **On a data-dependent choice of the tuning parameter appearing in certain goodness-of-fit tests**
Presenter: **Leonard Santana**, North West University, South Africa
Co-authors: James Allison

A data-dependent method is proposed for choosing the tuning parameter appearing in many recently developed goodness-of-fit test statistics. The new method, based on the bootstrap, is applicable to a class of distributions for which the null distribution of the test statistic is independent of unknown parameters. No data-dependent choice for this parameter exists in the literature; typically, a fixed value for the parameter is chosen which can perform well for some alternatives, but poorly for others. The performance of the new method is investigated by means of a Monte Carlo study, employing three tests for exponentiality. It is found that the Monte Carlo power of these tests, using the data-dependent choice, compares favourably to the maximum achievable power for the tests calculated over a grid of values of the tuning parameter.

E1083: **Resampling methods for randomized response techniques**
Presenter: **Beatriz Cobo**, University of Granada, Spain
Co-authors: Raghunath Arnab, Maria del Mar Rueda, Antonio Arcos

A survey is a research method that is based on questioning a sample of individuals. The interest in sample surveys studies often focuses on sensitive or confidential aspects to the interviewees. Because of this, the typical problem that arises is social desirability, which is defined as,the tendency of respondents to answer based on what is socially acceptable. For this reason, many respondents refuse to participate in the survey or provide false answers or conditioned responses, causing the accuracy and reliability of the estimates to be altered in a major way. Randomized response technique introduced by Warner is a possible solution for protecting the anonymity of the respondent and is introduced to reduce the risk of escape or no response sensitive questions. A problem with these techniques is the variance estimate, since it depends on the random mechanism chosen and need to calculate the second-order inclusion probability of each pair of sample units. In some complex sampling designs, this is very complicated.An alternative is to obtain the variance using resampling techniques, such as jackknife. We show that the classical jackknife estimator underestimates the variance estimator, so we propose a correction for this estimator.

E1078: **Modelling finger force produced from different tasks using linear mixed models with lme R function**
Presenter: **Marie-Jose Martinez**, University of Grenoble, France
Co-authors: Caroline Bazzoli, Frederique Letue

In experimental sciences, analysis of variance (ANOVA) is commonly used to explain one continuous response with respect to different experimental conditions, assuming independent observations and homoscedastic errors. In studies where individuals contribute more than one observation, such as longitudinal or repeated-measures studies, the linear mixed model provides then a better framework as an alternative to classical ANOVA. The considered data have been obtained from a biomechanical study carried out to understand the coordination patterns of finger forces produced from different tasks corresponding to different experimental conditions. This data cannot be considered independent because of within-subject repeated measurements, and because of simultaneous finger measurements. To fit these data, we propose a methodology focused on linear mixed models. Different random effects structures and different residual variance-covariance matrices are considered. We highlight how to use the lme R function to deal with such a complex modelling approach.

E1072: **Likelihood-free simulation-based optimal design with an application to spatial extremes**
Presenter: **Markus Hainy**, Johannes Kepler University Linz, Austria
Co-authors: Werner Mueller, Helga Wagner

A novel method is developed to find the optimal experimental design for problems where the likelihood is intractable, but simulation from the likelihood is feasible. To approximate the expected utility criterion, we make use of ABC (approximate Bayesian computation) methods. We detail the approach for a model on spatial extremes (Schlather model), where the goal is to find the optimal design for efficiently estimating the

parameters governing the dependence structure. The method is applied to determine the optimal design of weather stations for maximum annual summer temperatures.

E1096: **Statistical modelling of quarterly record of rainfall distribution in South West Nigeria**
Presenter: **Iyabode Oyenuga**, The Polytechnic Ibadan, Nigeria
Co-authors: Benjamin A Oyejola, Victor A Oni, Johnson T Olajide
Modeling distribution of rainfall in South West Nigeria is examined. Quarterly rainfall distribution data of Ibadan as a case study were collected for a period of 43 years (1971-2013) from Nigeria Meteorological Agency (NMA) quoted in Central Bank of Nigeria (CBN) bulletin. The time series analysis was used to model and forecast the quarterly rainfall. The time plot shows there are seasonal cycles of the series, we used Alkaike information criterion to detect auto-regressive (AR), moving average (MA) and auto-regressive moving average ARMA models of the best order. It was shown that AR(4), MA(4) and ARMA(2,2) have the least Akaike infomation criterion. These models were then used to forecast for the quarterly rainfall for five years.

E1110: **Estimation of running distance of top tennis players from video images and its applications**
Presenter: **Toshio Sakata**, Kyushu University, Japan
Co-authors: Kosuke Okusa, Hiriyasu Sakamoto, Yukiyasu Yoshinaga, Toshio Onishi
Recently, analysis of sports data is flourishing not only in the commercial community but also the statistical one. Data is a kind of big data, and there are programs for analyzing data about: service ace rates or net success rate and double fault rates etc., in a word, about match records. We view tennis players' data differently, and estimate how long and how efficiently they are running on a court in a game. Is this a factor of player's strength? To answer this we collect videos of tennis matches from internet and trace the locus of player's moving and calculate the distance of his running on the court. Methodologically, in each segmented image from a video, the position of a player's feet are automatically detected on a projective transformed court and the mid of them is traced. These running data are incorporated in a regression equation for match win rates or player's ranking as independent variables and the regression coefficients are estimated. All are analyzed by several software including the statistical software R, which shows a new possibility of R in sports data analysis.

E1247: **Frontal view gait analysis using weighted center of gait silhouette based on the scale registration**
Presenter: **Wataru Isaka**, Chuo University, Japan
Co-authors: Kosuke Okusa, Toshinari Kamakura
The problem of analyzing and classifying frontal view gait video data is studied. We focus on the shape scale changing in the frontal view human gait, we estimate scale parameters using the statistical registration and modeling on a video data. To demonstrate the effectiveness of our method, we apply our model to the frontal view gait video data. We analyze the balance between left and right arm swing based on the scale corrected silhouette's weighted center. As a result, our model shows good performance for the scale estimation and classification of human gait.

E1296: **Application of circulation type classification in analysis of links between atmospheric circulation and precipitation**
Presenter: **Jan Kysely**, Institute of Atmospheric Physics AS CR, Czech Republic
Co-authors: Eva Plavcova
Links between large-scale atmospheric circulation and surface climate variables represent important benchmark for assessing quality of climate models. If basic properties of the observed links are not captured in a climate model simulation for the recent climate, credibility of future projections is low for such a model. We study the circulation-to-precipitation links in an ensemble of regional climate models over several regions of Europe, by making use of a circulation type classification based on the Jenkinson scheme and circulation indices describing flow direction, strength and vorticity (derived from gridded mean sea level pressure). The ability of the climate models to reproduce the observed links between circulation and daily precipitation statistics (including extremes) is evaluated for individual seasons. Confidence intervals for the observed characteristics are determined by resampling methods. Results for regional climate models are compared with those for driving global climate models. We also discuss differences between applicability of the circulation type classification and simple circulation indices.

E1309: **The exponentiated modified Weibull long term hazard model in the context of breast cancer**
Presenter: **Gleici Perdona**, USP, Brazil
Co-authors: Hayala Souza, Francisco Louzada, Fernanda Peria
Studies on the frequentist properties of Exponentiated Modified Weibull Model for Long Term model are presented. This model embeds several existing lifetime models in a more general and flexible framework in terms of the hazard function. Moreover, the model has significant interpretations of its parameters related to a problem of breast cancer. Exact and interval estimation based on maximum likelihood and bootstrap resampling are presented. A data set from Cancer Register, which consist of 96 women diagnosed with advanced breast cancer between 2000 and 2010, treated on School of Medicine Clinic Hospital, Brazil was considered to exemplify the potential of the model.

EP03 Room First floor Hall POSTER SESSION III Chair: Cristian Gatu

E1134: **D-optimal designs for AFT models with Type I and random censoring**
Presenter: **M. Jesus Rivas-Lopez**, University of Salamanca, Spain
Co-authors: Jesus Lopez-Fidalgo, Rodrigo Del Campo
Accelerated Failure Time (AFT) models are fairly commonly used in the field of manufacturing, nevertheless there is an increasing interest in using AFT models for clinical trials as some literature reflects. These models focus on the direct effect of the explanatory variables on the survival function instead of the hazard function in the Proportional Hazards models. This property allows for an easier interpretation measuring directly the effect of the corresponding covariate on the survival time. Fisher Information Matrix for AFT models and approximate designs are computed, for Type I and random censoring. Optimal experimental designs will be computed for the Log-Logistic AFT model, but the procedures developed here can be extended to other AFT models in a similar way.

E1144: **rCOSA: a software package for clustering objects on subsets of attributes**
Presenter: **Maarten Kampert**, Leiden University, Netherlands
Co-authors: Jacqueline Meulman
rCOSA is a software package interfaced to the R language. It implements statistical techniques for clustering objects on subsets of attributes (COSA). The main output is a dissimilarity matrix that one can use for a variety of proximity analysis methods. Our package extends these COSA techniques. We add functions for hierarchical methods, multidimensional scaling, K-groups clustering, and data visualization. Of the many publications that cite the original COSA paper, only a small number actually use COSA. It is not an easy algorithm to program, and available software has been out-of-date. We introduce an up-to-date software package and a clear guidance for this advanced technique. The software package and related links are available for free at: https://github.com/mkampert/rCOSA

E1146: **A comparison of methods for analysing logistic regression models with both clinical and genomic variables**
Presenter: **Caroline Bazzoli**, University of Grenoble, France
Co-authors: Sophie Lambert-Lacroix
Prediction from high-dimensional genomic data is an active field in todays medical research. Most of the proposed prediction methods make use of

genomic data alone without considering established clinical covariates that often are available and known to have predictive value. Recent studies suggest that combining clinical and genomic information may improve predictions. We consider methods that simultaneously use both types of covariates, but applying dimension reduction only to the high-dimensional genomic variables. We first describe an approach based on partial least square method in linear regession context and propose new approaches for logistic regression models. We perform a comparison of the performance of several prediction methods combining clinical covariates and genomic data using simulation and a real data set.

E1158: **A Bayesian probit regression-based approach to address recording replications in the diagnosis of Parkinson's disease**
Presenter: **Jacinto Martin Jimenez**, Universidad de Extremadura, Spain
Co-authors: Lizbeth Naranjo, Carlos J. Perez, Yolanda Campos-Roca, Amparo Rubio

Several investigations have recently considered the use of acoustic parameters extracted from speech recordings as an objective and non-invasive tool to perform diagnosis and monitoring of Parkinson's Disease (PD). An experiment consisting of 40 patients from the Regional Association for PD in Extremadura (Spain) and 40 healthy controls have been performed. Repeated speech recordings were obtained from which several acoustic characteristics were extracted. The objective is to automatically discriminate healthy people from those with PD. Most of the published methodologies applied to perform classification of this kind of dataset fail to account for the dependent nature of the data. They artificially increase the sample size and lead to a diffuse criterion to define which subject is suffering from PD. Other methods summarize the characteristics in different ways and lose the structure of the dataset. A Bayesian probit regression-based approach that handles repeated measurements is proposed. The idea of using a data augmentation framework with latent variables is exploited to derive an efficient Gibbs sampling algorithm. The proposed approach is conceptually simple, computationally low-cost and easy-to-implement. The proposed methodology provided good results when it was applied to the dataset obtained from the experiment.

E1168: **Statistical analysis of human respiration based on complex signals with Doppler radar**
Presenter: **Yoshitomo Akimoto**, Chuo University, Japan
Co-authors: Kosuke Okusa, Toshinari Kamakura

The interest is in detecting human activities like respiration or heart beating, that is indispensable for our living. Japanese white paper tells that the number of aged people living alone is increasing recently and that this tendency is worldwide. We study the distribution of the complex signals from Doppler radar, which is characterized by the distribution on the 8-digit character on complex plane. This distribution is similar to the distribution on the circle. We discuss about this projected distribution on the 8-digit character for efficient estimates of human respiration patterns. We also investigate the property of this distribution based on simulation and show real sensing data example.

E1169: **Functional data analysis of running data by 3D-motion capture**
Presenter: **Takenori Sakumura**, Chuo University, Japan
Co-authors: Yoshitomo Akimoto, Kosuke Okusa, Toshinari Kamakura

Recently running has become very popular for health. More than 20,000 runners have been taking part in Boston and Tokyo marathon in recent years. People like running because it is a very easy sport and does not require any other large facilities except shoes and running wares. However, running is sometimes very dangerous and may result in death. We investigate 3D running models obtained by the 3D motion capture system. The data set is created by projecting 3D locations of 19 markers to the 3D human skeleton models. The measurements are arranged on the timeline and multivariate time-series. We use function data modeling for the observed data of joints locations when the subject runs on the treadmill under the 3D motion capture system. The 3D modeling is basically an approximation of trajectories of joints locations owing to functions such as radial basis functions, which are very efficient to approximate our data. We discuss trajectory changing when some loads are applied to subjects and also investigate statistical interactions of human joints in the body.

E1226: **Bayesian adaptation**
Presenter: **Catia Scricciolo**, Bocconi University, Italy

Nonparametric curve estimation is a fundamental problem intensively studied in a Bayesian framework only in the last decade, with more than a ten-years delay over the ponderous progress made in the frequentist literature. Adaptive estimation is a main theme: it accounts for designing a prior on a function space so that the posterior contracts at the truth at minimax-optimal rate. The rate has the desirable property of automatically adapting to the unknown regularity level of the estimandum: the correct rate stems, whichever the true value of the regularity parameter, even if knowledge of it is not a priori available to be exploited in the definition of the prior. As the amount of data grows, the posterior learns from the data so that the derived estimation procedure, despite lack of knowledge of the smoothness, performs as well as if the regularity level were known and this information could be incorporated into the prior. Adaptation may be thus regarded as an oracle property of the prior. We propose a categorization of the main approaches to design priors for minimax-optimal rate-adaptive estimation, and present some results in different statistical settings: density estimation and linear regression with unknown error distribution.

E1190: **Influence diagnostics under a model-based approach in survey sampling**
Presenter: **Inmaculada Barranco-Chamorro**, University of Sevilla, Spain
Co-authors: Juan Luis Moreno-Rebollo, Jose Antonio Mayor-Gallego, Maria Dolores Jimenez-Gamero

Robust methods and influence diagnostics are two different and complementary strategies to deal with the effect of individual data values on the summaries of statistical techniques. In survey sampling, deletion-type diagnostics are of interest to assess the effect of omitting an observation on linear estimators of population parameters. We focus on results under model-based inference. So a case-deletion diagnostic for a class of linear penalized calibrated estimators is proposed. Expressions for the weights and estimators are obtained. From them, an expression for a case-deletion diagnostic is obtained based on quantities related to the full data set. An application to small area estimation and numerical examples are included.

E1271: **A nonparametric kernel approach to interval-valued data analysis**
Presenter: **Kee-Hoon Kang**, Hankuk University of Foreign Studies, Korea, South
Co-authors: Cheolwoo Park, Yongho Jeon

An exploratory data analysis approach is proposed when data are observed as intervals in a nonparametric regression setting. The interval-valued data contain richer information than single-valued data in the sense that they provide both center and range information of the underlying structure. Conventionally, these two attributes have been studied separately as traditional tools can be readily used for single-valued data analysis. We propose a unified data analysis tool that attempts to capture the relationship between response and covariate by simultaneously accounting for variability present in the data. It utilizes a kernel smoothing approach, which is conducted in scale-space so that it considers a wide range of smoothing parameters rather than selecting an optimal value. It also visually summarizes the significance of trends in the data as a color map across multiple locations and scales. We demonstrate its effectiveness as an exploratory data analysis tool for interval-valued data using simulated and real examples.

E1143: **Longitudinal analysis of tumor markers CA 15-3 and CEA of breast cancer patients from Braga's hospital**
Presenter: **Ana Borges**, Universidade do Minho, Portugal
Co-authors: Ines Sousa, Luis Castro

Allied to an epidemiological study of population of the Senology Unit of Braga's Hospital that have been diagnosed with malignant breast cancer, the progression in time of two tumor markers is described: Carcinoma Antigen 15-3 (CA 15-3) and the Carcinoembryonic antigen (CEA). Our main

purpose is to describe the progression of these two tumor markers as a function of possible risk factors. As so, to understand how these risk factors influence that progression. The two response variables, CA15-3 and CEA, were analyzed separately making use of longitudinal models, testing different correlation structures. The same covariates used in a survival model, previous adjusted in an earlier study, were tested in the longitudinal model fitted. The reference time used was time since diagnosis for breast cancer. For diagnostic of the models fitted we have used empirical and theoretical variograms and plots of fitted and residuals. To model the fixed term of the longitudinal model we have tested for a changing point on the effect of time on the tumor markers. A longitudinal model was also fitted only to the subset of patients that died, using the reference time as time since date of death.

Sunday 7.12.2014 16:55 - 18:35 Parallel Session M – CFE

CSI01 Room Sala Convegni VOLATILITY MODELLING Chair: Giuseppe Storti

C504: **Roughing it up some more: Jumps and co-jumps in vast-dimensional price processes**
Presenter: **Sebastien Laurent**, AMU, France
Co-authors: Kris Boudt, Rogier Quaedvlieg

An estimator of the ex-post covariation of log-prices under asynchronicity, microstructure noise and finite activity jumps is proposed. The estimator is based on the CholCov previously introduced, and therefore ensures a positive semidefinite estimate. Monte Carlo simulations confirm good finite sample properties. In the application we demonstrate the benefits of decomposing the quadratic covariation into a continuous and a jump component. Based on adaptations of a range of popular multivariate volatility models, we show the decomposition leads to improved fit and forecasting abilities.

C860: **Modeling eigen-dynamics of realized covariances**
Presenter: **Giampiero Gallo**, University of Florence, Italy
Co-authors: Francesco Calvori, Fabrizio Cipollini

A new approach for modeling Realized Covariance matrices for forecasting purposes is proposed. Denoting as X_t the $k \times k$ daily Realized Covariance matrix at time t ($t = 1, \ldots, T$), the modeling strategy is inspired by its spectral decomposition $X_t = \sum_{i=1}^{k} \lambda_{t,i} q_{t,i} q'_{t,i}$. The eigenvalues $\lambda_{t,i}$'s, being non-negative, can be modeled one by one, in subsets or jointly: via MEM (Multiplicative Error Model) in the first case, via vector-MEM in the remaining cases. The eigenvectors q_t's are modeled one at a time via a novel formulation which is able to reproduce some of their empirical stylized facts. In case of large k's, since different representations may give substantially equivalent X_t but may have different degrees of predictability, we propose a data pre-processing procedure to adjust eigenvectors relative to close eigenvalues. The approach is applied to nine sectoral ETF's representing the S&P500 index and to the components of the S&P100.We perform a comparison between our approach and previous suggestions in the literature.

C1000: **Comparing the accuracy of multivariate density forecasts in selected regions of the copula support**
Presenter: **Cees Diks**, University of Amsterdam, Netherlands
Co-authors: Valentyn Panchenko, Oleg Sokolinskiy, Dick van Dijk

A testing framework is developed for comparing the predictive accuracy of competing multivariate density forecasts with different predictive copulas, focusing on specific parts of the copula support. The tests are framed in the context of the Kullback-Leibler Information Criterion, using (out-of-sample) conditional likelihood and censored likelihood in order to focus the evaluation on the region of interest. Monte Carlo simulations document that the resulting test statistics have satisfactory size and power properties for realistic sample sizes. In an empirical application to daily changes of yields on government bonds of the G7 countries we obtain insights into why the Student-t and Clayton mixture copula outperforms the other copulas considered; mixing in the Clayton copula with the t-copula is of particular importance to obtain high forecasting accuracy in periods of jointly falling yields.

CS24 Room A2 NONSTATIONARY TIME SERIES AND FINANCIAL BUBBLES Chair: Christian Francq

C201: **Investigating long run linear relationships between non constant variances of economic variables**
Presenter: **Hamdi Raissi**, Universite Europeenne de Bretagne, France
Co-authors: Junichi Hirukawa

A two step methodology for investigating long run linear relationships between the non constant variance structures of time series variables is developed. Noting that the processes we consider have time-varying variances our tools are based on a wild bootstrap technique. As a possible application we propose to use our methodology for improving variables forecasting. The theoretical outputs are illustrated by means of Monte Carlo experiments. The existence of linear variance relations between U.S. economic variables is also studied.

C672: **Statistical inference in semiparametric locally stationary ARCH models**
Presenter: **Lionel Truquet**, ENSAI, France

Semiparametric versions of the univariate time-varying ARCH(p) model previously introduced and studied are considered. For a given nonstationary data set, a natural question is to determine which coefficients capture the nonstationarity and then which coefficients can be assumed to be non time-varying. For example, when the intercept is the single time-varying coefficient, the resulting model is close to a multiplicative volatility model in the sense of previous works. Using kernel estimation, we will first explain how to estimate the parametric and the nonparametric component of the volatility and how to obtain an asymptotically efficient estimator of the parametric part when the noise is Gaussian. The problem of testing whether some coefficients are constant or not is also addressed. In particular, our procedure can be used to test the existence of a second-order dynamic in this nonstationary framework. Our methodology can be adapted to more general linear regression models with time-varying coefficients.

C678: **Explosive bubble modelling by noncausal process**
Presenter: **Jean-Michel Zakoian**, CREST, France

The linear mixed causal and noncausal autoregressive processes provide often a better fit to economic and financial time series than the standard causal linear autoregressive processes. The noncausal autoregressive process with heavy-tailed errors possesses a nonlinear causal dynamics, which includes unit root, bubble phenomena, or asymmetric cycles often observed in financial markets. The noncausal autoregressive process with heavy-tailed errors provides a new model for explosive multiple bubbles and their transmission in a multivariate dynamic framework. By considering a mixed autoregressive process with Cauchy errors, we also explain why standard unit root tests will not detect such explosive bubbles.

C1152: **Understanding commodity futures prices: fundamentals, financialization and bubble characteristics**
Presenter: **Roderick McCrorie**, University of St Andrews, United Kingdom
Co-authors: Christopher Gilbert, Isabel Figuerola-Ferretti

The purpose is to provide a systematic investigation of the empirical properties of a range of key commodity futures prices, providing a robust evidence base upon which to assess various explanations of commodity futures price movements. We utilize a recently developed statistical methodology that allows date-stamping of the origination and collapse of bubbles, in a form where critical values are robustified to allow for possibly different data spans and sampling frequencies. Our substantive findings are that commodity markets are prone to bubble-like phenomena and were especially so during the financial crisis in late 2007 and early 2008. We identify bubbles around that time in energy, grains, and oils and oilseeds, but not in livestock or softs; and the bubbles lasted for up to six months. Our key finding is that bubble incidence is more closely related to the supply and demand for market fundamentals than it is to financialization. We conclude that the evidence base that underlies regulators' actions in the light of the financial crisis "to diminish, eliminate, or prevent excessive speculation" in commodity futures markets is only weakly supportive at best.

CS28 Room M2 FINANCIAL ECONOMETRICS Chair: Christophe Chorro

C320: **Stress testing engineering: the real risk measurement?**
Presenter: **Bertrand Hassani**, Pantheon Sorbonne, France

Stress testing is used to determine the stability or the resilience of a given financial institution by deliberately submitting. We focus on what may lead a bank to fail and how its resilience can be measured. Two families of triggers are analysed: the first one stands in the impact of external (and / or extreme) events, the second one stands on the impacts of the choice of inadequate models for predictions or risks measurement; more precisely on models becoming inadequate with time because of not being sufficiently flexible to adapt themselves to dynamical changes.

C439: **From historical to risk neutral volatility: a GARCH approach**
Presenter: **Christophe Chorro**, Universite Paris 1, France
Co-authors: Florian Ielpo, Thuy An Nguyen thi

The aim is to study the ability of GARCH Option Pricing models to reproduce the VIX index trajectory. Using explicit Implied VIX formulas under the class of Gaussian or non-Gaussian GARCH models, we compare the numerical performances between different estimation and modeling approaches.

C546: **Measuring the time-varying dependence across financial markets: a contamination analysis of equity and sovereign CDS markets**
Presenter: **Hayette Gatfaoui**, University Paris one, France
Co-authors: Philippe de Peretti, Lorenzo Frattarolo

The interrelations between daily stock market indices and CDS spreads of major European countries are studied. We use cross-correlation tests, which are based on the innovations of time series, rather than the time series themselves. They allow for capturing the dynamics of bivariate or higher order systems among which non-correlation and non-causality. Therefore, we propose a fine-tuned analysis of spillover and feedback effects within stock markets, within CDS markets, and between stock and CDS markets. Our study encompasses three main steps. First, we model the dynamics of each series' volatility using either Markov Switching or Changing Point GARCH representations. Second, cross correlation tests are implemented over each regime of volatility. Finally, using rolling windows, we build a temporal directed network and analyze its evolution over time using different indicators. During the period spanning from 1999 to 2014, the whole dependence and causal structure changes over time. Moreover, excess dependence does have some predictive contents concerning future crises.

C698: **Lattice based techniques for GARCH option hedging**
Presenter: **Juan-Pablo Ortega**, CNRS Universite de Franche-Comte, France
Co-authors: Alexandru Badescu, Matthew Couch

Despite their success in the modeling of asset prices, GARCH processes are still rarely used for derivative hedging purposes. We present several results that we have obtained regarding this problem. In particular, we describe an approach that uses a lattice approximation for computing the hedge ratios required in the local quadratic hedging scheme for European options. We provide numerical experiments to compare the results of our strategy with those obtained using other standard evaluation methods such as Monte Carlo simulation under various natural probability measures that arise in the problem.

CS37 Room I2 COMMON FEATURES IN FINANCE AND MACROECONOMICS Chair: Joao Victor Issler

C151: **Inferring the predictability induced by a persistent regressor in a predictive threshold model**
Presenter: **Jean-Yves Pitarakis**, University of Southampton, United Kingdom
Co-authors: Jesus Gonzalo

A distributional theory for detecting predictability induced by a persistent variable is developed. Our framework is that of a predictive regression model with threshold effects and our goal is to develop operational and easily implementable inferences when one does not want to impose a priori restrictions on the parameters other than the slopes corresponding to the persistent predictor. Differently put our tests for the null hypothesis of no predictability against threshold style predictability across two regimes remain valid without the need to know whether the remaining parameters of the model are characterised by threshold effects or not (e.g. shifting versus non-shifting intercepts). One interesting feature of our setting is that our test statistic remains unaffected by whether some nuisance parameters are identified or not.

C567: **Measuring macroeconomic uncertainty: output growth and inflation**
Presenter: **Ana Galvao**, University of Warwick, United Kingdom
Co-authors: Mike Clements

It is found that model estimates of the term structure of ex ante or perceived macro uncertainty are more in line with realized uncertainty than survey respondents perceptions for both inflation and output growth. Survey estimates contain short-term variation in short-horizon uncertainty which is less evident in the model-based estimates. We consider the extent to which these short-term variations coincide with short-term movements in stock market uncertainty.

C606: **Do monetary and productivity shocks explain aggregate fluctuations?**
Presenter: **Osmani Guillen**, Banco Central do Brasil, Brazil
Co-authors: Joao Victor Issler, Afonso Arinos de Mello Franco-Neto

The objective is to estimate separately the welfare costs of the uncertainty stemming from business-cycle fluctuations and from economic-growth variation, when monetary and productivity shocks (respectively, transitory and permanent shocks) hit consumption simultaneously. Separating these welfare costs requires dealing with Dynamic stochastic general equilibrium (DSGE) model to explain comovements of aggregate time series over the business-cycle and to perform policy analysis. The product of the separation of these welfare costs are degenerate multivariate distributions that are treated by the application of Levi's Continuity Theorem and the Disintegration Theorem. Under Normality, we show that the parameters of the original marginal distributions are not affected, providing the means for calculating separately the welfare costs of business-cycle fluctuations and of economic-growth variation. Our empirical results show that, if we consider only transitory shocks, the welfare cost of business cycles is small. On the other hand, we found that the welfare cost of economic-growth variation is relatively large. Most of the results hold for a variety of models, and the majority countries studied.

C720: **Microfounded forecasting**
Presenter: **Joao Victor Issler**, Getulio Vargas Foundation, Brazil
Co-authors: Wagner Gaglianone

A microfounded framework is proposed to investigate a panel of forecasts (e.g. model-driven or survey-based) and the possibility to improve their out-of-sample forecast performance by employing a bias-correction device. Following a previous work, we theoretically justify the modeling of forecasts as a function of the conditional expectation, based on the optimization problem of individual forecasters. This approach allows us to relax the standard assumption of mean squared error (MSE) loss function and, thus, to obtain optimal forecasts under more general functions. However, different from these authors, we apply our results to a panel of forecasts, in order to construct an optimal (combined) forecast. In this sense, a feasible GMM estimator is proposed to aggregate the information content of each individual forecast and optimally recover the conditional expectation. Our setup can be viewed as a generalization of the three-way forecast error decomposition, and as an extension of the bias-corrected average forecast, of previous works. A real-time forecasting exercise using Brazilian survey data illustrates the proposed methodology.

CS49 Room P2 DEVELOPMENT ON SEASON ADJUSTMENT AND SEASONAL MODELLING Chair: Gianluigi Mazzi

C196: **An application of ramp effect to modelling of a crisis**
Presenter: **Sylwia Grudkowska**, National Bank of Poland, Poland

Sudden changes in trend-cycle, caused by e.g. the impact of the crisis, often lead to the significant and persistent deterioration in the quality of the seasonal adjustment (SA) results from Automatic Model Identification (AMI) procedure. Such effects increase uncertainty in SA results. Therefore, there is a strong users' demand for additional tools that amend the RegARIMA models chosen by AMI. The aim is to present the results of an application of the improved algorithm for identification of a ramp effect to the real time series derived from Eurostat's statistics database. The study, conducted on economic indicators for 35 countries, shows that the algorithm vastly improves the quality of results of standard automatic adjustment. The work also investigates the impact of Ramp effect on the RegARIMA model, size of revisions and detection of turning points. The usage of the new SA tool, called JDemetra+, enabled us to integrate the algorithm with the TRAMO/SEATS method.

C328: **Calendar effects in time series analysis**
Presenter: **Dominique Ladiray**, INSEE, France
Co-authors: Fabien Guggemos

Practically all economic time series are computed and published according to the Gregorian calendar, a calendar based on the motion of the earth around the sun. This solar calendar also brings rhythm to our lives and usually has a deep impact on the economy. The most well known and important calendar effect is seasonality. But most economic indicators are also linked, directly or indirectly, to a daily activity which is usually summed up and reported each month or each quarter. In this case, the number of working days, which varies from one month to another in a quasi predetermined way, can explain some short-term movements in the time series. Apart from the day composition of the month, other calendar effects such as public holidays or religious events may also affect the series. Religious events are often closely linked to other calendars and their dates, expressed in the Gregorian calendar, may not be fixed (Easter, Ramadan, Chinese New Year, etc.). We present some algorithms and facts about calendars and the mathematics that allow defining a relevant set of regressors to take into account calendar effects in the analysis of a series.

C545: **Seasonal adjustment of short time-series: a comparative study**
Presenter: **Enrico Infante**, EUROSTAT, Luxembourg
Co-authors: Gian Luigi Mazzi

Economic time-series can be or become short due to a variety of reasons: production of new indicators, changes in methodologies, definitions, nomenclatures, etc. While users always ask for seasonally adjusted figures, their production can be very problematic and their quality not necessarily in line with the requirement of official statistics. The problem does not arise when series are very short (three years or less) since most used SA methods cannot be applied. For series longer than three years, SA data can be obtained but results can be volatile and subject to relevant revisions. The aim is to present a comparative study based on the use of the most known SA methods on a selected group of macro-economic indicators. For each indicator we select three sample timespans (short, medium and long) and we compare the quality of the SA data on the common period to the three timespans. A variety of parametric and non-parametric SA quality measures is used and synthetic comparative tables are derived. The attention is focused on the stability of the SA pattern when changing the timespan, as well as on the relative performance of alternative SA methods when dealing with short time-series.

C668: **Seasonal adjustment in real-time: a comparison of X-13ARIMA-SEATS and CAMPLET**
Presenter: **Barend Abeln**, Investment consultant, Netherlands
Co-authors: Jan Pam Jacobs

Seasonality in macroeconomic time series can obscure movements of other components in a series that are operationally more important for economic and econometric analyses. Indeed, in practice, one often prefers to work with seasonally adjusted data to assess the current state of the economy and its future course. Recently, two most widely used seasonal adjustment methods, Census X-12-ARIMA and TRAMO-SEATS, merged into X-13ARIMA-SEATS to become a new industry standard. We compare and contrast X-13ARIMA-SEATS with a real-time seasonal adjustment program called CAMPLET, an acronym of its tuning parameters. CAMPLET consists of a simple adaptive procedure which separates the seasonal component and the non-seasonal component from an observed time series. Once this process has been carried out there will be no need to revise these components at a later stage when more time series data become available, in contrast with other seasonal adjustment methods. The main features of CAMPLET are described, in addition to a brief review of X-13ARIMA-SEATS. We apply the X-13ARIMA-SEATS and CAMPLET methods to two real time series: U.S. non-farm payroll employment, and operational income of Ahold. The main findings are that both methods generally produce similar seasonal adjusted figures. In addition, X-13ARIMA-SEATS does not produce large revisions in a (quasi-) real-time analysis in non-farm payroll employment, but CAMPLET does not show any revisions at all. CAMPLET non-seasonals fall deeper in 2009 than X-13ARIMA-SEATS seasonal adjusted figures. Finally, our second illustration involving operating income of Ahold shows that X-13ARIMA-SEATS does not pick up changes in seasonal patterns, in contrast to CAMPLET.

CS56 Room D2 MODELING MULTIVARIATE FINANCIAL TIME SERIES Chair: Edoardo Otranto

C223: **How interdependent are systemic risk indicators? A network analysis**
Presenter: **Daniele Frison**, European Central Bank, Spain
Co-authors: Christian Brownlees

Following the financial crises in the US and Europe, hundreds of systemic risk indicators have started being scrutinized by regulators to take pulse of the economy. When studying such an extensive collection of indices, it is natural to ask to which extent the indicators comove. We analyze a panel of 156 systemic risk indicators from the ESRB risk dashboard from January 1999 to December 2013. We study cross-sectional dependence using a factor-network approach: dependence among indicators is determined by a set of common factors and a network of pairwise partial dependence relations among individual pairs. The empirical analysis shows that network interdependence is prominent while factors explain a relatively small proportion of overall covariation. Macro, credit and funding risk categories are the most central and highly interconnected indicators in the network. Network linkages are stronger among indicators in the same risk categories rather than indicators belonging to the same geographic area. Finally, the network centrality analysis shows that macro-imbalance indicators like current account balance, residential property prices, corporate sector indebtedness and bank loans-to-deposits ratios are the most interdependent systemic risk bellwethers of the dashboard.

C525: **Conditional correlation models with nonstationary variances and correlations**
Presenter: **Cristina Amado**, University of Minho, Portugal
Co-authors: Annastiina Silvennoinen, Timo Terasvirta

The issue of allowing for nonstationarity in the volatilities and correlations in conditional correlation models is addressed. For this purpose, we introduce nonstationarity in the individual variances and correlations in two steps. Firstly, the variance equations are modelled such that the long-run dynamics of volatility change smoothly over time. Secondly, the correlation structure is assumed to evolve slowly over time between two extreme states of constant correlations. An empirical application to daily return series of the European, Japanese and US stock market indices markets illustrates the functioning of the model in practice.

C647: **Contagion through common exposure and systemic risk**
Presenter: **Fulvio Corsi**, Ca Foscari University Venice and City University London, Italy

Co-authors: Fabrizio Lillo, Davide Pirino

The aim is to study the common risk exposures of financial institutions as a bipartite network where one set of nodes are the financial institutions and the other the investment assets. While the literature on direct contagion effects has grown at a fast pace in recent years the one on indirect contagion through portfolio holdings is still sparse. The goal is to identify and quantify the contagion effects among financial institutions through the bipartite network and propose a systemic risk measure which accounts for indirect contagion through common assets.

C085: **A new approach to high-dimensional volatility modelling**

Presenter: **Christian Hafner**, UCL Louvain-la-Neuve, Belgium

Co-authors: Oliver Linton

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.

CS84 Room Q2 TAIL RISKS Chair: Marco Geraci

C064: **Short selling in the tails**

Presenter: **Marco Valerio Geraci**, Universite libre de Bruxelles, Belgium

Co-authors: David Veredas, Tomas Garbaravicius

Despite the academic and practitioners' consensus that short selling has mostly positive effects, regulators applied short selling bans during times of crisis, on the grounds that asset prices are vulnerable to (potentially predatory) short selling. Our objective is to make sense of these concerns by studying the association between short selling and price drops in extreme circumstances. We develop a proxy for short selling activity based on securities lending data for the stocks of the largest European and North American banks and insurance companies. We find that the linear relation between short selling and returns is very low in normal times, which is consistent with the literature, and leads our conclusions towards the goodness of short selling. However, when we examine the relation in the tails we discover an uncommonly high negative relationship. To quantify this finding we adopt TailCoR, a measure of tail correlation. We show that during short selling bans the (negative) relationship between short selling and returns decreases, but only modestly. In contrast, long-term policies contrasting abusive short selling adopted by SEC with regulation SHO appear to have been more successful in obtaining a gradual and sustainable decrease in the vicious relation uncovered.

C068: **Tail risk dynamics in stock returns: Observation-driven approach and links to the business cycle**

Presenter: **Daniele Massacci**, Einaudi Institute for Economics and Finance, Italy

The dynamics of tail risk in stock returns over the business cycle are analyzed. An econometric specification is proposed for the conditional distribution of extreme returns: the model builds on the peaks over the threshold from extreme value theory; the observation-driven approach is followed to specify the laws of motion for the parameters as functions of observable variables. Application of the model to daily returns from size-sorted U.S. decile stock portfolios shows the highly countercyclical dynamics of tail risk.

C129: **Ranking the stars**

Presenter: **Federico Nucera**, Luiss Guido Carli, Italy

Co-authors: Bernd Schwaab, Andre Lucas, Siem Jan Koopman

A methodology is provided to combine different systemic risk rankings of financial institutions to achieve a robust overall ranking for policy or asset allocation purposes. We consider a number of important risk rankings from the literature and disentangle their common (signal) and idiosyncratic (noise) parts. We apply our methodology to 113 financial sector firms in the European Union from 2002-2013, and demonstrate that our combined ranking is less volatile than the individual input rankings and leads to less turnover at the top. The combined ranking forecasts financial firms' equity returns out of sample, and serves as a basis of a profitable asset allocation strategy. We further argue that combined risk rankings are reliable inputs into policy and regulatory discussions.

C260: **Score driven time varying tail risk**

Presenter: **Xin Zhang**, Sveriges Riksbank, Sweden

Co-authors: David Veredas

Time variation is introduced into the tail risk measurement in heavy-tailed distributions. The framework is based on extreme value theory with dynamic tail index parameter driven by the score of the partial likelihood, i.e. the tail index parameter is a function of past observations. As far as the partial likelihood satisfies standard regularity conditions, the model provides an unbiased estimate of the tail dynamics. Our approach is a robust and natural way for modeling the dynamic tail risk. We show its good performance in capturing the tail risk dynamics with Monte Carlo simulations and an application with daily returns.

CS85 Room E2 FUNDS PERFORMANCE MEASUREMENT Chair: Spyridon Vrontos

C299: **Density forecasting of hedge fund returns**

Presenter: **Ekaterini Panopoulou**, University of Kent, United Kingdom

Co-authors: Spyridon Vrontos

Hedge funds' exposures to various financial and macroeconomic risk factors are investigated by employing quantile predictive regression models. The quantile regression framework enables us to unveil the entire conditional distribution of individual hedge fund returns and assess the risk-return profile of each hedge fund/ trading strategy. We categorize hedge fund strategies in (i) directional strategies that take direct market exposure, (ii) non-directional strategies that aim at minimizing market risk and (iii) semi-directional strategies that try to diversify market risk. We employ hedge fund data from the Barclays hedge fund database for the period January 1994 to December 2013. Our findings suggest that considerable heterogeneity exists with respect to the degree of exposures to risk factors of different hedge fund investment styles. Finally, we further exploit this heterogeneity in dynamic optimal fund of fund construction.

C309: **Hedge fund predictability and optimal asset allocation**

Presenter: **Theologos Pantelidis**, University of Macedonia, Greece

Co-authors: Ekaterini Panopoulou, Spyridon Vrontos

The degree of both return and volatility hedge fund predictability is revealed using a regime switching framework. Optimal combinations of regime switching model forecasts allow us to capture the stylized facts of hedge fund returns and construct superior hedge fund return forecasts in the presence of parameter instability and model uncertainty. Our dataset consists of individual hedge fund data from the Barclays hedge fund database for the period January 1994 to December 2013. Our extensive set of predictors contains the Fung and Hsieh factors, factors related to style investing and to investment policies, macro related / business indicators variables and market-oriented factors. The economic value of the proposed predictability models is investigated by studying its effects on asset allocation and active portfolio management.

C629: **Late trading in mutual fund shares - the sequel**

Presenter: **Moritz Wagner**, University of Auckland, New Zealand

Co-authors: Dimitri Margaritis

New evidence of late trading activities in the mutual fund markets of France, Germany and the Netherlands is provided. We find that investors who are allowed to trade late can earn substantial returns between 22% and 47% annually. Evidence of such illicit and abusive trading practices was uncovered in 2003 during a large scandal in the US fund industry. Our preliminary findings suggest that late trading may still persist in European markets.

C690: **Performance evaluation of funds**
Presenter: **Spyridon Vrontos**, University of Westminster, United Kingdom
One of the most controversial questions arising in fund management is whether fund managers create additional value to the performance of the funds under their management and if they are able to deliver superior performance when compared with the market. The recent Financial Crisis of 2007-2009 has revived the interest in this area since the majority of the funds experienced significant losses during the crisis period. The performance fees, i.e. the fees that a client account may be charged by the investment manager that manages its assets, are in the case of funds among the highest of the market. Through the performance evaluation of funds, we examine the existence or not of fund managerial skill and if the high fees are justified by their performance. A variety of different models that have been proposed in the literature are applied and compared.

CS89 Room B2 ANALYSING COMPLEX TIME SERIES Chair: Matteo Barigozzi

C022: **Dynamic factor models, cointegration, and error correction mechanisms**
Presenter: **Matteo Barigozzi**, London School of Economics, United Kingdom
Co-authors: Marco Lippi, Matteo Luciani
Dynamic factor models are studied when the factors F_t are I(1) and singular, i.e. rank$(F_t) <$ dim(F_t). By combining the classic Granger Representation Theorem with recent results on singular stochastic vectors, we prove that, for generic values of the parameters, F_t has an Error Correction representation with two unusual features: (i) the autoregressive matrix polynomial is finite, (ii) the number of error-terms is equal to the number of transitory shocks plus the difference between the dimension and the rank of F_t. This result is the basis for the correct specification of an autoregressive model for F_t. Estimation of impulse-response functions is also discussed. Results of an empirical analysis on a US quarterly database support the use of our model.

C122: **Local composite quantile regression smoothing for Harris recurrent Markov processes**
Presenter: **Degui Li**, University of York, United Kingdom
Co-authors: Runze Li
The aim is to study the local polynomial composite quantile regression (CQR) smoothing method for the nonlinear and nonparametric models under the Harris recurrent Markov chain framework. The local polynomial CQR regression method is a robust alternative to the widely-used local polynomial method, and has been well studied in stationary time series. We relax the stationarity restriction on the model, and allow that the regressors are generated by a general Harris recurrent Markov process which includes both the stationary (positive recurrent) and nonstationary (null recurrent) cases. Under some mild conditions, we establish the asymptotic theory for the proposed local polynomial CQR estimator of the mean regression function, and show that the convergence rate for the estimator in the nonstationary case is slower than that in the stationary case. Furthermore, a weighted type local polynomial CQR estimator is provided to improve the estimation efficiency, and a data-driven bandwidth selection is introduced to choose the optimal bandwidth involved in the nonparametric estimators. Finally, we present numerical studies to illustrate the methodology and theory.

C576: **Classifying complex time series databases by periodic components**
Presenter: **Maria Lucia Parrella**, University of Salerno, Italy
Co-authors: Francesco Giordano, Michele La Rocca
The automatic statistical modelling of large databases of time series may be a very hard problem when the time series show complex features, such as nonlinearity, nonstationarity, high frequency and periodic components. Classification and clustering of such "complex objects" may be particularly useful for the areas of model selection and pattern recognition. As a consequence, time series clustering represents a first mandatory step in temporal data mining research.We consider the case when the time series show periodic components of different frequency. As an example with real data, we consider the time series from a large database provided by an electric company. Time series refer to the electricity consumption of the business and industrial firms. The main features of the observed time series are: cycles, long memory and stationarity. These peculiarities and the large number of time series in the database make the analysis complex, and point towards highly automated procedures which are able to: (i) efficiently detect the structure of the series and (ii) reduce the computational burden and the amount of data to be stored. A novel clustering algorithm based on spectral techniques is proposed. The performance of the proposed procedure is investigated through a simulation study.

C826: **A review on nonlinear regression analysis of irregularly located spatial time-series data**
Presenter: **Zudi Lu**, University of Southampton, United Kingdom
Larger amounts of time-series data with more complex structures collected at irregularly spaced sampling locations are becoming more prevalent in a wide range of disciplines. With few exceptions, however, practical statistical methods for modeling and analysis of such data remain elusive. We provide a review on some developments and progress of the research. In particular, we will look at some semiparametric models, such as a class of spatio-temporal autoregressive partially nonlinear regression models, which permits possibly nonlinear relationships between responses and covariates. In addition, the dependence structure is stationary over time but nonstationary over space, while the sampling grids can be irregular. We develop a computationally feasible method for estimation and thus enable our methodology to be applicable in practice. Asymptotic properties of our proposed estimates are established and comparisons are made, in theory and via simulations, between estimates before and after spatial smoothing. For illustration, our methodology is applied to investigate housing prices in relation to mortgage rates in the United States from 1991 to 2012, with a threshold structure found to be helpful for prediction.

CS91 Room N2 ESTIMATION OF MACROECONOMIC MODELS WITH TIME VARYING VOLATILITY Chair: Marco Del Negro

C088: **Large dynamic panels with stochastic volatility**
Presenter: **William McCausland**, University of Montreal, Canada
Large dynamic panels are useful for macroeconomic forecasting. Models with factor structure are convenient for dimension reduction. It is desirable to include financial variables along with measures of real activity, but it has proven difficult to accommodate the time varying volatility that these variables exhibit. We propose new inferential methods for dynamic factor models in data-rich environments. Both the conditional mean and variance of observed series have a dynamic factor structure. We use MCMC techniques to simulate from the joint posterior distribution of factors and parameters. We use standard methods to draw conditional mean factors from their posterior distribution. We apply the HESSIAN method for drawing stochastic volatility factors from their conditional posterior distribution; we draw each stochastic volatility factor series as a single Gibbs block, and the various series one at a time. This approach is numerically efficient and allows more flexible statistical dependence of volatility factors than approaches widely used in the literature. We apply our new techniques to a large panel of real activity and financial variables.

C101: **Large VARs with time varying volatility**
Presenter: **Marco Del Negro**, Federal Reserve Bank of New York, United States

Co-authors: Daniel Greenwald

A novel approach to the estimation of large VARs with time varying volatility is proposed.

C107: **Volatility shocks and business cycles**

Presenter: **Carlos Montes-Galdon**, Columbia University, United States

The contribution of the volatility of several macroeconomic variables to the business cycle is analyzed. A structural vector autoregressive model augmented with stochastic volatility that has first moment effects in those variables is used. This method allows us to make a distinction between the effect of changing the size of macroeconomic shocks and the effect of pure volatility shocks, that can be also understood as changes in macro-uncertainty. Using this model, we find that changes in volatility in a finance-related proxy variable are one of the main drivers of business cycles in the United States, but uncertainty in inflation, output and monetary policy also have modest effects in the variance of the main macroeconomic variables. Finally, we show how to estimate the model using Hamiltonian Monte-Carlo techniques, since standard stochastic volatility methods can not be used in this case.

C300: **Estimating dynamic equilibrium models with stochastic volatility**

Presenter: **Pablo Guerron**, Federal Reserve Bank of Philadelphia, United States

A novel method to estimate dynamic equilibrium models with stochastic volatility is proposed. First, we characterize the properties of the solution to this class of models. Second, we take advantage of the results about the structure of the solution to build a sequential Monte Carlo algorithm to evaluate the likelihood function of the model. The approach, which exploits the profusion of shocks in stochastic volatility models, is versatile and computationally tractable even in large-scale models, such as those often employed by policy-making institutions. As an application, we use our algorithm and Bayesian methods to estimate a business cycle model of the U.S. economy with both stochastic volatility and parameter drifting in monetary policy. Our application shows the importance of stochastic volatility in accounting for the dynamics of the data.

CS99 Room O2 STATISTICAL METHODS FOR MODELLING SPATIO-TEMPORAL RANDOM FIELDS Chair: Simone Padoan

C360: **Nonparametric simulation of the multivariate max-stable random vectors**

Presenter: **Philippe Naveau**, CNRS-IPSL, France

Co-authors: Giulia Marcon, Simone Padoan, Pietro Muliere

Many applications require simulation tools to reproduce the dependence of multivariate structure. This demand is particularly important for extreme events. Theoretically, multivariate extreme value theory suggests that the dependence structure among component size maxima should be non-parametric, but still obeys specific constraints. In this context, we propose a non-parametric simulation scheme to generate max-stable random vectors. Our algorithm is based on the estimation of the Pickands dependence function using Bernstein Polynomials. We discuss the properties of the proposed simulation method and illustrate its performance with a simulation study. We show the applicability of this framework in high-dimensional problems, analysing the dependence structure of heavy rainfall in France.

C479: **Fused Lasso for spatial and temporal quantile function estimation**

Presenter: **Ying Sun**, KAUST, Saudi Arabia

Co-authors: Montse Fuentes

Quantile functions are important in characterizing the entire probability distribution of a random variable, especially when the tail of a skewed distribution is of interest. The aim is to introduce new quantile function estimators for spatial and temporal data with a fused Lasso penalty to accommodate the dependence in space and time. This method penalizes the difference among neighboring quantiles, hence it is desirable for applications with features ordered in time or space without replicated observations. To better characterize local dependence, temporal and spatial fused adaptive Lasso estimators are also proposed by introducing adaptive weights to the fused Lasso penalty. The performance of the proposed methods are evaluated by simulations. Our methodology is applied to particulate matter (PM 2.5) data from the Community Multiscale Air Quality (CMAQ) model to characterize the upper quantile values which are crucial to studying spatial association between PM 2.5 concentrations and adverse human health effects.

C644: **Non-stationary dependence structures for spatial extremes**

Presenter: **Raphael Huser**, King Abdullah University of Science and Technology, Saudi Arabia

Co-authors: Marc Genton

Max-stable processes are natural models for spatial extremes, because they provide suitable asymptotic approximations to the distribution of maxima of random fields. In the recent past, several parametric families of stationary max-stable models have been developed, and fitted to various types of data. However, a recurrent problem is the modelling of non-stationarity. While it is fairly straightforward to build non-stationary models for marginal distributions, it is much less obvious to model non-stationarity in the dependence structure of extremal data, and there have been very few attempts to address this important issue so far. We discuss non-stationarity modelling for max-stable processes, with a view towards space-time applications, and show how inference can be performed using pairwise likelihoods. We also illustrate the methodology with an application to environmental data.

C695: **High-order composite likelihood inference for multivariate or spatial extremes**

Presenter: **Marc Genton**, KAUST, Saudi Arabia

Co-authors: Stefano Castruccio, Raphael Huser

Likelihood inference for multivariate and spatial extremes is among the most challenging problems in computational statistics, and current approaches typically rely on less expensive composite likelihoods constructed from small subsets of data. We show how, with state-of-the-art computational facility and an extensive use of parallel computing, it is possible to perform full likelihood inference for small data sets and to efficiently evaluate high-order composite likelihoods. With extensive simulations we assess, for the first time,the loss of information of composite likelihood estimators with respect to a full likelihood approach for some widely-used multivariate and spatial extreme models, and also discuss how to choose low-order composite likelihood truncation in order to improve estimation efficiency.

CS107 Room F2 ENERGY ECONOMICS Chair: Francesco Ravazzolo

C1041: **Oil and the great moderation in a regime switching framework**

Presenter: **Vegard Larsen**, BI Norwegian Business School, Norway

Co-authors: Hilde C Bjornland

The role of oil prices in reducing US macroeconomic volatility is revisited. We address the question by extending a New-Keynesian model similar to one developed by Blanchard and Gali to allow for parameters and volatility of shocks to change throughout the sample. Using a Markov Switching Rational Expectations (MSRE) framework, all the break points between the different regimes are determined endogenously within the model. Doing so we find that smaller and/or fewer oil price shocks did not play any major role in explaining the so called "Great Moderation". Indeed, oil prices have had several periods of lower volatility, yet these periods do not coincide with the volatility reduction of the "Great Moderation".

C1067: **Commodity prices, fiscal policy design and economic activity**

Presenter: **Leif Anders Thorsrud**, Norwegian Business School BI, Norway

Co-authors: Hilde Bjornland

In countries where the resource revenue constitutes a large component of total government revenues public spending will be directly sensitive to fluctuations in the price of the resource. For this reason Norway and Australia adapted explicit fiscal strategies during the early 2000's to shield their economies from disturbances in the commodity market. We develop a Bayesian Dynamic Factor Model with time varying parameters and analyze how fiscal policy design affects the way commodity prices influence public spending and economic activity across time in Norway and Australia. The model allows for resource movements and spending effects through a large panel of variables at the sectoral level, while also controlling for disturbances to global activity. Unsurprisingly, and in accordance with the changes in fiscal strategies, we find large time variation in how systematic fiscal policy responds to disturbances in the commodity market. Surprisingly, our results show that during the recent period of booming commodity prices procyclical fiscal policy works to exacerbate the fluctuations in economic activity in both countries. However, in the same period, but after a global activity shock which also increase commodity prices, Norway runs pro-cyclical policy, while Australia runs counter-cyclical fiscal spending, thereby insulating the economy from a spending boom.

C1092: **Oil price shocks and the UK economy, 1990-2005**

Presenter: **Marco Lorusso**, University of Glasgow, United Kingdom

Co-authors: Charles Nolan

The purpose is to estimate via Bayesian technique an open economy two-bloc DSGE model in order to study the transmission and impact of oil shocks on the UK economy and the rest of the world. Our sample choice encapsulates the Non-Inflationary, Consistently Expansionary (NICE) decade and corresponds to the period in which UK was a net oil exporter. The observed variables for the foreign bloc are constructed using trade weights and aggregating data of the main trading partners of UK. In line with previous literature, our findings confirm the overall importance of global oil demand and supply factors for the UK. More specifically, the source of oil price shocks matters. Contrary to previous literature that predicted an appreciation of the British currency against other currencies in the presence of an increase in the oil price, we find something different. In response to a drop in foreign oil production British sterling appreciates, whereas a reduction of foreign oil intensity implies a depreciation of the UK currency. In general, negative shocks to foreign oil supply and intensity imply an improvement of the UK trade balance reflecting the UK's position as an oil producer.

C1201: **Monetary policy in a simple new Keynesian model of a small open oil-exporting economy**

Presenter: **Martin Seneca**, Norges Bank, Norway

Co-authors: Andrea Ferrero

How should monetary policy respond to an oil price shock in a small open oil-exporting economy? We build a simple model of a small open economy that is totally dependent on exports of oil for foreign currency revenue to provide an answer to this question. Motivated by Norway, we let the non-oil economy be shielded from the use of oil revenues through a sovereign wealth fund and a fiscal policy rule, while home firms have a monopoly on the supply of intermediate input to oil extraction. We derive a loss function as a second-order approximation to household utility and consider optimal monetary policy. While there are substantial spill-overs from the off-shore oil sector to the mainland economy, the dependence on oil does not introduce additional trade-offs for monetary policy. The central bank should respond to an oil boom by increasing interest rates to bring domestic inflation down. But the size of the trade-off between domestic inflation and the welfare-relevant output gap depends on the oil sector, and off-shore activity introduces cost-push effects from all shocks in the model. The sovereign wealth fund does not affect the nature of the trade-off as long as government spending is exogenous.

Sunday 7.12.2014 16:55 - 19:00 Parallel Session N – ERCIM

ES18 Room H1 STATISTICAL METHODS FOR SPATIAL EPIDEMIOLOGY Chair: Peter Congdon

E284: **Small-area spatioemporal analysis of pedestrian injuries in New York City 2001-2010**
Presenter: **Charles DiMaggio**, Columbia University, United States

The aim was to measure the spatiotemporal risk of pedestrian injury in New York City at the census tract level over a recent 10-year period. Crash data on 140,835 pedestrian injuries in 1,908 census tracts from 2001 to 2010 were obtained from the New York City Department of Transportation. Injury counts within census tracts were analyzed with a Bayesian hierarchical spatiotemporal model using Integrated Nested Laplace Approximations (INLA). The yearly rate of of pedestrian crashes decreased 16.2% over the study period, from 23.7 per 10,000 to 16.2 per 10,000, with much of the decrease occurring during the first four years. Every one unit increase in a social deprivation index was associated with a 19% increase in pedestrian injury risk (Incidence Density Ratio (IDR) = 1.19, 95% CrI 1.16, 1.23), and for every one standardized unit increase in traffic density, there was a 20% increase in pedestrian injury risk (IDR=1.20, 95% CrI 1.15, 1.26). Some census tracts in New York City did not benefit from the over-all decrease in risk, or experienced sporadic years of increased risk compared to the city as a whole.

E371: **Exact calculation of multiplicity-adjusted p-values of scan statistics in spatial epidemiology**
Presenter: **Satoshi Kuriki**, The Institute of Statistical Mathematics, Japan
Co-authors: Kunihiko Takahashi, Hisayuki Hara

In spatial epidemiology, the multiplicity-adjusted p-values of spacial scan statistics to detect hotspots are estimated by Monte Carlo simulation. However, this simulation method cannot provide a precise estimate when the true p-value is small. We propose an algorithm to calculate the exact p-value under multinomial distribution. In other words, we propose an "exact test" for detecting hotspots. The outline of our method is as follows: (i) We first construct an undirected graph which represents the spatial information of the locations where the data are observed, then (ii) extract a Markov structure among spatial scan statistics, and finally (iii) calculate the multiplicity-adjusted p-values by successive integration (summation) technique based on the Markov structure. In step (ii), the theory of chordal graph in graphical models is used. The technique of step (iii) is similar to the one used in change-point analysis. The computational complexity of proposed algorithm is evaluated. The problem of choosing a scan window is discussed in view of the computational time. Illustrative numerical examples are presented using a real data set and simulated data sets. These examples show the usefulness as well as the limitation of the proposed algorithm.

E486: **Data cloning in Bayesian disease mapping**
Presenter: **Ying MacNab**, University of British Columbia, Canada

Complex hierarchical models, Gaussian Markov random field priors, and statistical computation are at the essential core of Bayesian disease mapping. Data cloning will be presented, discussed, and illustrated for its potential role in Bayesian disease mapping. In particular, via data cloning, we discuss identification in complex model, the roles and options of hierarchical prior and prior selections, and statistical computing. We illustrate how data cloning theory and methods can facilitate deeper understanding of various competing models and prior options and bring new insight into the nature of intelligent smoothing and related statistical computing.

E505: **Geographic variation in population attributable fractions: adjusting for spatial confounding**
Presenter: **Peter Congdon**, QMUL, United Kingdom

The contribution of a particular risk factor to a disease may be quantified using the population attributable fraction (PAF). Implications of confounder adjustment in multilevel data for spatial variability in PAFs are considered. The focus is on the PAF for increased BMI as a risk factor for diabetes, with adjustment for both individual and contextual (spatial) confounders, and allowing for spatial heterogeneity in BMI impacts. Contextual adjustment includes known neighbourhood factors (e.g. area deprivation), and unobserved spatially clustered influences. A logit regression method to estimate PAFs allows for the impact of continuous BMI on diabetes, without simply dichotomising the BMI effect (e.g. between obese and non-obese patients). It is still important to consider the form of the exposure-outcome relationship, and a model with spatially varying linear impacts of BMI on diabetes is compared with one allowing potentially nonlinear gradients.

E731: **Bayesian geostatistical models for veterinary parasitological surveys under informative sampling**
Presenter: **Annibale Biggeri**, University of Florence, Italy
Co-authors: Lorenzo Cecconi, Dolores Catelan

An application of Bayesian model-based geostatistics in Veterinary Epidemiology is shown. The aim is to use posterior predictive probabilities of infection to tailor future prevalence surveys. We exemplify the proposed approach applying posterior predictive infection probabilities and related uncertainty to draw informative samples from the 8794 geo-referenced sheep farms of the Campania region (southern Italy). Parasitological data came from a first cross-sectional survey carried out to study the spatial distribution of selected helminths in sheep farms. A grid sampling was performed to select the farms for coprological examinations. We specified a Bayesian kriging model and designed three different preferential surveys: weighting by posterior predictive probabilities of infection, by prediction uncertainty or combining both criteria. We show how to analyze the second surveys and discuss the proposed samping strategy as a useful tool to address targeted infection control treatments and surveillance campaigns. This approach is easily extendable to other fields of research.

E759: **Profiling high risk areas in spatial surveillance**
Presenter: **Corrado Lagazio**, University of Genova, Italy
Co-authors: Dolores Catelan, Annibale Biggeri

Identification of areas at unusual high or low risk of disease is important for spatial surveillance, but it is only partly addressed by disease mapping. We use the classical Poisson-Gamma model for disease mapping to show how standard model checking procedures, namely cross-validated posterior predictive distributions, may also be used to highlight hot-spots (outliers) of disease. Evaluations will be based on a full range of informative prior distributions under a null hypothesis, supposed to be true for most of the observations, of absence of excess risk.

ES23 Room N1 DEPENDENCE MODELS AND COPULAS: APPLICATIONS Chair: Wolfgang Trutschnig

E1277: **Comparison of conditional distributions in portfolios of dependent risks**
Presenter: **Miguel Angel Sordo**, University of Cadiz, Spain
Co-authors: Alfonso Suarez-Llorens, Alfonso José Bello

Given a portfolio of risks, we study the marginal behavior of the i-th risk under an adverse event, such as an unusually large loss in the portfolio or, in the case of a portfolio with a positive dependence structure, to an unusually large loss for another risk. By considering some particular conditional risk distributions, we formalize, in several ways, the intuition that the i-th component of the portfolio is riskier when it is part of a positive dependent random vector than when it is considered alone. We also study, given two random vectors with a fixed dependence structure, the circumstances under which the existence of some stochastic orderings among their marginals implies an ordering among the corresponding conditional risk distributions.

E366: **Dynamic correlation matrices based on vines**
Presenter: **Benjamin Poignard**, ENSAE - CREST, France

Co-authors: Jean-David Fermanian

New dynamics for correlation matrices within the multivariate GARCH family are proposed, namely the Vine-GARCH model. The idea comes from the ability of partial correlations to parameterize nicely any correlation matrix. To do so, we specify a hierarchical structure of dependence thanks to a partial correlation vine, which enables us to represent any correlation matrix. Then we specify independent partial correlation trajectories. These partial correlation processes allow for generating automatically positive definite correlation matrices. Consequently, we avoid any normalization step and we do not rely on intricate parametrizations to satisfy constraints such as symmetry and positive definiteness. Furthermore, the Vine-GARCH approach tends to be parsimonious because the parametrization in terms of partial correlation vine enables us to set constraints to some dynamics without altering the other ones.

E568: **Impulse response analysis in a nonlinear world**

Presenter: **Enrico Foscolo**, Free University of Bozen-Bolzano, Italy

An impulse response function measures the time profile of the effect of a shock on the behavior of a time series, say Y_t. The prevalent form of the impulse response function (shortly, IRF) is defined as the difference between two realizations of Y_{t+j} identical up to $t-1$. It provides an answer to the question what is the effect of a shock of size v_t hitting the system at time t on the system at time $t+j$, given that no other shocks hit the system. We provide new insights into the impulse response analysis in at least two aspects. First, we propose a nonparametric bootstrap-based procedure for the computation of the conditional expectation of Y_{t+j}. Second, we involve the concept of the Copula function for dealing with the temporal dependence. In this framework empirical methods and Copulas play a key role in order to avoid restrictions on the form (linear or nonlinear) of the model for Y_t.

E723: **Switching GAS Copula models for systemic risk assessment**

Presenter: **Leopoldo Catania**, Sapienza University of Rome, Italy

Co-authors: Mauro Bernardi

Recent financial disasters emphasised the need to accurately predict extreme financial losses and their consequences for the institutions belonging to a given financial market. The ability of econometric models to predict extreme events strongly relies on their flexibility in modelling the highly nonlinear and asymmetric dependence observed in financial returns. We develop a new class of flexible Copula models where the evolution of the dependence parameters follow a Markov-Switching Generalised Autoregressive Score (S-GASC) dynamics. Maximum Likelihood estimation is performed using the Inference Functions for Margins (IFM) approach and a new version of the Expectation–Maximisation (EM) algorithm specifically tailored to this class of models. The S-GASC models are then used to estimate the Conditional Value-at-Risk (CoVaR), which is defined as the VaR of a given asset conditional on another asset (or portfolio) being in financial distress. We theoretically prove that the CoVaR risk measure of the one–step ahead predictive distribution of the S-GASC models is consistent with respect to the first order stochastic dominance criterion. An empirical investigation shows that the proposed S-GASC models outperform competitors using several VaR backtesting procedures.

E231: **Directional multivariate Value at Risk and copulas**

Presenter: **Raul Andres Torres Diaz**, Carlos III de Madrid, Spain

Co-authors: Rosa E. Lillo, Henry Laniado

In Insurance and Finance, the concept of Value at Risk (VaR) as a risk measure is widely used due to its good properties and easy interpretation. However, its extension to the multivariate framework is not unique since the definition of a multivariate quantile is not unique either and it depends on the partial order considered in R or on extending each of the properties of the univariate quantile. We introduce a definition of directional multivariate value at risk MVaR based on a notion of directional multivariate quantile recently introduced in the literature. The directional approach allows the manager to consider the external information or risk preferences in her/his analysis. The proposed multivariate VaR is a vector-valued measure with the same dimension as the underlying risk portfolio. We provide the main properties of this financial risk measure and the relation with some families of copulas for which it is possible to obtain closed forms for the proposed VaR. Comparisons with other alternative multivariate VaR given in the literature are also analyzed in terms of robustness.

ES30 Room A1 ADVANCES IN ROBUST DATA ANALYSIS Chair: Luis A. Garcia-Escudero

E471: **Industrial development and well-being of population: Robust regression with compositional data**

Presenter: **Valentin Todorov**, UNIDO, Austria

Co-authors: David Kepplinger, Shyam Upadhyaya

Industrial development affects people's lives in various ways. The level of industrialization, often measured by the manufacturing value added per capita, is highly correlated with many social indicators, including the Human Development Index (HDI). The relation between the level of industrialization and the HDI gives a basic overview, but to support policy-making and to gain a deeper understanding of the relation and its mechanics, a more detailed analysis is needed. One very important question is how the size of different sectors of the manufacturing industry influences the well-being of population. To reduce the influence of outliers on the results of the analysis we apply robust regression using MM-type estimates. The data are compositional in their nature, since the value added in the different manufacturing sectors add up to the total value added of the manufacturing industry. It would be therefore more realistic to assume that the relevant information is only contained in the ratios between the industrial sectors and adapt the robust regression to compositional data. In order to perform inference, the distribution of the parameters can be estimated using bootstrapping but since bootstrapping robust regression estimates can be problematic fast and robust bootstrap was developed.

E514: **Parameters selection in robust clustering**

Presenter: **Luis Angel Garcia-Escudero**, Universidad de Valladolid, Spain

Co-authors: Alfonso Gordaliza, Agustin Mayo-Iscar

TCLUST provides a clustering methodology that can be used for doing unsupervised classification when robustness is a main concern. This methodology is very flexible and exhibits nice theoretical/robustness performance. However, due to this high flexibility, the use of the TCLUST methodology requires specifying three parameters when applying it. Namely, the number of clusters, the trimming level and a constant controlling the maximum allowable difference in cluster scatters. These parameters are interrelated and a fully unsupervised choice of these three parameters is surely not advisable. However, it is clearly useful to rely on tools which can help the user to make a proper choice for them depending on the data set at hand. Some graphical tools have already been developed with this purpose. We review these graphical tools and some new more numerical (less subjective) tools are also commented.

E588: **Robust estimation in partially linear errors-in-variables models**

Presenter: **Ana Maria Bianco**, Universidad de Buenos Aires and CONICET, Argentina

Co-authors: Paula Spano

Errors-in-variables models derive from the analysis of regression models where the covariates have been measured with error. Among semi-parametric models, partially linear models have been extensively used due to their flexibility to model linear components in conjunction with non-parametric ones. We focus on partially linear models where the covariates of the linear part are measured with additive errors. We consider a robust family of estimators of the parametric and nonparametric components. The estimation procedure is based on a three-step algorithm, where robust orthogonal estimators of the linear component are combined with robust smoothing. We study some asymptotic properties of the proposal. A simulation study allows us to compare the performance of the proposed estimators with the classical approach for moderate size samples.

E735: **Finding the number of groups in model-based clustering through the restricted likelihood**
Presenter: **Marco Riani**, University of Parma, Italy
Co-authors: Luis Angel Garcia-Escudero, Alfonso Gordaliza, Agustin Mayo-Iscar, Andrea Cerioli

Finding the appropriate value for the number of groups is a crucial and difficult task in any clustering method. Distance-based methods often rely on heuristics like the Calinski and Harabasz index, while suitable versions of the likelihood, like the Integrated Classification Likelihood (ICL), are used in model-based clustering. However, the likelihood is known to be unbounded in the general heteroschedastic model for multivariate normal data, and restrictions on the group-specific covariance matrices are required. The goal is to develop a likelihood-based criterion for estimating the number of groups which takes into account these restrictions. Our proposal is based on an analytic derivation of the effect of popular constraints in heteroschetastic model-based clustering. Its superiority over classical ICL is shown through an extensive simulation study.

E192: **A Von Mises approximation for the small sample distribution of the asymmetrically trimmed mean. Applications**
Presenter: **Alfonso Garcia-Perez**, UNED, Spain

Recently, a very accurate approximation for the small sample distribution of the symmetrically trimmed mean has been obtained. We review this approximation and extend it to the asymmetrically trimmed mean. Thanks to this very accurate approximation, the new tool of the p value line can be used to choose the trimming fractions. Some applications of this are given to other more complex problems such as LTS regression.

ES34 Room B1 REGRESSION MODELS FOR EXTREME VALUES Chair: Stephane Girard

E045: **Estimation of extreme conditional quantiles through power transformation**
Presenter: **Huixia Judy Wang**, George Washington University, United States
Co-authors: Deyuan Li

The estimation of extreme conditional quantiles is an important issue in numerous disciplines. Quantile regression (QR) provides a natural way to capture the covariate effects at different tails of the response distribution. However, without any distributional assumptions, estimation from conventional QR is often unstable at the tails, especially for heavy-tailed distributions due to data sparsity. We develop a new three-stage estimation procedure that integrates QR and extreme value theory by estimating intermediate conditional quantiles using QR and extrapolating these estimates to tails based on extreme value theory. Using the power-transformed QR, the proposed method allows more flexibility than existing methods that rely on the linearity of quantiles on the original scale, while extending the applicability of parametric models to borrow information across covariates without resorting to nonparametric smoothing. In addition, we propose a test procedure to assess the commonality of extreme value index, which could be useful for obtaining more efficient estimation by sharing information across covariates. We establish the asymptotic properties of the proposed method and demonstrate its value through simulation study and the analysis of a medical cost dataset.

E058: **Lawson's algorithm and estimation of extreme conditional quantiles**
Presenter: **Keith Knight**, University of Toronto, Canada

Lawson's algorithm is an iteratively reweighted least squares algorithm that can be used to compute L_∞ estimates in a linear regression model. We will explore the use of a modification of this algorithm to compute parametric and non-parametric estimates of extreme conditional quantiles.

E075: **Estimation of the conditional tail index in presence of random covariates**
Presenter: **Laurent Gardes**, University of Strasbourg, France
Co-authors: Gilles Stupfler

Studying extreme events is relevant in numerous fields of statistical application (hydrology, actuarial science among many others). Most of the time, extreme value analysis focuses on heavy-tailed distributions. In this situation, the so-called tail index is an important parameter that controls the behavior of the tail distribution and is thus of primary interest particularly to estimate extreme quantiles. The estimation of the tail index is considered in the presence of a finite-dimensional random covariate. Uniform weak consistency and asymptotic normality of the proposed estimator are established and some illustrations on simulations are provided.

E140: **Averaged extreme regression quantile**
Presenter: **Jana Jureckova**, Charles University in Prague, Czech Republic

The extreme regression quantile in the linear model can be calculated in two steps: either by linear programming or to estimate the vector of slope components by a special R-estimator, and the intercept component as the maximal residual of responses from the R-estimator. The averaged extreme regression quantile, with weights equal to the regressors, after a standardization with the true parameter is equal to the weighted mean of p largest order statistics of model errors. These identities help to understand better the nature of the extreme regression quantiles.

E076: **Data envelope fitting with constrained polynomial splines**
Presenter: **Abdelaati Daouia**, Toulouse School of Economics, France
Co-authors: Hohsuk Noh, Byeong U. Park

Estimation of support frontiers and boundaries often involves monotone and/or concave edge data smoothing. This estimation problem arises in various unrelated contexts, such as optimal cost and production assessments in econometrics and master curve prediction in the reliability programs of nuclear reactors. Very few constrained estimators of the support boundary of a bivariate distribution have been introduced in the literature. They are based on simple envelopment techniques which often suffer from lack of precision and smoothness. We develop a novel constrained fit of the boundary curve which benefits from the smoothness of quadratic spline approximation and the computational efficiency of linear programs. Using cubic splines is also feasible and more attractive under multiple shape constraints; computing the optimal spline smoother is then formulated into a second-order cone programming problem. Both constrained quadratic and cubic spline frontiers have a similar level of computational complexity to the unconstrained fits and inherit their asymptotic properties. The utility of this method is illustrated through applications to some real datasets and simulation evidence is also presented to show its superiority over the best known methods.

ES52 Room G1 RECENT DEVELOPMENTS IN THE DESIGN OF EXPERIMENTS Chair: Chang-Yun Lin

E033: **Optimal blocked orthogonal arrays**
Presenter: **Chang-Yun Lin**, National Chung Hsing University, Taiwan

Blocking is an important technique to reduce the noises introduced from uncontrollable variables. Many optimal blocking schemes have been proposed in the literature but there is no single approach that can be applied to various blocked designs. We construct a mathematical framework using the count function (or the so-called indicator function) and develop a comprehensive methodology which allows us to select various optimal blocked orthogonal arrays: regular or non-regular designs with qualitative, quantitative or mixed-type factors of two, three, higher or mixed levels. Under this framework, most existing approaches are special cases of our methodology.

E292: **Recent developments in optimal experimental designs for functional MRI**
Presenter: **MingHung Kao**, Arizona State University, United States

Functional magnetic resonance imaging (fMRI) is one of the leading brain mapping technologies for studying brain activity in response to mental stimuli. For neuroimaging studies utilizing this pioneering technology, high-quality experimental designs help to collect informative data to make precise and valid statistical inference about functions of the brain, and are very important. We present some recently developed analytical and

computational results on fMRI designs with a focus on the former. The performance of some traditionally popular fMRI designs as well as some newly developed fMRI designs is discussed.

E468: **Response surface methodology using split-plot definitive screening designs**
Presenter: **Po Yang**, University of Manitoba, Canada
Co-authors: Chang-Yun Lin
Definitive screening designs are a new class of three-level designs. We investigate the performance of definitive screening designs in split-plot structures for one-step response surface methodology. The result of the projection eligibility and the study of D-efficiency and I-efficiency show that split-plot definitive screening designs perform well when the number of important factors is small. To reduce the risk of being unable to fit second-order models for response surfaces, we provide the column indexes of projections. Experimenters can assign potentially important factors to those columns to avoid ineligible projections. An example is presented to demonstrate how to analyze data for response surface methodology using the split-plot definitive screening design.

E645: **Algorithms for factorial designs generation**
Presenter: **Roberto Fontana**, Politecnico di Torino, Italy
An overview of recent algorithms for orthogonal fractional factorial and *D*-optimal designs generation is provided. In the first part we focus on *minimum size* orthogonal fractional factorial designs (OFFDs) generation. We present an algorithm based on the joint use of polynomial counting function and complex coding of levels. It is worth noting that the algorithm puts no restriction either on the number of levels of each factor or on the orthogonality constraints and so it can be applied to a very wide range of designs, including mixed level orthogonal arrays. Then we study optimal *saturated* designs, mainly *D*-optimal designs. Widely used statistical software make available efficient algorithms for finding an optimal design according to the user's specifications. Nevertheless, these algorithms do not guarantee a *global* optimal design. Indeed, they start from an initial randomly chosen design and find a local optimal design. If the initial design is changed, the optimum found will, in general, be different. A natural question arises. Should we stop at the design found or should we run the algorithm again in search of a better design? Very recent methods and software for discovery probability are used to support the decision to continue or stop the sampling. Finally, we briefly discuss an algorithm for uniform random generation of Latin squares and Sudoku designs, which is based on the maximum cliques of properly defined graphs.

E1035: **Computational complexity in determining efficient fractional factorial designs in presence of model uncertainty**
Presenter: **Subir Ghosh**, University of California, United States
Determining an efficient fractional factorial design for a model known to describe adequately the data to be collected from an experiment is challenging but determining an efficient design in the presence of model uncertainty is even more challenging in terms of its computational complexity. The aim is to go into highlighting these complexities and give the lists of efficient designs. The efficiency performance comparisons are made for the designs in the model certainty with the model uncertainty.

E1273: **Design and analysis of discrete choice experiments for models with response time**
Presenter: **William Li**, University of Minnesota, United States
Co-authors: Stefano Barone, Alberto Lombardo, Deqiang Zou
In choice experiments respondents undergo a questionnaire which is nowadays mostly submitted through the internet. One method of analysis adopts the Multinomial Logit (MLN) model. We show that the MLN analysis can be enhanced by using an additional response which can be easily observed and recorded by electronically submitted questionnaires. In practice, modern survey platforms like "Qualtrics" can be used to record the so called "response latency", i.e. the time taken by the respondent to make the choice and select the most preferred profile in the choice set. We use the response latency model to deduce the relative weight of importance of the profiles for each choice set and respondent. This type of response can be used in place of the simpler and less informative dichotomous choice variable in the MLN model. As a result, a more reliable estimate of the optimal profile can be obtained, implying lower risks for new investments and marketing decisions. A comparison between the proposed model and the existing MNL model is made.

ES55 Room F1 NEW SOFTWARE DEVELOPMENTS FOR COMPETING RISKS AND MULTI-STATE MODELS Chair: Luis Meira-Machado

E037: **Fitting regression models to interval-censored observations of illness-death models in R**
Presenter: **Thomas Alexander Gerds**, University of Copenhagen, Denmark
Interval censored observations occur in survival analysis when it is not possible to observe the event time exactly. For example, in medical statistics the date of onset of cancer can typically only be determined to lie within the period between two subsequent clinical examinations of the patient. We consider interval censored event times gathered from repeated follow-up examinations in the case where different subjects are allowed to have different examination schedules. When the aim is to analyze factors that affect the event rate (e.g., the rate of cancer), competing risks, such as death due to other causes, need to be taken into account. The statistical problem is particularly challenging when in subjects for whom the competing event is observed, it remains unknown whether or not the event of interest has occurred in the period since the last clinical examination. Thus, even if the interest focuses on parameters that describe the risk of the event, say onset of cancer, we often need to consider an illness-death model for a proper analysis. The aim is to introduce the R-packages prodlim and SmoothHazard which implement tools for graphical presentation of the data and models, algorithms to obtain the nonparametric maximum likelihood estimator, and parametric as well as semi-parametric regression models which we adapted to this context.

E055: **An overview of multistate modeling with R**
Presenter: **Maja Pohar Perme**, University of Ljubljana, Slovenia
Event history analysis deals with data obtained by observing individuals over time focusing on events occurring for the individuals. Thus, typical outcome data consist on times of occurrence of events and on the types of events which occurred. Frequently, an event may be considered as a transition from one state to another and, therefore, multi-state models will often provide a relevant modeling framework for event history data. We will give a review of such methods and their implementation in R with emphasis on regression models for both transition intensities and transition- and state occupation probabilities.

E086: **Sparse multistate modelling**
Presenter: **Holger Reulen**, University of Goettingen, Germany
Analysis of 'time to event' is commonly known as the specification of risks for a single type of transition from one unique entry state to one unique exit state of a categorical system over time. Multistate models generalize this concept by allowing for the estimation of risks for multiple transition types. This may lead to a highly parametrized estimation task if each of the transition type specific risks is specified with several distinct covariate effects. Aiming for the estimation of effects equal to zero is a first option to get to a more parsimonious representation of a multistate model. By the estimation of two effects equal to each other, we yield a second option to decrease the parameter space dimensionality of the model. More formal, we have the opportunities to estimate $\beta_{p,q} = 0$, and/or $\beta_{p,q} = \beta_{p,q'}$, with transition types $q \neq q'$, and covariates x_p. The aim is to introduce two recently developed estimation approaches (and their accompanying R packages) providing sparse multistate modelling by the above principles: functional gradient descent boosting for multistate models (R-package `gamboostMSM`) and estimation of multistate models with L_1-penalized effects and effect differences (R-package `penMSM`).

E111: **The mstate package for non- and semi-parametric multi-state models, applied to a treatment success model**
Presenter: **Liesbeth de Wreede**, LUMC, Netherlands
Co-authors: Marta Fiocco, Johannes Schetelig, Hein Putter

The mstate package in R has been developed for the analysis of non- and semi-parametric (Cox model based) competing risks and multi-state models. The package contains functions for all steps of the analysis, from constructing the model and preparing the data to plotting patient-specific curves of the predicted transition probabilities. Models with different covariate effects (either the same for all transitions or transition-specific) and different baseline hazards assumptions (either different for all transitions or equal for some) can be fitted by means of mstate. Moreover, dynamic predictions, incorporating information about intermediate events, can also be computed. We will illustrate several modelling choices and the use of mstate for a multi-center cohort of CLL (chronic lymphocytic leukemia) patients who have undergone a stem cell transplantation. This dataset, collected by the European Society for Blood and Marrow Transplantation, contains information about pre-transplant risk factors and post-transplant events. We will focus on the probability of treatment success and the impact of covariates on this outcome. Our example shows the utility of applying a multi-state model for the analysis of complex clinical outcomes. However, careful consideration of clinical aspects, data quality and model choices are necessary to make applications successful.

E744: **Application of multistate models in hospital epidemiology: a software perspective**
Presenter: **Tobias Bluhmki**, University of Ulm, Germany
Co-authors: Martin Schumacher, Harbarth Stephan, Jan Beyersmann, Arthur Allignol

The use of multistate models has started to rise in hospital epidemiology. This is motivated by the fact that they provide the tools both to show that timing of events must be modelled and to actually model this timing. Indeed, when considering hospital-acquired infections (HI), patients are usually free of infection when they are admitted to hospital but may acquire one as time progresses. HI constitutes a major complication that can be severe both in terms of mortality and morbidity and often leads to a prolonged length of stay (LoS) in the hospital, which is one of the main drivers for extra costs due to infections. An estimator of the excess LoS associated with an infection can be obtained as a functional of the matrix of transition probabilities, which is in turn based on the transition hazards. The use of the flexible pseudo observation regression technique for the excess LoS was recently proposed. We illustrate the use of these methods and associated software with a recent study on HI consisting of more than 600,000 patients admitted in 10 west-European hospitals between 2011 and 2012. Considerations regarding the size of the data will also be given.

ES70 Room M1 NONPARAMETRIC FUNCTIONAL DATA ANALYSIS Chair: Alicia Nieto-Reyes

E291: **Modeling repeated functional data, with applications to trajectories of fertility**
Presenter: **Hans-Georg Mueller**, University of California Davis, United States
Co-authors: Kehui Chen , Pedro Delicado

Repeatedly observed functional data are encountered in various applications. These include demographic trajectories observed for each calendar year. A previous conditional double functional principal component approach to represent such processes poses complex problems for both theory and applications. A simpler and more interpretable approach can be based on a marginal rather than conditional functional principal component representation of the underlying function-valued processes. An additional assumption of common principal components leads to the special case of a simple tensor product representation. For samples of independent realizations of the underlying function-valued stochastic process, this approach leads to straightforward fitting methods for obtaining the components of these models. The resulting estimates can be shown to satisfy asymptotic consistency properties. The proposed methods are illustrated with an application to trajectories of fertility that are repeatedly observed over many calendar years for 17 countries.

E389: **Correlation median for functions**
Presenter: **Rosa Lillo**, Universidad Carlos III de Madrid, Spain
Co-authors: Dalia Valencia, Juan Romo

The correlation median for functions is introduced as a correlation coefficient that yields a representative curve of dependence between two samples of functional data. This coefficient is a robust alternative to the cross-correlation studied by Ramsay and Silverman. It is necessary to define both the median absolute deviation from the median (MAD) and the comedian to the functional context before introducing the definition of correlation coefficient. The MAD and comedian are constructed using the median of the data instead of the mean of the data. Since the most extended possibility for calculating the median of a set of curves is based on a depth measure, the MAD and comedian for functions are defined in terms of the modified band depth. Some properties of the functional correlation median are provided as well as a robustness analysis. These concepts are illustrated with simulated and real data.

E519: **Selection of variable and classification boundary by functional logistic regression**
Presenter: **Hidetoshi Matsui**, Kyushu University, Japan

Penalties with an ℓ_1 norm provide solutions in which some coefficients are exactly zero and can be used for selecting variables in regression settings. We focus on the form of the ℓ_1 penalty in logistic regression models for functional data, in particular, their use in classifying functions into three or more groups while simultaneously selecting variables or classification boundaries. Data from repeated measurements are represented by basis expansions, and the functional logistic regression model is estimated by the penalized maximum likelihood method with the help of ℓ_1 type penalties. By extending the existing penalties, we propose a new class of penalties that appropriately estimate and select variables or boundaries for the functional multiclass logistic regression model. The value of the tuning parameter included in the penalized likelihood is selected by a model selection criterion. Analysis of real data show that the form of the penalty should be selected in accordance with the purpose of the analysis.

E665: **Registration of distributions with Wasserstein distance**
Presenter: **Jean-Michel Loubes**, University of Toulouse, France
Co-authors: Helene Lescornel

A parametric deformation model for distributions is considered. More precisely, we assume we observe J sample of random variables which are warped from an unknown distribution template, i.e we observe $X_{ij} = \varphi_{\theta_j^\star}(\varepsilon_{ij}) \quad 1 \leq i \leq n, \quad 1 \leq j \leq J$, where $\theta_j^\star$ is the unknown deformation parameter in $\Lambda \subset \mathbb{R}^p$, associated with the j-th sample $(X_{1j}, \ldots, X_{nj})$, and ε_{ij} are i.i.d r.v.'s. We tackle the problem of estimating the individual deformation parameters $\theta_j^\star$. For this, we construct a registering criterion based on the Wasserstein distance to quantify the alignment of the distributions. We prove consistency of the empirical estimators.

E738: **Additive mixed models for generalized functional data**
Presenter: **Fabian Scheipl**, Ludwig-Maximilians-Universitaet Muenchen, Germany
Co-authors: Jan Gertheiss, Sonja Greven

The aim is to propose and evaluate an extensive framework for additive regression models for correlated functional responses from exponential families whose natural parameter varies smoothly over the functions' arguments. Our proposal allows for multiple partially nested or crossed functional random effects with flexible correlation structures as well as linear and nonlinear effects of functional and scalar covariates that may vary smoothly over the argument of the functional response. It accommodates densely or sparsely or irregularly observed functional responses and

predictors which may be observed with additional error and includes both spline-based and functional principal component-based terms. Estimation and inference in this framework is based on standard generalized additive mixed models, allowing us to take advantage of established methods and robust, flexible algorithms.

ES76 Room D1 STATISTICAL METHODS FOR INSURANCE AND FINANCE Chair: Cira Perna

E448: **The three-way Lee-Carter model to detect the leading causes of death**
Presenter: **Maria Russolillo**, University of Salerno, Italy
Co-authors: Giuseppe Giordano, Steven Haberman

In the last century developed countries have been affected by two complementary trends: an increasing survival in higher ages and a smaller number of births. The combination of these two phenomena largely impacts both on the costs of the social security public systems and on the societies dealing with the rising pension and healthcare costs. In spite of the common trend given by the ageing population and the growing longevity, the mortality rates are also influenced by gender, countries, ethnicity, income, wealth, causes of death and so on. According to the WHO a "right" recognition of the causes of death is important for forecasting more accurately mortality. In this framework we intend to investigate the main causes of death impacting on the upcoming human survival, throughout a Multi-dimensional Data Analysis approach to the Lee Carter model of mortality trends. In a previous paper, we stated that the crude mortality data can be considered according to several criteria. We take into account a three way array holding mortality data structured by time, age-group and causes of death. The model decomposition we propose is a modified version of the classical Lee Carter model allowing for three-way data treatment, analysis of residuals, and graphical representation of the different components. A case study based on actual data will be discussed.

E792: **Mortality for causes-specific deaths in Mortgage Life Insurance**
Presenter: **Vincenzo Passannante**, University of Salerno, Italy
Co-authors: Valeria D'Amato, Marilena Sibillo

Many forms of insurance contracts secure a mortgage and this specific set is identified as Mortgage Life Insurance. One of the latest trends is to ensure the mortgage by an insurance policy covering the risk that the borrower suffers from particular critical diseases. The frame is the Accelerated Critical Illness, a policy just providing a benefit in the case of death and acceleration in the event of a critical illness. The aim is to link the Mortgage to a Critical Illness insurance policy covering the risk of cause-specific death. The new product will be structured and its main quantitative characteristic will be evaluated.

E793: **Some remarks on predictive accuracy in survival forecasting**
Presenter: **Albina Orlando**, National Research Council CNR, Italy
Co-authors: Emilia Di Lorenzo, Michele La Rocca, Marilena Sibillo, Cira Perna

Accuracy in survival forecasting is particularly required in financial valuations concerning pension annuity portfolios. These business are spread over very wide time intervals and are typically characterized by high numbers of contracts. Moreover, the longevity phenomenon, even if randomly moving, is in any case increasing and has to be taken into account, in order to avoid very dangerous underestimation of future liabilities. These two elements make the survival forecasting the insurer can perform at the issue time a crucial issue, very exiting and tackled in literature. The work moves just from the consideration of the stochastic character of the longevity phenomenon inside the more general survival trend. The analysis is based on the study of the ratios between the observed (real) data and the corresponding ones got as outcomes of the survival model chosen among the others. The aim to have ratios coinciding with the unity is realized in the stochastic environment in requiring that the process of the ratios tends to 1 in the long term. We construct a stochastic process for the ratios usable as a correction factor for the output of the survival model. Some graphs illustrate the idea and its practical impact.

E959: **Rapidly bounding the exceedance probabilities of high aggregate losses**
Presenter: **Isabella Gollini**, University of Bristol, United Kingdom
Co-authors: Jonathan Rougier

The task of assessing the righthand tail of an insurer's loss distribution is considered for some specified period, such as a year. We present and analyse six different approaches: four upper bounds, and two approximations. We examine these approaches under a variety of conditions, using a large event loss table for US hurricanes. For its combination of tightness and computational speed, we favour the Moment bound. We also consider the appropriate size of Monte Carlo simulations, and the imposition of a cap on single event losses. We strongly favour the Gamma distribution as a flexible model for single event losses, for its tractable form in all of the methods we analyse, its generalisability, and because of the ease with which a cap on losses can be incorporated.

E859: **Some critical insights on the unbiased efficient frontier a la Bodnar & Bodnar**
Presenter: **Claudio Pizzi**, University Ca Foscari of Venice, Italy
Co-authors: Marco Corazza

In a recent paper concerning the determination of the unbiased efficient frontier of a risky asset portfolio, the authors: - Derive an unbiased estimator of the efficient frontier; - Show that, with respect to this unbiased efficient frontier, the usual sample efficient frontier is overoptimistic in the sense that the latter underestimates the variance of each efficient portfolio; - Present an exact F-test on the efficient frontier. The main assumptions underlying these results are: 1. All the considered assets are risky; 2. The returns of these assets are independently distributed; 3. The returns of these assets are normally distributed. Both the independence of the returns and their normality are questionable. In particular, the independence appears quite unrealistic. With respect to this framework, we present the following critical insights: - We study the asymptotic behavior of the unbiased estimator of the efficient frontier; - We analyze the operational effectiveness of the unbiased estimator of the efficient frontier by a bootstrap approach; - We investigate some anomalous behavior of the unbiased estimator of the efficient frontier.

ES93 Room I1 INFERENCE UNDER IMPRECISE KNOWLEDGE Chair: Barbara Vantaggi

E236: **Estimation of numerical model discrepancy in inverse problems by particle methods**
Presenter: **Erkki Somersalo**, Case Western Reserve University, United States
Co-authors: Daniela Calvetti, Oliver Ernst

In inverse problems, when recast in the form of Bayesian inference, the likelihood density requires a careful modeling of the measurement uncertainties which, in addition to the exogenous noise, comprise also the model discrepancy. In applications arising from physics, an accurate theoretical model for the forward problem may be available, and the model discrepancy is mostly related to a numerical approximation of the model. An example of such problem is the electrical impedance tomography (EIT) problem, in which the forward model is given by a partial differential equation (PDE) based on electrostatics, and its solution is approximated using finite element methods (FEM). We discuss the statistical approximation of the numerical discrepancy due to coarse FEM discretization, and present a fast and light numerical scheme based on sequential approximate sampling algorithm that leads to an approximation of the unknown electrical impedance as well as the statistics of the modeling error.

E183: **Compatibility of conditional distributions**
Presenter: **Pietro Rigo**, University of Pavia, Italy
Co-authors: Patrizia Berti, Emanuela Dreassi

In various frameworks, to assess the joint distribution of a k-dimensional random vector $X = (X_1, \ldots, X_k)$, one selects some putative conditional distributions $Q_1, \ldots, Q_k$. Each Q_i is regarded as a possible (or putative) conditional distribution for X_i given $(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_k)$. Well known examples are Gibbs sampling, multiple imputation and spatial data modeling. Another example, even if controversial, is Bayesian inference via improper priors. However, these (and similar) procedures make sense only if $Q_1, \ldots, Q_k$ are compatible, in the sense that there is a joint distribution P for X with conditionals $Q_1, \ldots, Q_k$. Three types of compatibility results are discussed: (i) the X_i are assumed to take values in compact sets; (ii) the Q_i are supposed to have densities with respect to reference measures; (iii) the law P with conditionals $Q_1, \ldots, Q_k$ is requested to belong to some given class C of distributions. Two choices for C are considered, that is, $C = \{\text{exchangeable laws}\}$ and $C = \{\text{laws with identical univariate marginals}\}$.

E562: **The role of finite additivity in multiple prior Bayesian inference using a 2-mononotone prior capacity**
Presenter: **Davide Petturiti**, University of Perugia, Italy
Co-authors: Barbara Vantaggi

Bayesian inferential procedures essentially rely on the choice of a single complete prior probability but in some situations this task is quite problematic. Such an issue can be faced by replacing a single prior probability with a class of priors and then working with their envelopes. Following this line, in the general framework of coherent finitely additive conditional probabilities, we start from a likelihood function on a partition $\mathcal{L}$ and a prior information expressed by a 2-monotone capacity on the algebra $\mathcal{A}$ generated by $\mathcal{L}$. We provide a closed form expression for the lower envelope of the set of all the possible "posterior probabilities", first considering all the coherent extensions and then restricting to the subset of fully $\mathcal{L}$-disintegrable extensions. For fully $\mathcal{L}$-disintegrable extensions we mean any extension which is $\mathcal{L}$-disintegrable with respect to a conditional prior whose conditioning event ranges in the algebra generated by $\mathcal{A}$ and the "observations". Some considerations on the countable additive case are drawn.

E642: **Similarity measures taking into account the power of interactions among attributes**
Presenter: **Giulianella Coletti**, University of Perugia, Italy
Co-authors: Marco Baioletti, Davide Petturiti, Valentina Poggioni

The focus is on similarity measures for fuzzy data, i.e., for objects described in terms of degree of membership of attributes. The measures we introduce can model situations in which the attributes have different relevance, also taking into account the possible interactions among attributes expressed as weights. Following this idea we propose a generalization of the most known similarity measures for binary data and their classical fuzzy counterparts. This generalization imply some choices, starting from the t-norm to compute the fuzzy intersection and the fuzzy measure of fuzzy sets. We obtain a family of similarity measures with a high expressive power and in which it is possible to convey background knowledge about the application domain (through the weights). This expressiveness is due to the great freedom in assessing a large number of parameters but then this could cause a difficulty in eliciting the weights. To face this problem it is necessary to put some constraints on the weight function in order to reduce the number of parameters to be specified and we propose a method to automatically learn the weights from a classified dataset. We provide experimental analyses to show the effectiveness of the given measures.

E368: **Probabilistic inference and syllogisms**
Presenter: **Giuseppe Sanfilippo**, University of Palermo, Italy
Co-authors: Niki Pfeifer, Angelo Gilio

Traditionally, syllogisms are arguments with two premises and one conclusion which are constructed by propositions of the form "All S are P" and "At least one S is P" and their respective negated versions. We discuss probabilistic notions of the existential import and the basic sentences type. We develop an intuitively plausible version of the syllogisms that is able to deal with uncertainty, exceptions and nonmonotonicity. We develop a new semantics for categorical syllogisms that is based on subjective probability. Specifically, we propose de Finetti's principle of coherence and its generalization to lower and upper conditional probabilities as the fundamental corner stones for the new semantics. Coherence allows for dealing with partial and imprecise assessments. Moreover, it is especially suitable for handling zero antecedent probabilities (i.e., here conditioning events may have probability zero): This is relevant for studying the probabilistic interpretation of the existential import. Then, we generalize our probabilistic interpretation of the basic syllogistic concepts to construct probabilistic versions of selected syllogisms. Finally, we relate them to inference rules in nonmonotonic reasoning.

ES102 Room Q1 BAYESIAN NETWORKS IN OFFICIAL STATISTICS Chair: Paola Vicard

E392: **Bayesian networks for official statistics purposes: focus on data integration**
Presenter: **Mauro Scanu**, ISTAT, Italy
Co-authors: Paola Vicard

Official statistics is a general word that includes all the actions that a National Statistical Institute should perform to produce official data, from data collection up to data dissemination. Many of the problems that an official statistician should tackle can be represented as a decision problem that can be subject to the availability of new knowledge. This is the case of data imputation and editing, weights computation in post-stratification and the use of paradata in responsive designs. The objective is to review the latest developments of the use of Bayesian networks for official statistics purposes, with a focus on the problem of data integration. Two data integration problems will be considered: record linkage, when the data sets to merge observe two overlapping sets of units so that linkages should refer to observations belonging to the same unit, and statistical matching, when the sets of units observed in the two data sets are not overlapping, as in the case of two sample surveys. The aim is to give a unified view of the two problems in terms of Bayesian networks.

E282: **Advances in integrative causal analysis and connections to statistical matching**
Presenter: **Ioannis Tsamardinos**, University of Crete, Greece
Co-authors: Sofia Triantafillou, Lagani Vincenzo, Anna Roumpelaki

Integrative Causal Analysis (INCA) is the problem of inducing causal structure from multiple, heterogeneous datasets. The datasets can be heterogeneous for several reasons such as (a) being obtained under different experimental conditions or (b) measuring different variables. The main idea is to identify the set of possible causal models that simultaneously fit to all available datasets and reason with this set of causal models, e.g., identify their invariant characteristics. The presented INCA algorithms are solved by converting them to satisfiability problems (SAT) and exploiting the significant body of research in SAT solvers for efficiency. Among others, these ideas provide a new approach and a different perspective to statistical matching based on causality, substituting the Conditional Independence Assumption with causal-based ones such as Faithfulness of the distribution to a causal model. In large-scale experiments, algorithms based on INCA are shown to be complementary to statistical matching approaches and outperform them when the number of common variables is low on some tasks. INCA algorithms could also be used potentially to match observational with experimental datasets.

E577: **On increasing information quality (InfoQ) by calibration of official statistics**
Presenter: **Ron Kenett**, KPA Israel and University of Turin Italy, Israel
Co-authors: Luciana Dalla Valle

Information quality, or InfoQ, is defined by Kenett and Shmueli as "the potential of a data set to achieve a specific (scientific or practical) goal by using a given empirical analysis method". This concept is more broad and articulated than data and analysis quality. InfoQ is based on

the identification of four interacting components: the analysis goal, the data, the data analysis, and the utility, and it is assessed through eight dimensions: 1) data resolution, 2) data structure, 3) data integration, 4) temporal relevance, 5) generalizability, 6) chronology of data and goal, 7) operationalization and 8) communication. The use of copulas and Bayesian Networks allows us to calibrate official statistics and organizational or administrative data, strengthening the quality of the information derived from a survey, and enhancing InfoQ. It also provides an ability to conduct studies with dynamic updates using structured and unstructured data thus enhancing several of the InfoQ dimensions. It is illustrated, with different case studies, a novel strategy to increase InfoQ based on the integration of official statistics data using copulas and Bayesian Networks.

E449: **Correcting measurement errors with Bayesian networks when the variables are continuous and discrete**
Presenter: **Mauro Mezzini**, University Roma Tre, Italy
Co-authors: Daniela Marella, Paola Vicard

Measurement error is the difference between a feature value provided by the respondent and the corresponding true but unknown value. Measurement error is one of the main nonsampling error sources. Object-oriented Bayesian networks (OOBNs) have been recently proposed to model and correct measurement errors. In particular, the measurement error in a categorical variable is described by a mixed measurement model implemented in a BN. We apply this model to 2008 Survey on Household Income and Wealth (SHIW), conducted by Banca d'Italia. In particular, financial assets in SHIW are affected by misreporting of financial amounts with a prevalence of underreporting. The measurement error model is estimated using a validation sample. Specifically, the probability of an error is estimated from the validation sample using a BN model. The variables are mixed continuous and discrete. Hence, we need to discretize before learning the BN. This is a challenging problem since discretization should be performed accounting for the association structure; otherwise, if the continuous variables are independently discretized, their relation structure could be dramatically altered. Here we propose a methodology to learn the network while discretizing. Once the error probability is estimated, the overall measurement error model is implemented and microdata are imputed.

E329: **Bayesian network structural learning for complex survey data**
Presenter: **Flaminia Musella**, Roma Tre University, Italy
Co-authors: Daniela Marella, Paola Vicard

Bayesian Networks (BNs) are multivariate statistical models satisfying sets of conditional independence statements. Recently, BNs have been applied to official statistics problems. The association structure can be learnt from data by a sequence of independence and conditional independence tests using the PC algorithm. The learning process is based on the assumption of independent and identically distributed observations. This assumption is almost never valid for sample survey data since most of the commonly used survey designs employ stratification and/or cluster sampling and/or unequal selection probabilities. Then the design may be not ignorable and it must be taken into account in the learning process. Alternative procedures of Bayesian network structural learning for complex designs are becoming of interest. A PC correction is proposed for taking into account the sampling design complexity. In most cases, the design effects are provided only for the cells and for specific marginals of the contingency table. In such a situation the first-order Rao Scott corrections can be computed for those loglinear models admitting an explicit solution to the likelihood equations. Therefore, we focus on decomposable models and the subset of hierarchical loglinear models, typically used to investigate the association structure in terms of (conditional) independence between categorical variables.

ES113 Room C1 ADVANCES IN ORDINAL AND PREFERENCE DATA Chair: Jose Fernando Vera

E128: **Modelling "don't know" responses in rating scales**
Presenter: **Marica Manisera**, University of Brescia, Italy
Co-authors: Paola Zuccolotto

A probabilistic framework for the treatment of "don't know" (DK) responses in surveys aimed at investigating human perceptions through expressed ratings is proposed. The rationale behind the proposal is that DK is a valid response to all extents because it informs about a specific state of mind of the respondent, and therefore, it is not correct to treat it as a missing value, as it is usually treated. The actual insightfulness of our model depends on the chosen probability distributions. The required assumptions of these distributions first pertain to the expressed ratings and then to the state of mind of DK respondents toward the ratings. Regarding the former, we worked in the CUB model framework, while for the latter, we proposed using the Uniform distribution for formal and empirical reasons. We show that these two choices provide a solution that is both tractable and easy to interpret, where DK can be taken into account by simply adjusting one parameter in the model. A case study on Eurobarometer data is proposed. Results show that correcting for DK can be important in empirical studies when the study aims to compare several items, or groups of people expressing evaluations on a single item.

E249: **A new dimension for democracy: egalitarianism in the rank aggregation problem**
Presenter: **Federico Ricci Tersenghi**, Sapienza University of Rome, Italy
Co-authors: Pierluigi Contucci, Emanuele Panizzi, Alina Sirbu

In an election with three or more candidates, that are ranked by voters in order of preference, there are no univocal criteria for the selection of the winning (consensus) ranking and the outcome is known to depend sensibly on the adopted rule. Building upon XVIII century Condorcet theory, whose idea was to maximize total voter satisfaction, we propose the addition of a new basic principle (dimension) to guide the selection: satisfaction should be distributed among voters as equally as possible. With this new criterion we identify an optimal set of rankings. They range from the Condorcet solution to the one which is the most egalitarian with respect to the voters. We show that highly egalitarian rankings have the important property to be more stable with respect to fluctuations and that classical consensus rankings (Copeland, Tideman, Schulze) often turn out to be non-optimal. The new dimension we have introduced provides, when used together with that of Condorcet, a clear classification of all the possible rankings. By increasing awareness in selecting a consensus ranking our method may lead to social choices which are more egalitarian compared to those achieved by presently available voting systems.

E318: **A heuristic algorithm for the consensus ranking problem**
Presenter: **Sonia Amodio**, University of Naples Federico II, Italy
Co-authors: Antonio D'Ambrosio, Roberta Siciliano

Preference rankings virtually appear in all field of science (behavioural sciences, machine learning, decision making and so on). The well-known social choice problem consists in trying to find a reasonable procedure to use the aggregate preferences expressed by subjects (usually called judges) to reach a collective decision. This problem turns out to be equivalent to the problem of estimating the consensus (central) ranking from data that is NP-hard. A branch and bound algorithm has been previously proposed to calculate the consensus ranking given n rankings expressed on m objects. We propose a new algorithm to find the consensus ranking that is perfectly equivalent to the previous algorithm in terms of solutions reached but permits a remarkable saving in computational time.

E732: **A three-way structural equation multidimensional scaling model**
Presenter: **J Fernando Vera**, University of Granada, Spain

In the context of least squares MDS allowing transformations, a Structural Equation Multidimensional Scaling model is proposed for three-way one-mode dissimilarity data, in order to estimate the unknown subjacent dissimilarity matrix while simultaneously the objects are represented in a low dimensional space. Considering different dissimilarity matrices as indicators of the unobserved dissimilarities, an alternating estimation procedure is proposed in which the unknown conjoint dissimilarity matrix is estimated in a covariance structure framework. The model also addresses the

important situation of two-way one-mode asymmetric dissimilarity data, in order to estimate the unknown symmetric subjacent dissimilarity matrix while simultaneously the objects are represented in a low dimensional space.

E971: **The effects of objective and subjective conjoint vulnerabilities on life satisfaction**
Presenter: **Alfonso Piscitelli**, Federico II University of Naples, Italy
Co-authors: Adolfo Morrone

One of the most important aspects in evaluating the quality of life is to take into account the effect of having multiple disadvantages since they far exceed the sum of their individual effects. The microdata of the annual Istat survey on "Everyday life" have information on the "joint distribution" of the most salient features of quality of life. We address the problem of measuring the cumulative effects of objective and subjective conjoint vulnerabilities using as a target variable the subjective well-being which is viewed as multidimensional variable influenced by objective situation of distress. We analysed the features of quality of life in the microdata of the survey on "Everyday life", using nonlinear regression. The linear regression is restricted to reveal only relations showing a linear trend. Because of this restriction, regression methods that can deal with nonlinear relations have become more and more popular. Estimation of regression models for nonlinear relations is more computationally complex and intensive than for the linear regression model. The results show different conditions which define the subjective well-being, identifying the contributions of the situations of distress on life satisfaction.

ES120 Room P1 CHANGE-POINT DETECTION, DEPENDENCE MODELLING AND RELATED ISSUES Chair: Ivan Kojadinovic

E298: **Change-point tests based on U-quantiles**
Presenter: **Daniel Vogel**, Ruhr Universitaet Bochum, Germany
Co-authors: Martin Wendler

Functional limit theorems for U-quantiles of stationary, but potentially dependent series are presented. We introduce the notion of near epoch dependence in probability (PNED), a very general weak dependence condition, which, in contrast to the traditional L_2 near epoch dependence, does not require the existence of any moments, making it particularly appealing in the area of robust statistics. The motivation is to develop robust change-point tests. Many popular robust estimators, like the Hodges-Lehmann estimator or the Q_n scale estimator, are U-quantiles. They are known to combine excellent robustness properties with a high efficiency at the Gaussian model. Our invariance principles provide the basis for the construction of change-point tests for a variety of robust estimators, including the ones mentioned. As an example, we consider in particular the problem of robustly detecting changes in the mean of a dependent series by means of the Hodges-Lehmann estimator. We illustrate the benefit of the new method at several data examples. Crucial for any application is the estimation of a long run variance parameter, for which a general solution is given.

E751: **From non-sequential to sequential**
Presenter: **Stanislav Volgushev**, Ruhr University Bochum, Germany
Co-authors: Xiaofeng Shao

In time series analysis, statistics based on collections of estimators computed from subsamples play a role in many applications. Proving results about the joint asymptotic distribution of such statistics is challenging, since it typically involves a nontrivial verification of technical conditions and tedious case-by-case asymptotic analysis. We provide a technique that allows to circumvent those problems in a general setting. Our approach consists of two major steps: a probabilistic part which is mainly concerned with weak convergence of sequential empirical processes, and an analytic part providing general ways to extend this weak convergence to smooth functionals of the sequential empirical process. The methodology is illustrated for several examples such as self-normalization and change-point detection.

E677: **Extreme value copula estimation based on block maxima of a multivariate stationary time series**
Presenter: **Axel Buecher**, Ruhr-Universitaet Bochum, Germany
Co-authors: Johan Segers

The core of the classical block maxima method consists of fitting an extreme value distribution to a sample of maxima over blocks extracted from an underlying series. In asymptotic theory, it is usually postulated that the block maxima are an independent random sample of an extreme value distribution. In practice however, block sizes are finite, so that the extreme value postulate will only hold approximately. A more accurate asymptotic framework is that of a triangular array of block maxima, the block size depending on the size of the underlying sample in such a way that both the block size and the number of blocks within that sample tend to infinity. The copula of the vector of componentwise maxima in a block is assumed to converge to a limit, which, under mild conditions, is then necessarily an extreme value copula. Under this setting and for absolutely regular stationary sequences, the empirical copula of the sample of vectors of block maxima is shown to be a consistent and asymptotically normal estimator for the limiting extreme value copula. Moreover, the empirical copula serves as a basis for rank-based, nonparametric estimation of the Pickands dependence function. The results are illustrated by a simulation study.

E317: **Goodness-of-fit tests for multivariate copulas under strong mixing based on a dependent multiplier bootstrap**
Presenter: **Betina Berghaus**, Ruhr University Bochum, Germany
Co-authors: Axel Buecher

In recent years, copulas became a standard tool for modeling multivariate stochastic dependence. We present goodness-of-fit tests for the copula based on observations from a strongly mixing, stationary time series. Being based on a dependent multiplier bootstrap, the method is a suitable adaptation of an existing testing procedure for i.i.d. data sets and shares the same computational benefits compared to methods based on a parametric bootstrap. As a measure of the goodness-of-fit, we consider a distance between the empirical copula function C_n and C_{θ_n}, where θ_n denotes a suitable parametric estimator and where C_{θ_n} is the corresponding copula within the parametric class. Besides being able to test the goodness-of-fit of the contemporaneous dependence, the new results can also be applied to validate models for the copulas describing the serial dependence. The latter is in particular useful for modeling data by Markovian copula models. The finite-sample performance of the testing procedure is illustrated by a simulation study.

E660: **A single-index model for conditional copula modeling**
Presenter: **Olivier Lopez**, CREST ENSAE Paris Tech, France
Co-authors: Jean-David Fermanian

Semiparametric estimation of a conditional copula function is performed, using a single-index model. This model consists of assuming that only an unknown linear combination of the covariates is required to compute the conditional copula function. This allows us to perform dimension reduction, while keeping some nonparametric flexibility. We provide asymptotic results for the consistency of the index parameter and of the final copula estimator.

ES124 Room E1 SURVIVAL AND RELIABILITY Chair: Juan Eloy Ruiz-Castro

E432: **Goodness-of fit for randomly censored data based on maximum correlation**
Presenter: **Ewa Strzalkowska-Kominiak**, Universidad Carlos III de Madrid, Spain
Co-authors: Aurea Grane

The maximum correlation, ρ^+, is the Frechet-Hoeffding upper bound of the correlation coefficient corresponding to the bivariate distribution

function $H^+(x,y) = \min(F_1(x), F_2(y))$. It is a measure of agreement between the distribution functions, F_1 and F_2, since $\rho^+ = 1$ if and only if $F_1 = F_2$ up to a scale and location change. We study the goodness-of-fit test based on ρ^+ in the context of randomly censored data under a general censoring mechanism and also in the case of the Koziol-Green model. We prove the asymptotic properties of the proposed test statistic. Additionally, we present an extensive simulation study on the empirical power which shows a good performance of our approach and its advantages over the most famous Pearson-Type test.

E612: **Inference based on local likelihoods for regression models of non-proportional hazards and excess hazards**

Presenter: **Giuliana Cortese**, University of Padua, Italy

Various extensions to the Cox proportional hazards model have been proposed to allow for nonproportionality of the hazards. In particular, some flexible hazards models include regression coefficients that may change over time. Recently, an approach based on local partial likelihoods has been proposed to estimate time-varying coefficients. Inferential procedures (confidence bands and hypothesis testing) based on local partial likelihood are presented as a diagnostic tool for investigating departure from the Cox model. Moreover, a flexible regression model for the excess hazard is presented as a special case and inference based on local linear likelihoods is also described for this model. Excess hazards model are employed when information on different specific causes of failure is missing or incomplete, and expresses the excess risk due to a specific event over the expected risk in the background population. In order to provide a better variance-bias trade-off, a semiparametric version of those models is considered, where both time-constant and time-varying coefficients can be included. The local linear estimation technique is then applied together with profile likelihood methods. The performance of those models and the related inference are investigated via simulation studies. An application to a breast cancer dataset from the Norwegian cancer registry is presented to illustrate advantages and disadvantages of the procedures from a practical point of view.

E343: **Matrix transformations in the analysis of systems subject to repairable and nonrepairable failures**

Presenter: **Charles Wells**, University of Dayton, United States

Motivation and a conceptual framework for systems which are subject to repairable and non-repairable failures is introduced. System metrics appropriate for such systems may not include classic measures such as reliability or availability. As a result, alternative system metrics are examined. Under the assumption that the time until a non-repairable failure follows a continuous distribution of phase-type, it is shown that matrix-valued Laplace transformations arise as a natural consequence of the calculation of system metrics. Some interesting properties of these transformations are provided, and computational examples are presented to demonstrate the tractability of the computations.

E827: **The discrete conditional phase-type model to predict, and plan for, retinopathy of prematurity**

Presenter: **Rebecca Rollins**, Queens University Belfast, United Kingdom

Co-authors: Adele H. Marshall, Karen Cairns, Eibhlin McLoone, Sarah Chamney

Retinopathy of prematurity (ROP) is a disease in which retinal blood vessels of premature infants fail to grow and develop normally, and is one of the major causes of childhood blindness in the developing and developed world. The infants at risk of this disease are cared for in a neonatal unit which, with its specialist care, is one of the most expensive wards to run and must be managed with the greatest efficiency. The Discrete Conditional Phase-type (DC-Ph) distribution consists of two components; the conditional component measuring the inter-relationships between covariates and the survival component which models the survival distribution using a Coxian phase-type distribution. Coxian phase-type distributions are a special type of Markov model that can represent survival as a series of transient states which terminate in the one absorbing state when the event of interest has been reached. For the first time a support vector machine (SVM) capable of classifying multiple outcomes is used to represent the conditional component of the DC-Ph model to identify the infant's risk of developing ROP along with the Coxian phase-type distribution to model and predict the infant's length of stay. The resulting model will serve as an aid to optimise ward resources.

E244: **A discrete piecewise multi-state model in survival: application to breast cancer**

Presenter: **Juan Eloy Ruiz-Castro**, University of Granada, Spain

Co-authors: Mariangela Zenga

The multi-state models have been considered in the survival field for modeling illnesses that evolve by passing through several stages. Multi-state models can be developed by applying several techniques, such as non-parametric, semi-parametric and stochastic processes, especially Markov processes. When the development of an illness is being analyzed, the evolution is followed in a periodic form. Medical revisions take place in discrete times and a panel data can be built. A discrete non-homogeneous piecewise Markov process is constructed for modeling and analyzing multi-state illness with a general number of states. The model is built and some measures of interest, such as survival function, transition probabilities, spent mean times in a group of states, conditional probability of state change, etcetera are given. We show that the time up to any state is phase-type distributed. The results are achieved in a matrix algebraic form and algorithms are shown. The model is applied for analyzing the behavior of breast cancer. The results have been implemented computationally with Matlab.

Monday 8.12.2014 08:45 - 10:05 Parallel Session O – CFE

CS09 Room N2 VOLATILITY MEASURING, MODELING AND APPLICATIONS Chair: Massimiliano Caporin

C450: **Network banks exposures and variance spillovers in the Euro area**
Presenter: **Monica Billio**, University of Venice, Italy
Co-authors: Massimiliano Caporin, Lorenzo Frattarolo, Loriana Pelizzon

A new model of European equity volatility and volatility spillovers based on financial bank connectedness is proposed. The system of connections among banks is extracted from banks foreign exposures provided by the BIS consolidated banking statistics from 2006 till 2013. Volatility and volatility spillovers are estimated using a proximity structure BEKK model where the proximity matrices are the row-normalized banks foreign exposures observed quarterly. The use bank exposures network, as proximity matrices, allows us to capture second order spillovers effects on the European equity volatility from the volatilities of other countries and from correlations. Those effects, due to the dimensionality curse, are usually neglected in modeling equity volatilities with a number of series greater than 2. Our analysis shows also a structural change in the volatility persistence at the end of 2006 driven by the changes in the structure of the network exposures. The persistence eigenvector analysis shows the relevance of Ireland banks exposures in the volatility spillover potentially due to the large number of foreign banks legally located in Ireland.

C511: **Markov switching models for volatility: filtering, approximation and duality**
Presenter: **Maddalena Cavicchioli**, University of Modena and Reggio Emilia, Italy
Co-authors: Monica Billio

A dual and novel relationship in the estimation of Markov Switching (MS) processes for volatility is shown. It is well known that MS-GARCH models suffer from path dependence which makes the estimation step unfeasible with usual Maximum Likelihood procedure. However, by rewriting the MS-GARCH model in a suitable linear State Space representation, we are able to give a unique framework to reconcile the estimation obtained by Kalman Filter with that coming from auxiliary models proposed in the literature. This result provides the missing link to justify the validity of these approximations of the original MS-GARCH model when conducting estimation. Reasoning in the same way, we present a linear Filter for MS-Stochastic Volatility (MS-SV) models on which different conditioning sets yield more flexibility in the estimation. Applications on simulated data and on short-term interest rates show the feasibility of the proposed approaches.

C077: **Heterogeneous investment horizons and jump risk in financial markets**
Presenter: **Deniz Erdemlioglu**, IESEG School of Management, France
Co-authors: Nikola Gradojevic

The transition of realized jump risk across heterogeneous trading horizons ranging from minutes to weeks is investigated. To capture such dynamics, we propose an estimation procedure that identifies jump risk regimes (high and low) between multiple time periods (short and long). Applying this method to high-frequency data on stocks, bonds and currencies, we find evidence of regime-switching and scale-dependent jump risk. While the duration of high jump risk is rather short at intra-day time intervals, jump risk regimes are more persistent at long investment horizons. Across all markets, the results indicate that the European bond market is more sensitive to negative (left-tail) jump risk, relative to the US bond market. We discuss the implications of the empirical results for the specification of jump risk (and premia) embedded in the VIX index.

CS29 Room O2 SOLUTION AND ESTIMATION OF NON-LINEAR GENERAL EQUILIBRIUM MODELS Chair: Luca Guerrieri

C108: **Measuring ambiguity aversion**
Presenter: **Mohammad Jahan-Parvar**, Federal Reserve Baord of Governors, United States
Co-authors: A. Ronald Gallant, Hening Liu

Recently proposed Bayesian statistical methods are used to estimate a class of asset pricing models featuring smooth ambiguity preferences and to assess the properties of the estimated model. The used statistical method is a refinement of the Generalized Scientific Model Bayesian methods, which can accommodate highly nonlinear macro-finance models. We provide the first measurement of the size of ambiguity aversion for smooth ambiguity preferences. One major issue with our exercise is the time needed for estimation. Our estimation is more time-consuming compared with similar studies. Our empirical findings imply that: (1) Standard Bayesian model comparisons favor smooth ambiguity aversion preferences over the Epstein-Zin's recursive preferences. (2) The estimate of ambiguity aversion parameter is reasonable. (3) Our benchmark model produces forecasts of the equity premium, consumption growth rates, and short-rates that are much closer to the data compared with the alternative model.

C497: **Sovereign risk and spillovers: Untangling the web in Europe**
Presenter: **Nora Traum**, North Carolina State University, United States
Co-authors: Huixin Bi

The aim is to characterize the extent to which macroeconomic fundamentals explain the rapid increase in long-term interest rate spreads of Eurozone nations' sovereign bonds against German sovereign bonds. We build a New Keynesian monetary union model, whose features include a banking sector, explicit fiscal policies of union members, and cross-union member trade and asset linkages. Each government finances transfers and expenditures by collecting distortionary income taxes and issuing bonds. We use Bayesian inference to estimate parameters of a first order approximation of the model in the post-EMU, pre-crisis period (1999 to 2007). We then turn to analyzing the model in the crisis period. We model fiscal limits representing the maximum levels of debt-to-GDP ratios that governments are politically able to service to be stochastic, and assume a limit's distribution follows a logistical function. We use the monotone mapping method for global solutions to solve the full nonlinear model incorporating the fiscal limits and quantify the various transmission channels of sovereign risk "spilling over" from one union member to another. Finally, we examine how the central bank should respond given the sovereign risk premia and what types of monetary policies help reduce the risks of spillovers.

C924: **Collateral constraints and macroeconomic asymmetries**
Presenter: **Luca Guerrieri**, Federal Reserve Board, United States

Using Bayesian methods, a nonlinear general equilibrium model is estimated where occasionally binding collateral constraints on housing wealth drive an asymmetry in the link between housing prices and economic activity. The estimated model shows that, as collateral constraints became slack during the housing boom of 2001-2006, expanding housing wealth made a small contribution to consumption growth. By contrast, the housing collapse that followed tightened the constraints and sharply exacerbated the recession of 2007-2009. The empirical relevance of this asymmetry is corroborated by evidence from state- and MSA-level data.

CS41 Room E2 APPLIED ECONOMIC ISSUES Chair: Lynda Khalaf

C394: **Within- and cross-country price dispersion in the Euro area**
Presenter: **Fabio Rumler**, Austrian Central Bank, Austria
Co-authors: Adam Reiff

Using a comprehensive data set on retail prices across the Euro area, within- and cross-country price dispersion in European countries is analyzed. First, we study price dispersion over time, by investigating the time-series evolution of the coefficient of variation, calculated from price levels.

Second, since we find that cross-sectional price dispersion by far dominates price dispersion over time, we study price dispersion across space and investigate the role of geographical barriers (distance and national borders). We find that (i) prices move together more closely in locations that are closer to each other; (ii) cross-country price dispersion is by an order of magnitude larger than within-country price dispersion, even after controlling for product heterogeneity; (iii) a large part of cross-country price differences can be explained by different tax rates, income levels and consumption intensities. In addition, we find some indication that price dispersion in the Euro area has declined since the inception of the Monetary Union.

C1264: **Non-standard confidence limits for ratios and tipping points, with applications to dynamic regressions and panel data**
Presenter: **Lynda Khalaf**, Carleton, Canada
Co-authors: Jean-Thomas Bernard, Ba Chu, Marcel Voia

Multivariate Fieller methods are extended beyond standard estimators. Special cases include dynamic univariate and panel regression with persistent covariates, long run relations and cointegration equations, and dynamic polynomial panels. For these models, we prove that the limiting distribution of statistics underlying multivariate Fieller methods is the same and standard whether covariates have unit roots, near unit roots or roots far from one. Consequently, the standard Fieller-type inferential framework can be used without taking a stand on the extent of persistence in covariates, and regardless of associated discontinuities in limiting distributions of underlying estimators. Proofs are derived more generally for asymptotically mixed normal estimators with unknown possibly random nuisance parameters. Taken collectively, results exploit fundamental properties of mixed Gaussian asymptotics that seem to have escaped notice for inference on ratios. The case of curve turning points is considered as an empirically relevant problem of interest, examples of which in economics include Kuznet, Laffer, Rahn, Engel, Beveridge curves, as well as statistical Phillips, Yield and wage curves. Since dynamic polynomials are often considered to fit curve turning or tipping points, we extend the MLE-based heterogenous panel method to accommodate polynomial covariates with possible unit roots. In this context, stability conditions that preserve the definition of long-run relations are derived leading to simple directives regarding the lag structure of covariates. A simulation study illustrates the finite sample properties of the proposed methods in dynamic polynomial panels. Our proposed methods work well with small samples even at the boundary of stability restrictions.

C1305: **Banks, lending and the transmission of monetary shocks**
Presenter: **Demetris Koursaros**, Cyprus University of Technology, Cyprus

The aim is to examine the way financial intermediaries can amplify the effects of monetary and other shocks, in an attempt to understand the behavior of macroeconomic variables during the recent financial crisis. I built a theoretical model where loan officers' perceived uncertainty on the projects seeking finance is procyclical. Loan officers evaluate their lending decisions each period which makes them perceive that their ability to find good loan opportunities is weakened during recessions and strengthened during expansions. Due to the non-convex nature of bank cash flows from a loan, changes in the ability of loan officers' to value loans coming from variations in uncertainty, result in sharp loan contractions during recessions. During a recession, uncertainty increases and banks find it difficult to identify worthy lending opportunities, contracting the flow of funds to entrepreneurs, exacerbating the downfall. Modern economies depend on financial intermediation and a change in uncertainty affects the lending and interest rate decisions of banks, which appears to have great impact on their growth and well-being.

CS69 Room B2 RISK PREMIA TIME SERIES Chair: Jeroen V.K Rombouts

C574: **On the estimation of variance risk premia**
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. We criticize these results and propose a class of flexible structural time series models which, using signal extraction techniques, estimate more precisely the market price of risk. We advocate the inclusion of interactions, 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.

C833: **The dynamics of the equity risk premium**
Presenter: **Christian Dorion**, HEC Montreal, Canada
Co-authors: Harjoat S. Bhamra, Alexandre Jeanneret

The dynamics of the equity risk premium through the lens of a consumption-based structural model are studied. The economy consists of heterogeneous firms that choose their optimal default policy and capital structure in the presence of agency conflicts. Firms refinance dynamically and exhibit time variation in financial leverage and default risk. We bring the model to the data and estimate a model-implied monthly measure of equity risk premium at the market level over the period 1929-2012. We find that the equity risk premium is countercyclical, varies with financial and economic uncertainty, depends on past default and refinancing clusters, and eventually helps predict future excess stock returns.

C1222: **Sparse change-point model**
Presenter: **Jeroen Rombouts**, ESSEC Business School, France

Change-point specifications constitute flexible models that capture structural changes by allowing for switches in the model parameters. Nevertheless most models suffer from an over-parametrization issue since typically only one latent state variable drives the breaks. This implies that all parameters have to change when a break happens. We introduce sparse change-point processes, a new approach for detecting which parameters change over time. We propose shrinkage prior distributions allowing to control model parsimony by limiting the number of parameters which evolve from one structural break to another. Well-known applications are revisited to emphasize that many popular breaks are, in fact, due to a change in only a subset of the model parameters. It also turns out that seizable forecasting improvements are made over standard change-point models.

CS75 Room C2 QUANTITATIVE RISK MANAGEMENT Chair: Mike K.P. So

C054: **Forecasting co-volatilities via factor models with asymmetry and long memory in realized covariance**
Presenter: **Manabu Asai**, Soka University, Japan
Co-authors: Michael McAleer

Modelling covariance structures is known to suffer from the curse of dimensionality. In order to avoid this problem for forecasting, a new factor multivariate stochastic volatility (fMSV) model for realized covariance measures is proposed that accommodates asymmetry and long memory. Using the basic structure of the fMSV model, the dynamic correlation MSV model, the conditional/stochastic Wishart autoregressive models, the matrix-exponential MSV model, and the Cholesky MSV model are extended. Empirical results for 7 financial asset returns for US stock returns indicate that the new fMSV models outperform existing dynamic conditional correlation models for forecasting future covariances. Among the new fMSV models, the Cholesky MSV model with long memory and asymmetry shows stable and better forecasting performance for one-day, five-day and ten-day horizons in the periods before, during and after the global financial crisis.

C517: **The Japanese Taylor rule estimated using quantile regressions**
Presenter: **Jauer Chen**, National Taiwan University, Taiwan
Co-authors: Masanori Kashiwagi
Quantile regressions to estimate the Japanese Taylor rule are conducted using a sample that includes recent observations pertaining to Japan's zero interest rate policy. To address censoring and endogeneity, we compute censored quantile instrumental variable estimators. Our estimates indicate that the inflation coefficient tends to decrease moving toward the right tail of the conditional distribution for lower quantiles. This pattern is reversed in uncensored quantile regressions and in the estimation results of the literature using Japanese data prior to the zero interest rate policy, indicating the importance of the information provided by the recent Japanese data and the consideration of censoring.

C804: **Bayesian analysis of max-stable models via hybrid sampling methods**
Presenter: **Ka Shing Chan**, The Hong Kong University of Science and Technology, China
Co-authors: Mike K.P. So
Precipitation risk is possibly one of the most important threats from climate change. No matter it falls as snow or rain, it can greatly increase the flood risk and cause huge social and economic losses. A natural approach to model such extreme spatial events is through max-stable processes, which are extensions of multivariate extreme value theory into stochastic process setting. One limitation of using such an approach is that the joint likelihood of max-stable processes is usually analytically intractable. We develop a hybrid MCMC scheme for Bayesian analysis of several max-stable models, including the Smith, Schlather, Brown-Resnick and Wadsworth-Tawn models. Real data analysis is preformed using the precipitation data in the Greater China region.

CS102 Room M2 CONTRIBUTIONS IN FINANCIAL ECONOMETRICS II Chair: Marco Lippi

C261: **Correlation aggregation in high frequency financial data**
Presenter: **Yang Liu**, University of Amsterdam, Netherlands
Co-authors: Peter Boswijk
A new approach to correlation structure modelling using high frequency measures is proposed. Instead of augmenting the lower frequency structure with classical intraday realized measures such as the realized correlation and realized volatility, we derive several high frequency measures through correlation aggregation in the framework of the generalized autoregressive score model (GAS), which we label the "HF-GAS" model. The high frequency dynamic assumption enables the HF-GAS model to extract more information from intraday data and estimate high and low frequency conditional correlation simultaneously. Our empirical study on high frequency indexes shows the superiority of the HF-GAS model over the daily GAS model and over the daily GAS model augmented with a realized correlation measure, both in-sample and out-of-sample.

C1153: **A conservative test for the lag structure of assets' realized volatility dynamics**
Presenter: **Constantin Roth**, University of Saint Gallen, Switzerland
Co-authors: Francesco Audrino, Lorenzo Camponovo
A conservative test is constructed to investigate the optimal lag structure for forecasting realized volatility dynamics. The testing procedure relies on the theoretical results recently shown. It is based on the ability of the adaptive least absolute shrinkage and selection operator (adaptive lasso) in combining efficient parameter estimation, variable selection, and valid finite sample inference in one step. In an application to several constituents of the S&P500 index we show that (i) the optimal significant lag structure is time-varying and subject to drastic regime shifts that seem to happen across assets simultaneously; (ii) in many cases the relevant information for prediction is included in the first 22 lags, corroborating previous results about the accuracy and the difficulty to outperform out-of-sample the heterogeneous autoregressive (HAR) model; and (iii) some common features of the optimal lag structure can be identified across assets belonging to the same market segment.

C1180: **Forecasting interest rates using geostatistical techniques**
Presenter: **Giuseppe Arbia**, Catholic University of the Sacred Heart, Italy
Co-authors: Michele Di Marcantonio
Geostatistical spatial models are widely used in many applied fields to forecast data observed on continuous three-dimensional surfaces. Applications are traditionally confined to fields like geography, geology, meteorology, agriculture and epidemiology and others. We propose to extend their use to finance and, in particular, to forecasting the term structure of interest rates. We present the results of an empirical application where we apply the proposed method to forecast Euro Zero Rates using the Ordinary Kriging method based on the anisotropic variogram, focusing on the period 2003-2014. The empirical results show that, for long-term maturities of interest rates, the model is characterized by good levels of predictions' accuracy. From a comparative point of view, our model proves to be more accurate than using forward rates implicit in the Euro Zero Rates curve as proxies of the market expectations. Finally, a comparison with other recent methods for forecasting yield curves is proposed. Our work contributes to the existing literature by adopting an innovative approach to analyze the term structure of interest rates for short-term forecasting purposes.

C1191: **The expectation hypothesis of the repo rates: new evidence from multiple hypotheses and heteroskedasticity control**
Presenter: **Vanessa Gunnella**, University of Bologna, Italy
The aim is to empirically test the Expectation Hypothesis (EH) of the term structure of the US repurchasing agreements (repo) rates, considered in a Vector Auto Regression (VAR) model. A multiple hypotheses approach is adopted, in order to jointly test all the statistical hypotheses implied by the EH, i.e. the long-run and short-run implications of the theory. Furthermore, the testing procedures are carried out by taking into account non-stationary volatility of the series and of the error terms through bootstrap inference, white correction and rolling windows analysis. Differently from previous results, overall evidence in favor of the statistical non-rejection of the EH is found. In particular, the rolling window analysis clarifies that the EH has been rejected only during periods of turmoil of the financial/repo markets.

CS104 Room A2 CONTRIBUTIONS TO BAYESIAN ECONOMETRICS I Chair: Roberto Leon

C1279: **Modelling conditional duration via flexible error distributions**
Presenter: **Rasika Yatigammana**, University of Sydney, Australia
Co-authors: Richard Gerlach
Empirical evidence suggests that duration data generally has a unimodal distribution. It has a very long right tail and the majority of transactions have durations close to zero. Frequently, the conditional distribution has been modelled by employing Exponential and Weibull distributions while Log-Normal, Generalised Gamma and Burr have also been considered. However, these distributions have been rejected due to lack of goodness of fit to real data. Ability to capture some of the features, especially the long right tail may have significance for forecasting and trading strategies. The literature has shown that distributions with either constant or strictly increasing or decreasing hazard functions do not capture the characteristics of durations that well. Hence it is proposed to use partitioned and mixture distributions with different parameters for each section/ component and unknown partition points for added flexibility. Combinations of different distributions such as Exponential and Weibull are considered in the context of Autoregressive Conditional Duration (ACD) models. A Bayesian approach is used in the estimation of distribution parameters employing an independent Gaussian proposal within a Markov chain Monte Carlo framework. After a simulation study on multiple series, the estimation is carried out on real data employing Log predictive likelihood ratios for model selection.

C1063: **Detecting time variation in the 'price puzzle': an improved prior choice for the time varying parameter VAR model**
Presenter: **Peter Reusens**, KU Leuven, Belgium
Co-authors: Christophe Croux

Bayesian estimators with different prior choices for the time variation of the coefficients of the Time Varying Parameter Vector Autoregression (TVP VAR) model are compared using Monte Carlo simulation studies. In particular, multiple simulation studies have been performed both for the local level setting and a three variable TVP VAR setting, for data generating processes that are typical for macroeconomic data. These simulation studies show that the frequently used prior choice for the time variation parameter only allows for a tiny amount of time variation in the coefficients and hence results in posterior estimates that largely underestimate the true time variation of the coefficients. Also, two more uninformative priors are proposed for the time variation coefficient for which the posterior estimate has a lower bias and a lower mean squared error. Finally, the estimator with improved prior choices is used to study the time varying response of inflation to an interest rate shock using a three variable TVP VAR model for the USA. Major time variation in the impulse response is found: while a significant 'price puzzle' is present for the period 1972-1979, the response of inflation to an interest rate shock is negative for most other time periods.

C227: **Sparse high-dimension multivariate autoregressive models**
Presenter: **Daniel Felix Ahelegbey**, University of Venice, Italy
Co-authors: Monica Billio, Roberto Casarin

A Bayesian graph-based inference to over-parameterized and high-dimension multivariate autoregressive (MAR) models is proposed. We propose an efficient Markov Chain Monte Carlo algorithm for joint inference of the lag order, the temporal dependence in the observed time series and the parameters of the MAR model. We refer to our model as Bayesian graphical MAR (BGMAR). We study the efficiency of the proposed BGMAR through comparisons with the least absolute shrinkage and selection operator (Lasso). The simulation results and the applications to real data show that the BGMAR produces a better representation of the temporal dependence and a higher predictive accuracy than the Lasso. The empirical applications to high-dimension time series considered in the macroeconomic and financial literature illustrate the effectiveness of BGMAR and its advantages over the Lasso.

C081: **Fully Bayesian inference for smooth threshold autoregressive (STAR)(k)-GARCH(s,q) models**
Presenter: **Darfiana Nur**, Flinders University Australia, Australia
Co-authors: Glen Livingston Jr, Irene Hudson

Regime switching volatility models have been used extensively in finance to explain time-varying volatility and contagion in financial markets. The focus is on the Smooth Threshold Autoregressive (STAR)(k)-GARCH(s,q) model, which is often applied in economics and finance. The aim is to develop a novel and efficient estimation method for this model under a fully Bayesian paradigm which includes the model orders (k,s,q). The joint posterior distributions will be shown, followed by the design of a posterior simulator using a combination of MCMC algorithms. Finally, simulation studies and a case study will be presented for the proposed MCMC posterior simulator.

CS106 Room F2 COMPUTATIONAL ECONOMETRICS I Chair: Christopher M. Otrok

C1202: **Estimating the time-varying parameters model: an alternative approach**
Presenter: **Stella Hadjiantoni**, Queen Mary University of London, United Kingdom
Co-authors: Erricos Kontoghiorghes

It is often possible that the regression coefficients are gradually changing. In such cases, the assumption for constant parameters is not valid. Herein, the ordinary linear regression model with time dependent coefficients is considered. A new method to estimate the time-varying parameters (TVP) model within the context of generalized least squares is investigated. The filtering and smoothing estimators are derived. The problem of adding and deleting observations from the TVP model is also examined within the context of rolling window estimation. The proposed method develops an equivalent model where previous computations are efficiently utilized. When removing observations from the model, the imaginary unit is used to delete the effect of the desired observations. This results in a non-positive definite dispersion matrix which comprises complex covariance values. The estimator of this model is obtained by solving a generalized linear least squares problem. An efficient algorithm is proposed to compute the estimator of the model. The algorithm avoids repeated computation of matrix inverses, as in the Kalman Filter method, which can be the reason for inaccurate results. The estimation method is also extended to estimate a system of related regressions with time-varying coefficients.

C1212: **A dynamic model of limit order market with possibly rational traders**
Presenter: **Martin Smid**, Institute of Information Theory and Automation, Czech Republic

N agents trading at a continuous-time limit order market are considered. In accordance with reality, the agents' actions do not take effect immediatelly but after a short random latency time. Each agent may be uncertain about the information of the other agents. We show that an optimal strategy of an informed rational agent who maximizes discounted consumption over an infinite time horizon while keeping the probability of bankruptcy (in a traded instrument or in money) at a prescribed level depends only on the current shape of the order book, the agent's inventory of the traded asset and the past information the others base their strategies on. If such agent is not informed, his strategy depends, in addition, on his uncertainty. Thus, any form of dependence on the past may persist on equilibrium at the market. We also present an empirical study supporting our findings given a simpler setting, in which the agent may post only a single limit order.

C1282: **A numerical estimation method for discrete choice models with non-linear externalities**
Presenter: **Fabio Vanni**, Scuola Superiore Sant Anna, Italy
Co-authors: Giulio Bottazzi, Ugo Gragnolati

The purpose is to present a stochastic discrete choice model with non-linear externalities and a related methodology to carry out inferential analysis on its parameters. Such framework allows us to disentangle the role of the intrinsic features of alternatives from the effect of externalities in determining individual choices. The stochastic process underlying the model is demonstrated to be ergodic, so that numerical methods are adopted to estimate the parameters by chi square minimization and evaluate their statistical significance. In particular, optimization rests on successive parabolic interpolation. Finally, the model and the related inferential analysis are applied to the case of firm localization using Italian sectoral census data.

C1230: **Reduction and composite likelihood estimation of non-scalar multivariate volatility models**
Presenter: **Lyudmila Grigoryeva**, Universite de Franche-Comte, France
Co-authors: Luc Bauwens, Juan-Pablo Ortega

Multivariate volatility models are widely used for the description of the dynamics of time-varying asset correlations and covariances. Among the well-known drawbacks of many of these parametric families one can name the so called curse of dimensionality and the necessary nonlinear parameter constraints that are difficult and sometimes unfeasible to handle. These features make the estimation of such models a nontrivial task and call for new methods and techniques to successfully tackle this problem in practice. The discussed model reduction approach is used to efficiently apply the composite likelihood estimation method for a newly introduced DVEC (Diagonal VEC) family and DCC (Dynamic Conditional Correlation) models with different types of intrinsic parametrization that are shown to be closed under the proposed reduction procedure. The linear and nonlinear parameter constraints problem at the time of estimation is treated using a Bregman-proximal trust-region optimization algorithm. Additionally, an empirical out-of-sample performance comparison of the models under study estimated using both the standard likelihood and the

composite likelihood is provided.

CS65 Room G2 REGIME CHANGE MODELING IN ECONOMICS AND FINANCE III Chair: Willi Semmler

C929: **Financial investment constraints. A panel threshold application to German firm level data**

Presenter: **Artur Tarassow**, Universitaet Hamburg, Germany

The attempt is to test whether financial supply-side shifts explain the low-investment climate of private firms in Germany. The core contention is that a firm's financial position contributes to its access to external finance on credit markets. Special emphasis is put on small and medium-sized enterprises as these are usually assumed to be more informationally opaque. The application of a non-linear panel threshold model makes it possible to group firms endogenously according to their financial position. Various observable balance sheet indicators such as leverage, interest coverage ratio or measures of solvency are used as potential threshold variables. The firm-level panel dataset covers the period between 2006 and 2012. We find strong evidence for a positive but non-linear nexus between cash flow and fixed investments, suggesting that financially fragile firms rely more heavily on internal funds. Surprisingly, firm size does not seem to be a relevant grouping variable.

C948: **Do central bank forecasts matter for professional forecasters?**

Presenter: **Jacek Kotlowski**, National Bank of Poland, Poland

The aim is to establish to what extent public information provided by the central bank affects the forecasts formulated by professional forecasters. We investigated empirically whether disclosing GDP and inflation forecasts by the National Bank of Poland in the period 2004-2013 could have lowered the cross-sectional dispersion of professional forecasters expectations. Our main finding is that by publishing projection of future GDP growth the central bank was influencing the median and reducing the dispersion of one-year-ahead GDP forecasts. Moreover, the estimation results indicate that the role of central bank forecasts as a focal point for the private agents strengthened after the onset of the global financial crisis. We also identified some determinants of the dispersion among the forecasters. In general, the cross-sectional dispersion of GDP forecasts is positively influenced by the volatility of industrial production and it increases around turning points. Moreover, the dispersion is higher during slowdowns than in recovery phases. Finally, the disagreement among forecasters grows with surprises when GDP and industrial production releases differ from the forecasts. On the other hand the release of CPI projection by the central bank affects neither the cross-sectional dispersion nor a level of forecasts formulated by professional forecasters.

C949: **What common factors are driving inflation in CEE countries?**

Presenter: **Aleksandra Halka**, Narodowy Bank Polski, Poland

Co-authors: Grzegorz Szafranski

The sources of time variation in consumer inflation is analyzed across ten Central and Eastern European (CEE) countries and five sectors (durables, semi-durables, non-durables, food, and services) in the period 2001-2013. With a multi-level factor model we decompose product-level HICP inflation rates into CEE-region wide, sector, country, country-sector specific and idiosyncratic components. The outcomes indicate that region-wide and country specific components of inflation are very persistent in contrast to sector and product-level components, which is in line with similar studies for core EU countries. Moreover, two region-wide factors explain about 17% of variance in monthly price changes, whereas the other common components explain below 10% each. The results are at odds with empirical evidence on the importance of sectoral price shocks in developed economies. This difference may be related to the conclusion that the first region-wide factor is associated with common disinflationary processes that occurred in CEE economies in the 2000s, whereas the second one reveals significant correlations with global factors, especially commodity prices and euro area price developments

C917: **A likelihood-based approach to estimate Markov-switching stochastic volatility models**

Presenter: **Frederic Karame**, University of Maine, France

It focuses on Markov-switching stochastic volatility models (with leverage). The presence of a second unobserved variable capturing the volatility regimes induces path-dependence that considerably complicates the model estimation. While the literature has developed many algorithms to deal with this issue, we propose a simple particle filtering approach to estimate this model using standard maximum likelihood inference. Our contribution relies on two approximations that use nearly the same idea for general linear state-space representations with Markov switching. The general structure of our method is based on a Hamilton-like filter to track unobserved regimes. As the conditional model remains nonlinear, we use particle filtering to approximate the unobserved log-volatility and provide an unbiased estimation of the conditional likelihood necessary for updating regime probabilities and estimating model parameters. We use the smooth resampling approach previously developed for three reasons: (i) to update particles; (ii) implement the recombination of the tree possibilities or collapsing and consequently circumvent the path dependence issue; (iii) implement likelihood-based inference on unknown parameters. After the description of our approach and some Monte Carlo experiments, we present the estimation results on the IBOVESPA index for comparison with a previous work.

Monday 8.12.2014 08:45 - 10:05 Parallel Session O – ERCIM

ES145 Room O1 DIMENSIONALITY REDUCTION AND VARIABLE SELECTION Chair: Hannu Oja

E1046: **Supervised dimension reduction based on moments**
Presenter: **Joni Virta**, University of Turku, Finland
Co-authors: Klaus Nordhausen, Hannu Oja

In supervised dimension reduction the aim is to reduce the dimension of the vector of explanatory variables $\mathbf{x}$ while still preserving all the information and dependency it has on a response y. Standard methods used for this purpose include sliced inverse regression (SIR) and sliced average variance estimation (SAVE), which are based on the diagonalization of scatter matrices of first and second moments, respectively. We now extend this approach twofold; first, to use also third and fourth moments, and second, to use several of the individual moments at once. By diagonalizing several moment-based scatter matrices at the same time the method allows the finding of multiple kinds of dependencies simultaneously. This is demonstrated in simulations where the superiority of utilizing several moments simultaneously over using individual moments is seen.

E1086: **Seeking relevant information from a statistical model**
Presenter: **Marcela Svarc**, Universidad de San Andres, Argentina
Co-authors: Ricardo Fraiman, Yanina Gimenez

A general procedure for selecting variables is introduced, which can be applied to several classical multivariate problems, including principal components, regression, classification, clustering, among others. The aim is to allow the identification of a small subset of the original variables that can "better explain" the model concerned through nonparametric relationships. The method typically yields some "noisy" uninformative variables and some variables that are strongly related because of their general dependence and our aim is to help understand the underlying structures in a given dataset. The asymptotic behaviour of the proposed method is considered and some real and simulated data sets are analyzed as examples.

E1299: **Dimension-reduced modeling of spatio-temporal processes**
Presenter: **Jenny Brynjarsdottir**, Case Western Reserve University, United States
Co-authors: Mark Berliner

The field of spatial and spatio-temporal statistics is increasingly faced with the challenge of very large datasets. The classical approach to spatial and spatio-temporal modeling is very computationally demanding when datasets are large, which has led to interest in methods that use dimension-reduction techniques. We focus on the modeling of two spatio-temporal processes where the primary goal is to predict one process from the other and where datasets for both processes are large. We outline a general dimension-reduced Bayesian hierarchical modeling approach where spatial structures of both processes are modeled in terms of a low number of basis vectors, hence reducing the spatial dimension of the problem. Temporal evolution of the processes and their dependence is then modeled through the coefficients of the basis vectors. We present a new method of obtaining data-dependent basis vectors that is geared towards the goal of predicting one process from the other. We apply these methods to a statistical downscaling example, where surface temperatures on a coarse grid over Antarctica are downscaled onto a finer grid.

E952: **Wind power forecast combination through dimension reduction techniques**
Presenter: **Marta Poncela Blanco**, Joint Research Center, Italy
Co-authors: Pilar Poncela Blanco

Forecast errors of wind power have a direct impact in the operational decision problems involved in the integration of this non dispatchable form of energy. As the relationship between wind and the generated power is highly nonlinear and time-varying, and given the available variety of models, it is very common to use alternative models based on different assumptions about the variables involved or based on different types of models for getting more than one prediction for the same time and forecast horizon. The objective is obtaining a single final prediction to increase forecasting accuracy. The traditional way for calculating the weighting factors of the combination is based on the evaluation of past errors of each of the original competing forecasts. However, our approach is to use multivariate reduction techniques for dealing with the forecast combination. This way, the weighting factor is not based on the forecast error. Instead, each weighting factor is based on the amount of information that each predictor, regarding the other forecasters, adds to the final combination.

ES148 Room M1 DISCRETE STATISTICS Chair: Wicher Bergsma

E909: **A multivariatie model for multinomial choices**
Presenter: **Koen Bel**, Erasmus University Rotterdam, Netherlands
Co-authors: Richard Paap

Multinomial choices of individuals are likely to be correlated. Nonetheless, econometric models for this phenomenon are scarce. A problem of multivariate multinomial choice models is that the number of potential outcomes can become very large which makes parameter interpretation and inference difficult. We propose a novel Multivariate Multinomial Logit specification where (i) the number of parameters stays limited; (ii) there is clear interpretation of parameters in terms of odds ratios; (iii) zero restrictions on parameters result in independence between the multinomial choices and; (iv) parameter inference is feasible using a composite likelihood approach even if the multivariate dimension is large.

E1104: **Some step-up procedures that control the false discovery rate for discrete tests**
Presenter: **Sebastian Doehler**, Darmstadt University of Applied Science, Germany

In many multiple testing problems control of the false discovery rate is desired. We develop some step-up procedures that control the false discovery rate under arbitrary dependence of the p-values. When the test statistics are discrete, these new procedures can provide considerable improvements over classical procedures like the Benjamini-Yekutieli procedure by incorporating discreteness in a straightforward way. We illustrate these new procedures with an application to pharmacovigilance data.

E1112: **An SGoF-type multiple test method for discrete data**
Presenter: **Irene Castro-Conde**, University of Vigo, Spain
Co-authors: Sebastian Doehler, Jacobo de Una-Alvarez

The recently proposed Sequential Goodness-of-fit-test (SGoF) procedure is an alternative to the FWER- and the FDR- strong controlling methods in high-dimensional multiple testing problems, and especially when weak effects are present. However, for discrete data this procedure may be very conservative due to the wrong assumption that the p-values follow the uniform distribution under the complete null hypothesis. We introduce a modified SGoF-type procedure called Discrete SGoF that takes into account the discreteness of the tests by assuming that the p-values follow a distribution stochastically larger than the uniform. We prove some theoretical results of the Discrete SGoF procedure regarding the weak control of the FDR and we study its performance in a simulation study. We see that Discrete SGoF is consistently closer to the nominal level than the original SGoF procedure and exhibits a higher power. In addition, Discrete SGoF maintains the conservativeness property of the SGoF-type methods and the property of having an increasing power when the number of tests increases. We also compare the Discrete SGoF procedure with the Benjamini and Hochberg FDR-controlling method. Finally, we illustrate the application of the Discrete SGoF procedure to a real data set.

E1249: **A technique to simultaneously monitor bivariate random variables defined on contingency tables**
Presenter: **Athanasios Sachlas**, University of Piraeus, Greece

Co-authors: Sotiris Bersimis

Multivariate statistical process control (MSPC) is nowadays used in many non-industrial fields. Data from such processes usually require the development of special monitoring procedures. We propose a technique in order to monitor bivariate random variables defined on contingency tables. Specifically, we propose a technique for monitoring simultaneously the agreement of two raters, measured by Cohen's kappa, and one percentage defined on the same contingency table. The method is based on an appropriate approximation and it is assessed numerically. We also present an application of the method on real data from a retailer company.

ES57 Room H1 RECENT APPLICATIONS OF GRAPHICAL MARKOV MODELS Chair: Nanny Wermuth

E099: **Long term effects of emotional parentification**

Presenter: **Jochen Hardt**, University of Mainz, Germany

Parentification means that a child takes care of a parent - instead of vice versa as it should be. We expected that parentified children would develop deficits in the perception of their own needs in order to survive emotionally in the family - i.e. they should be prone to somatization. The present analysis is based on two internet surveys of about 500 subjects each. Via e-mail, respondents were asked to fill out a set of questionnaires. Two indicators of possible somatization were chosen as primary response, vegetative symptoms and pain. A third primary response, depression, served as a control variable. Sequences of regressions were performed to analyse the data. Contrary to our expectations, depressive symptoms were highly associated to maternal and paternal parentification. Pain was associated with maternal parentification, vegetative symptoms only with paternal parentification. The hypothesis of a deficit in self perception had to be rejected. The associations of parentification to depression are equally strong as those to vegetative symptoms and pain.

E306: **Discrete graphical models in social mobility research**

Presenter: **Renata Nemeth**, Eotvos Lorand University of Budapest, Hungary

Co-authors: Tamas Rudas

Graphical models are well suited to model direct and indirect associations that are of central importance in many problems of sociology. Such relevance is apparent in research on social mobility. Graphical models are discussed in the framework of marginal models, which provides a basis for a unified view. The marginal modeling framework relies on parameters that capture aspects of associations among the variables that are relevant for the graph and, depending on the substantive problem at hand, may lead to a deeper insight than other approaches. A general version of path models for categorical data is also introduced. These models are applied to the social status attainment process in the USA, Hungary and Czechoslovakia at the end of the last century, and shows that policies in the latter socialist countries to prevent status inheritance had little success.

E575: **Triangular symmetric elements of the entropy in graphical modelling**

Presenter: **Joe Whittaker**, Lancaster University, United Kingdom

A graphical model is defined by its joint distribution. We define the symmetric elements of the entropy for an arbitrary distribution and concentrate on the third order, the triangular elements. These are found to have interesting properties for the interpretation of the graph and lead to modest efficiency gains in graphical model search. Examples are given appropriate to a joint Gaussian distribution.

E739: **Identification of principal causal effects using additional outcomes in concentration graphs**

Presenter: **Fabrizia Mealli**, University of Florence, Italy

Co-authors: Barbara Pacini, Elena Stanghellini

Unless strong assumptions are made, nonparametric identification of principal causal effects in causal studies can only be partial and bounds (or sets) for the causal effects are established. In the presence of a secondary outcome, recent results exist to sharpen the bounds that exploit conditional independence assumptions. More general results, though not embedded in a causal framework, can be found on concentration graphs with a latent variable. The aim is to establish a link between the two settings and to show that adapting and extending results pertaining to concentration graphs can help achieving identification of principal casual effects in studies when more than one additional outcome is available. Model selection criteria are also suggested. An empirical illustrative example is provided, using data from a real social job training experiment.

ES151 Room N1 DEPENDENCE IN APPLIED STATISTICS AND DATA ANALYSIS Chair: Concepcion Ausin

E840: **Default probability estimation via pair copula constructions**

Presenter: **Claudia Tarantola**, University of Pavia, Italy

Co-authors: Luciana Dalla Valle, Maria Elena De Giuli, Claudio Manelli

Following the growing financial uncertainty, there has been intensive research by financial institutions, regulators and academics to develop models for firm evaluation and default probability estimation. The existing methodologies differ on the available information and data used for assessing the firm value. They can be broadly classified in models based on market data and on accounting data. We propose a novel semi-Bayesian approach for probability of default estimation, that combines features of the previous two classes of models. Our approach is based on multivariate contingent claim analysis and pair copula constructions. For each considered firm, balance sheet data are used to assess the asset value, and to compute its default probability. The asset pricing function is expressed via a pair copula construction, and it is approximated via Monte Carlo simulations. We apply our methodology to the analysis of both defaulted and operative firms. We consider four well-known fraudulent bankruptcy cases (Cirio, Enron, Parmalat and Swissair). Furtehrmore, to test the efficacy of our methodology we examine also a non defaulted firm (Sysco Company), operating in the same period of the defaulted ones.

E1029: **Modeling dependencies in clustered binomial data**

Presenter: **Andreas Dartsch**, University of Rostock, Germany

Co-authors: Rafael Weissbach

Binary clustered data frequently occur in different fields of research such as financial econometrics, epidemiology or medical statistics. They are often analyzed as cluster count variables following unbalanced extended binomial distributions with intraclass correlation (ICC) being a parameter (Common Correlation Model) or being determined by choosing α and β in a Beta-Binomial Model. The results of studying and estimating ICC are displayed. Interestingly, ICC is identified and therefore, asymptotically, the likelihood function has a unique maximum at the true value of ICC. Nevertheless, for finite samples, the maximum likelihood estimator does not need to exist. Especially, for large clusters the likelihood is easily monotonous even for higher numbers of clusters. This implies that the maximum likelihood estimator will be on the edge of the feasible region. Emerging from this unsatisfying situation many different estimators for ICC were published in the past. The aim is to point out a few of them and compare their performance within two non-nested models by employing these methodologies to epidemiological and financial data as well as simulated data. In particular, a moment based direct estimator for both balanced and unbalanced designs is developed by inverting a generalized chi-squared statistic.

E1175: **Copula-based models for the analysis of glacier discharge at King George Island, Antarctica**

Presenter: **Concepcion Ausin**, Universidad Carlos III de Madrid, Spain

Co-authors: Mario Gomez, Carmen Dominguez

As consequence of Global Warming, with the gradual increasing of environmental temperature, the ice cap glaciers lose liquid mass by melting at the base. This is known as glacier discharge. We are interested in the construction of adequate models to describe the distribution of the glacier

discharge as a function of other meteorological variables such as the air temperature and humidity, among others. In particular, we concentrate on one of the measuring stations of glacier discharge within the GLACKMA project (www.glackma.es) located in the Collins ice cap in the King George Island in the Antarctica. Considering that the relationship between the meteorological variables is clearly not linear, we propose the use of copulas which allow for the construction of flexible multivariate distributions with non linear dependencies. In particular, we provide estimations of the conditional densities of the glacier discharge given different values for the temperature and humidity for a subset of the data collected in the Collins icecap corresponding to three months of an Austral Winter. This procedure seems to be more informative than classical regression approaches.

E1208: **Sums of dependent random variables and portfolio selection via copula modeling**
Presenter: **Klaus Herrmann**, KU Leuven, Belgium
Co-authors: Irene Gijbels

The impact of varying stochastic dependence structures when applying a summation operation to components of a random vector is studied. Therefore a random variable $Z = \sum_{i=1}^{d} \omega_i X_i$ derived from a random vector $\mathbf{X} = (X_1, \ldots, X_d)$, where $(\omega_i)_{i=1}^{d}$ are given weights, is considered. While for special cases, such as independence or joint normality of $\mathbf{X}$, properties of Z are well known, these are now studied more generally using arbitrary absolutely continuous copulas to model the dependence structure on $\mathbf{X}$. Formulas and efficient numerical algorithms to compute the density, distribution and quantile functions of Z are presented. These approaches are connected to integration over a surface in the unit square $[0,1]^d$ only determined by the margins of $\mathbf{X}$. In this new formulation the influence of the margins is clearly separated from the influence of the dependence structure, a key feature of the copula approach. While this approach is useful to generalize bivariate stress-strength or copula-GARCH models, the main attention is devoted to portfolio selection. Therefore the choice of weights $(\omega_i^*)_{i=1}^{d}$ that minimize the expected shortfall $\mathbb{E}[Z|Z \leq F_Z^{-1}(\beta)]$, for $\beta \in (0,1)$, is discussed.

ES68 Room F1 MARKOV PROCESSES Chair: Sebastian Engelke

E957: **Maximum likelihood estimation in the two-state Markovian arrival process**
Presenter: **Pepa Ramirez Cobo**, Universidad de Cadiz, Spain
Co-authors: Emilio Carrizosa

The Markovian arrival process (MAP) has proven to be a versatile model for fitting dependent and non-exponential interarrival times, with a number of applications to queueing, teletraffic, reliability or finance. The aim is to examine the maximum likelihood approach for inference for the two-state Markovian arrival process. Despite the apparent simplicity of the problem, some technical difficulties are found concerning the evaluation of the objective function. Such difficulties are overcome by the use of the canonical representation of the process, a proper re-scaling of the function and the choice of a particular starting point of the algorithm. Finally, an approach to compare different estimates of the process based on the Kullback-Leibler divergence is delineated.

E162: **Estimation of the transition probabilities in the illness-death model**
Presenter: **Luis Machado**, University of Minho, Portugal
Co-authors: Artur Araujo, Javier Roca-Pardinas

The major goal in clinical applications of multi-state models is the estimation of transition probabilities. The usual nonparametric estimator of the transition matrix for nonhomogeneous Markov processes is the Aalen-Johansen estimator. However, two problems may arise from using this estimator: first, its standard error may be large in heavy censored scenarios; second, the estimator may be inconsistent if the process is non-Markov. Happily, there have been several recent contributions that account for these problems. We consider the estimation of the transition probabilities, using TPmsm a software application for R. It describes the capabilities of the program for estimating these quantities using seven different approaches. In two of these approaches the transition probabilities are estimated conditionally on current or past covariate measures. The software is illustrated using data from real data sets.

E1198: **Modelling unobserved heterogeneity in quantile regression for longitudinal data**
Presenter: **Maria Francesca Marino**, Sapienza University of Rome, Italy
Co-authors: Marco Alfo', Nikos Tzavidis

Quantile regression has become a standard tool in the analysis of longitudinal data as it offers a thorough overview on the conditional distribution of a response variable given a set of covariates. When dealing with longitudinal studies, observations coming from the same individual are naturally dependent because of the presence of unobserved sources of heterogeneity. If such a dependence is not properly taken into consideration in the data analysis, misleading inferential conclusions can be easily drawn. A mixed hidden Markov quantile model for continuous longitudinal data is proposed. Time-constant and time-varying random parameters are considered in the model specification to jointly account for time-invariant and dynamic unobserved factors affecting the response variable distribution. The resulting model offers great flexibility being a generalization of the basic linear mixed quantile regression model and the standard hmm for quantiles frequently used in the quantile regression framework for longitudinal data. In order to face the numerical integration problem typically arising in the mixed hmm context, a non parametric maximum likelihood approach is applied. Parameter estimates are obtained via an EM algorithm and their computation is greatly simplified by exploiting the the forward and backward variables defined in the so-called Baum-Welch algorithm.

E1054: **Monitoring side effects in phase II comparative clinical trials**
Presenter: **Sotiris Bersimis**, University of Piraeus, Greece
Co-authors: Athanasios Sachlas

The case of adverse events in Phase II comparative trials has been given only a small amount of attention. In a previous work we proposed a design where the occurrence of an adverse event leads to the termination of the clinical trial. This rule is too strict and can be used only in case of severe adverse events. Using the Markov chain embedding technique, we study the case that we monitor the development of side effects considering that the side effect does not immediately stop the clinical trial but penalizes the treatment that caused it. Tables of expected sample size and power are presented. The numerical illustration showed a very good performance for the new design. The penalization of the treatment terminates the clinical trial early enough (after a few pairs and thus involving a small number of patients) with high power.

ES155 Room P1 MULTIVARIATE MODELLING Chair: Apostolos Batsidis

E911: **Distribution of the product of singular Wishart matrix and normal vector**
Presenter: **Stepan Mazur**, Lund University, Sweden
Co-authors: Taras Bodnar, Yarema Okhrin

A very useful formula is derived for the stochastic representation of the product of a singular Wishart matrix with a normal vector. Using this result, the expressions of the density function as well as of the characteristic function are established. Moreover, the derived stochastic representation is used in generating the elements of the product which leads to a considerable improvement in the computation efficiency. Finally, we present several important properties of the singular Wishart distribution, like its characteristic function and distributional properties of the partitioned singular Wishart matrix.

E1248: **Taking advantage of a symbolic representation of non-central Wishart distributions**
Presenter: **Elvira Di Nardo**, Basilicata, Italy

The computation of joint moments $E\left\{\mathrm{Tr}[W(n)H_1]^{i_1}\cdots\mathrm{Tr}[W(n)H_m]^{i_m}\right\}$, with $H_1,\ldots,H_m$ complex matrices, Tr the trace and $W(n)$ non-central Wishart distribution has different applications: in the study of asymptotic properties of the sample covariance matrix or in quantifying the performance of multidimensional signal processing algorithms. Indeed for $H_1,\ldots,H_m$ sparse matrices, joint moments of entries of $W(n)$ can be recovered. This computation is a very general task and object of in-depth analysis, due to the complexity of the cumbersome involved formulae. Usually symbolic languages are required involving derivatives of vector of matrix functions. A different symbolic method is introduced allowing to take advantage of the cyclic properties of traces. Thanks to this device, $\mathrm{Tr}[W(n)]$ is represented as convolution of its central component and a matrix of formal variables, whose entries are uncorrelated with those of the central component. Due to the cyclic properties of traces, the notion of necklace is fruitfully employed in the computation allowing to set up an efficient symbolic procedure. The algorithm has been implemented in `Maple 12`. Comparisons with other techniques proposed for computing moments of $W(n)$ are given.

E932: **Semi-parametric approaches for modelling of time-to-event outcomes**
Presenter: **Wirda Nisar**, University of Liverpool, United Kingdom
Co-authors: Catrin Tudur-Smith, Ruwanthi Kolamunnage-Dona

Cox proportional hazard models (Cox-PH) are widely used in survival analysis and play a significant role in modelling time-to-event outcomes to assess the significance of the risk or prognostic factors through hazard function. Cox-PH can be easily generalised to include random effects, where random effects act multiplicatively on the risk function. This model can be extended to more general framework to include nested or multilevel effects. In multilevel models the observations are treated as level-1 unit and the subjects as level-2 units. We explore and compare the performance of several semi-parametric approaches to estimate the model parameters under the multilevel random effect model framework for time-to-event data. The adequacies of the methods are investigated through simulation studies. We apply the methodologies on 5 randomised control trials investigating the use of two anti-epileptic drugs: Carbamazepine (CBZ) and Sodium Valproate (SV) to estimate the treatment effect and the variation of treatment effect across the 5 trials. We compare the results from both application and simulation methods to provide a deeper insight to the methodology.

E1111: **Linear transformations to symmetry**
Presenter: **Nicola Loperfido**, University of Urbino, Italy
Co-authors: Cinzia Franceschini

Random vectors with null third-order cumulants are obtained by projecting the data onto appropriate subspaces. Statistical applications include, but are not limited to, the robustification of Hotelling test against nonnormality. Our approach only requires the existence of the third-order moments and leads to normal transformed variables when the parent distribution belongs to well-known classes of sample selection models.

ES84 Room Q1 BAYESIAN METHODS I Chair: Pierpaolo De Blasi

E903: **Hierarchical Bayesian LASSO for a negative binomial regression**
Presenter: **Shuai Fu**, SUPSI, Switzerland
Co-authors: Giorgio Corani

Numerous researches have been carried out to explain the relationship between the count data y and numbers of covariates x through a generalized linear model (GLM). A hierarchical Bayesian LASSO solution is proposed using six different prior models to the negative binomial regression. Latent variables Z have been introduced to simplify the GLM to a standard linear regression model. The proposed models regard two conjugate zero-mean Normal priors for the regression parameters and three independent priors for the variance: the Exponential, Inverse-Gamma and Scaled Inverse-χ^2 distributions. Different types of priors result in different amounts of shrinkage. A Metropolis-Hastings-within-Gibbs algorithm is used to compute the posterior distribution of the parameters of interest through a data augmentation process. Based on the posterior samples and an original Double Likelihood Ratio Test (DLRT) statistic, an automatic variable selection has been performed to choose the most relevant covariates and shrink the insignificant coefficients to zero. Numerical experiments on a real-life data set show that Bayesian LASSO methods achieved significantly better predictive accuracy and robustness than the classical maximum likelihood estimation and the standard Bayesian inference.

E944: **The horseshoe estimator: posterior concentration around nearly black vectors**
Presenter: **Stephanie van der Pas**, Leiden University, Netherlands
Co-authors: Bas Kleijn, Aad van der Vaart

A variety of estimators is available for estimating the underlying mean in the multivariate normal mean model in the situation that the mean vector is sparse in the nearly black sense. The posterior mean derived from the horseshoe prior, a continuous shrinkage prior, has shown promising results in simulation studies. It is of interest because computational difficulties associated with the well-studied 'spike-and-slab' priors can be avoided. Theoretical results, within the frequentist framework where the data is generated according to a fixed mean vector, are presented. The first main result, concerning the horseshoe estimator as a point estimator, is that if the number of nonzero parameters of the mean vector is known, the horseshoe estimator attains the minimax ℓ_2 risk, possibly up to a multiplicative constant. Furthermore, under some conditions, the horseshoe estimator combined with an empirical Bayes estimate of the number of nonzero means still yields the minimax risk. The second main result, of interest for uncertainty quantifcation, is that the posterior distribution contracts around the true mean vector at the minimax rate. Bounds on the posterior variance indicate that the posterior distribution of the horseshoe prior may be more informative than that of other one-component priors, including the Lasso.

E1080: **Bayesian flux balance analysis for a complex model of brain cellular metabolism**
Presenter: **Daniela Calvetti**, Case Western Reserve University, United States
Co-authors: Erkki Somersalo

To understand the interplay between carbon balance and ammonium balance in the energetics of brain metabolism, we perform a flux balance analysis with a large, detailed model of a portion of the brain using a probabilistic framework and presenting the solution in the form of sample generated via MCMC. This framework, which is well suited for the utilization of non-quantitative and imperfect data, is naturally suited for extracting functional relations between the variables.

E1128: **Bayesian analysis of randomized response data**
Presenter: **Amanda Chu**, City University of Hong Kong, China
Co-authors: Ray Chung, Mike So

In business research and social sciences surveys, it is not uncommon to collect responses to sensitive questions. Randomized response techniques have been shown to be a useful statistical method to protect respondents' privacy while being able to elicit truthful answers to sensitive questions. However, when the number of survey questions is not small, analyzing relationship among different responses becomes nontrivial. A Bayesian approach is proposed aiming at performing multivariate analysis of multiple responses where some of them are sensitive. To deal with high-dimensionality issues when the number of questions is large, Bayesian shrinkage idea is adopted. Our method is illustrated by simulation and a business survey on information security.

ES131 Room E1 METHODOLOGICAL STATISTICS I Chair: Jean-Marc Bardet

E899: **Measures of asymmetry based on a necessary/necessary and sufficient conditions for symmetry**
Presenter: **Dimitrios Bagkavos**, Accenture, Greece
Co-authors: Prakash Patil, Andy Wood

It is common practice to make assertions about the symmetry or asymmetry of a probability density function based on coefficients of skewness. Since most coefficients of skewness are designed to be zero for a symmetric density, they do, overall, provide an indication of symmetry. However, skewness, as opposed to asymmetry, is primarily influenced by the tail behavior of a density function. Therefore, coefficients of skewness do not reliably calibrate asymmetry in the density curve. Two measures of symmetry are presented based on necessary/necessary and sufficient condition for a continuous probability density function to be symmetric. They result to coefficients of asymmetry, for a continuous probability density function on the scale of -1 to 1. We show through examples that the proposed measures do an admirable job of capturing the visual impression of asymmetry of a continuous density function. Further, we discuss how the same methodology can be used to develop a hypothesis test for testing the symmetry of a function.

E1055: **Bounds for the probability of causation: a new formulation in Mediation Analysis**
Presenter: **Rossella Murtas**, University of Cagliari, Italy
Co-authors: Philip Dawid, Monica Musio

Given even the best evidence for the dependence of an outcome variable on an exposure variable, we can typically only provide bounds for the "probability of causation" in the case of an individual who has developed the disease after being exposed. We show how these bounds can be improved or adapted if further information becomes available. In addition to the actual literature we provide a new formulation in the case of Mediation Analysis. Mediation studies the relationship between two variables when a third is involved in the causal pathway. In particular, we will show how this mediated effect can be bounded in the case of no direct effect.

E961: **A univariate multiple use calibration**
Presenter: **Martina Chvostekova**, Palacky University Olomouc, Czech Republic

Multiple use calibration problem consists of using a single fitted regression line to construct the interval estimates for future unobserved values of an explanatory variable corresponding to a sequence of future observations of a response variable. In the case of an unknown or an unlimited number of observations, the numerical results indicate that the multiple use of confidence intervals can be constructed by inverting the tolerance intervals. Marginal property of these multiple use confidence intervals is that at least γ proportion of them will contain the corresponding true value of an explanatory variable with confidence $1-\alpha$. We consider a fixed form of a band around the fitted regression line and the multiple use confidence intervals are determinated by inverting the suggested band. In the case of a simple linear regression model an explicit formula for computing the multiple use confidence intervals is proposed. The suggested multiple use confidence intervals satisfy the content and the confidence level requirement, the computation is fast and simple, an average width is lower than those constructed by using the tolerance intervals in a majority of the considered cases in a provided numerical comparison.

E1137: **Two-mode three-way asymmetric MDS using the log linear model**
Presenter: **Jun Tsuchida**, Doshisha University, Japan
Co-authors: Hiroshi Yadohisa

Two-mode three-way proximity data (object vs. object vs. condition or source) is observed in many fields, including marketing research, social psychology, and so on. The data consist of two or more square matrices of proximities between objects. Such data usually contain asymmetric structures, in which a proximity from object i to object j does not equal a proximity from object j to object i. Two-mode three-way asymmetric multidimensional scaling (MDS) is one method for analyzing such data. We propose a two-mode three-way asymmetric MDS technique that uses the log linear model considering self-similarity. Maximum-likelihood estimation can estimate the parameters of asymmetry using the log linear model. Therefore, with some regularity conditions, we can use information criteria such as the Akaike information criterion to select the effects of asymmetry. Moreover, the model can be extended for an n-mode m-way model.

ES139 Room A1 CONTRIBUTIONS TO ROBUST DATA ANALYSIS Chair: Jana Jureckova

E558: **Robust estimation in high-dimensional datasets**
Presenter: **Pietro Coretto**, Universita degli Studi di Salerno, Italy

Multivariate estimators of location/scale parameters 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 discuss and compare new contributions that jointly deal with robustness and high-dimensionality. Emphasis is given both on theoretical properties, as well as empirical performance. We also show how these new tools can improve classical methods for clustering and classification.

E1268: **Making sense of multivariate data that isn't normal**
Presenter: **Arne Bathke**, University of Salzburg, Austria
Co-authors: Solomon Harrar, Amanda Ellis, Markus Pauly, Frank Konietschke, Woodrow Burchett

Parametric and nonparametric methods for inference between multivariate data samples are presented. Application is demonstrated using the R package npmv. Unlike in classical MANOVA, multivariate normality is not required for the data. Different response variables may even be measured on different scales (binary, ordinal, quantitative). In addition to permutation tests and F approximations for the overall hypothesis test, we present a multiple testing procedure which identifies significant subsets of response variables and factor levels, while controlling the familywise error rate.

E1051: **Robust regression with constraints under heteroscedasticity and broken orthogonality condition**
Presenter: **Jan Amos Visek**, Charles University in Prague, Czech Republic

The instrumental variables are one possibility of how to cope with the break of orthogonality condition however they are vulnerable by contamination. A robust modification of them - the instrumental weighted variables (IWV) - is resistant to the contamination and is consistent even under heteroscedasticity. When the estimated regression coefficients are subject to some constraints, IWV are able to cope also with collinearity. The proof of their consistency (under heteroscedasticity) will be outlined but the main focus will be on demonstration that the estimation with constraints gives better results than the ridge regression and that it works even under high collinearity and heavy contamination.

ES133 Room I1 APPLIED STATISTICS AND DATA ANALYSIS I Chair: Thomas Kneib

E931: **Impact of unknown dropout reasons on estimation of treatment effect: application to MAGNETIC trial**
Presenter: **Ruwanthi Kolamunnage-Dona**, University of Liverpool, United Kingdom

MAGNETIC is the largest, randomised, double-blind, placebo-controlled study to date that compared the addition of three doses of nebulised MgSO4 or placebo to standard treatment in children aged between 2 and 16 years. The main objective of the trial was to determine whether the use of MgSO4, when given as an adjunct to standard therapy for 60 minutes in acute severe asthma in children results in a clinical improvement compared with standard treatment alone. Asthma severity was assessed using a validated Yung Asthma Severity Score (ASS) and was recorded at

randomisation, and at 20, 40, 60 minutes after each nebuliser treatment, and at further follow-up at 120, 180, and 240 minutes post randomisation. As with most longitudinal studies, MAGNETIC also encountered patient withdrawals (or "dropout") during the treatment and follow-up. The reasons for dropout were sometimes clearly mentioned, but in many instances these reasons were unknown. We explore the impact of dropout on estimation of the treatment effect using the joint modelling of longitudinal data and time to study dropout. To account for unknown reasons of dropout, we consider a more advanced extension of joint modelling where reasons of dropout is considered within completing risks settings.

E1214: **Two Bayesian approaches to estimate aortic stiffness from patient-specific CTA images**
Presenter: **Ilaria Bianchini**, CNR-Imati, Italy
Co-authors: Raffaele Argiento, Ettore Lanzarone

Arterial *stiffness* is the capability of a vessel to vary its section secondary to blood pressure variations. It is a significant predictor of cardiovascular morbidity and mortality. The aim is to estimate the arterial stiffness and its spatial variations in a given aortic region. Data are patient-specific geometrical information non-invasively derived from Computed Tomography Angiography (CTA) images, and associated pressures obtained by means of a lumped parameter model. Indeed, we observe radii and pressures along an aortic segment (i.e., aortic arch and descending aorta) over a cardiac cycle. A non-linear differential constitutive model relates these quantities, in which Young and Poisson moduli are the unknown parameters to estimate. Such parameters are estimated in two different ways using a Bayesian approach. On the one hand, we extend a previously developed MCMC algorithm based on a time series representation of data. On the other hand, we exploit the ODE penalized B-spline approach. Both approaches are validated on numerical datasets and applied to a real clinical case. The final goal is to support the adoption of non-invasive CTA imaging in the clinical practice as a standard tool for large scale diagnosis and treatment planning of aortic diseases.

E806: **Modeling a longevity index for Spanish population**
Presenter: **Pilar Requena**, Alcala University, Spain
Co-authors: Valeria D'Amato

Insurers and reinsurers are typical longevity risk takers. Nevertheless, the capital market provides a significant alternative, in particular for the risk management of pension funds and annuity providers. Several longevity-linked securities have been proposed in the literature: longevity bonds, survivor swaps, mortality futures, mortality options and longevity caps and floors, but only few of them were issued in the market. Significant attempts to create products providing an effective transfer of longevity risk have been made by practitioners and investment banks. The development of a liquid market in traded mortality needs a standardized and transparent measurement of longevity. Recently, many different indices have been proposed, which aim at increasing the transparency and improving the longevity market. In particular, we analyze the named LifeMetrics Index which has these features. The mortality metrics used in the index are obtained by non-parametric smoothing techniques. Specifically Beers' ordinary minimized fifth difference method and cubic spline smoothing. The results are data for evaluating current and historical levels of longevity. We propose to apply the index principles suggested by LLMA-2012 on Spanish national population data from 1990 to 2013. We compare the results with longevity data obtained by different methodologies during the same period.

E1224: **Performance of structured additive regression models (STAR) when dealing with complex breast cancer screening data**
Presenter: **Elisa Duarte**, Universidade de Santiago de Compostela, Spain
Co-authors: Bruno de Sousa, Carmen Cadarso-Suarez, Thomas Kneib, Vitor Rodrigues

Nowadays, the various areas of study that use statistics as a decision support tool will have to deal with large datasets. Thus, it is of the utmost importance that software used in these analyses be prepared to face this issue, managing the balance between the physical resources necessary to give timely reliable results. The aim is to show the performance of structured additive regression models (STAR) when dealing with a screening dataset, provided by the Portuguese Cancer League (LPCC), consisting of 212,517 women who attended the Breast Cancer Screening Program in central Portugal between 1990 and 2010. The variables in the study are: breast cancer diagnosis, year of birth, age of menarche and menopause, pregnancy and nursing status, use of oral contraceptives, family history of breast cancer, municipality purchasing power index (PPI), a woman's place of residence, and the interaction between age of menopause and age of menarche. The model will explore how the multiple factors will relate to the probability of a woman having breast cancer. BayesX version 2.1 was used to perform the analysis. The model converged in 14 hours, after 17 iterations, using an 8x Intel 3.33GHz machine, with 22Gb of RAM.

ES138 Room C1 CONTRIBUTIONS TO BIOSTATISTICS AND BIOINFORMATICS Chair: Kim-Anh Do

E532: **One-tailed asymptotic inferences for the difference of proportions**
Presenter: **Maria Alvarez Hernandez**, University of Granada, Spain
Co-authors: Antonio Martin Andres

The one-tailed confidence intervals for the difference d of two proportions are useful in the field of clinical trials, drug bioavailability studies, vaccine efficiency or conformity industrial testing. This type of inferences has been widely used in the literature, from the exact point of view (computationally intensive) and from the asymptotic point of view (a great number of these methods have been proposed in the current literature). In the inferences about one proportion, several authors proved that the performance of the same method may often vary greatly from one-tailed case to two-tailed case. Additionally to the practical importance, this is the reason why the analysis of one-tailed case is also required for d. The paper evaluates several asymptotic methods for obtaining a one-tailed confidence interval for the difference between two binomial proportions (most of them are new proposals) and comes to the conclusion that the optimal method is based on the arcsine transformation with the increase of data proposed by Anscombe.

E782: **Mediana: R package for power evaluation in clinical trials with multiplicity adjustment methods**
Presenter: **Gautier Paux**, Institut de Recherches Internationales Servier IRIS, France
Co-authors: Alex Dmitrienko

Clinical development programs and clinical trials should be designed to ensure that, if a treatment is effective, there is a high probability to detect an effect of a given size. In modern drug development, sponsors are interested in assessing the efficacy of a new treatment on multiple endpoints, evaluating multiple doses compared to a control, or determining a treatment effect in multiple subgroups. Due to this multiplicity of tests, the probability of erroneously claiming the effectiveness of a new drug, i.e. the Type I error rate, will be inflated and must be controlled to support reliable statistical inferences. In recent decades, multiple new methods for addressing multiplicity issues in clinical trials have been developed. In the context of clinical trials with multiple objectives, sample size and power calculations should reflect the multiple testing strategy to be used. However, general analytical expressions of the power function do not exist and power evaluation is simulation-based. The Mediana R package provides a general framework for the power evaluation in clinical trials with multiplicity issues. It is based on the Clinical Scenario Evaluation approach, allowing a quantitative assessment of operating characteristics of candidate designs and statistical methods to characterize their performance in multiple settings.

E940: **Inference of gene regulatory relationship in C.elegans embryo**
Presenter: **Jie Hu**, The Chinese University of Hong Kong, China
Co-authors: Xiaodan Fan

One fundamental question in biology is how a zygote develops into an embryo with different tissues. To approach this goal, experimental biologists have produced large-scale 4D confocal microscope movies of C.elegans embryos. However, the lack of principled statistical tools has hindered

the comprehensive analysis of these spatial-temporal data sets. Based on our previous detection of the expression onset of each gene in each cell lineage, we aim at developing a probabilistic framework for evaluating possible regulatory relationships among profiled genes. Regulatory models with one or two parents are considered. Time series of the parent(s) and the child are fitted into a dynamics model based on time-lagged regression. Bayesian methods are used for performing inference. An information-criterion type of measure is used to select statistically significant gene regulatory relationships. Known regulatory relationships in current literatures and databases are used to evaluate our results.

E235: **PLS-DA for metabolomical (compositional) data using the logratio approach**
Presenter: **Alzbeta Kalivodova**, Palacky University Olomouc, Czech Republic
Co-authors: Karel Hron, Peter Filzmoser

Results of metabolomical measurements are often expressed as observations carrying only relative information, called compositional data. They are formed by positive vectors, whose parts contain quantitatively expressed relative contributions of parts on a whole; specially, proportions or percentages are often used for their constant-sum constraint representation in practice. Partial least squares discriminant analysis (PLS-DA) is a popular method in chemometrics used for classification of high-dimensional data. In order to enable processing of compositional data using PLS-DA, the logratio methodology, leading to new coordinates with respect to natural geometric features of compositions, needs to be applied. Consequently, for the obtained estimates of regression parameters (metabolites) further statistical inference using bootstrap techniques can be employed. Theoretical developments will be applied to real-world data from targeted metabolomics, used to analyze the significance of the corresponding metabolites by applying PLS-DA, adjusted for dealing with compositional data, in order to highlight possible markers of inherited metabolic disorders. A comparison with other preprocessing approaches to metabolomical data will be made in order to highlight the positive aspects of the logratio methodology for the identification of important biomarkers.

ES61 Room D1 MIXTURE MODELS FOR MODERN APPLICATIONS II Chair: Geoff McLachlan

E863: **Model selection in overlapping stochastic block models**
Presenter: **Christophe Ambroise**, Laboratoire Statistique et Genome, Genopole, Evry, France

Networks are a commonly used mathematical model to describe the rich set of interactions between objects of interest. Many clustering methods have been developed in order to partition such structures, among which several rely on underlying probabilistic models, typically mixture models. The relevant hidden structure may however show overlapping groups in several applications. The Overlapping Stochastic Block Model has been developed to take this phenomenon into account. Nevertheless, the problem of the choice of the number of classes in the inference step is still open. To tackle this issue, we consider the proposed model in a Bayesian framework and develop a new criterion based on a non-asymptotic approximation of the marginal log-likelihood. We describe how the criterion can be computed through a variational Bayes EM algorithm, and demonstrate its efficiency by running it on both simulated and real data.

E1061: **Mixture model averaging for clustering**
Presenter: **Yuhong Wei**, McMaster University, Canada
Co-authors: Paul McNicholas

In mixture model-based clustering applications, it is common to fit several models from a family and report clustering results from only the 'best' one. In such circumstances, selection of this best model is achieved using a model selection criterion, most often the Bayesian information criterion. Rather than throw away all but the best model, we average multiple models that are in some sense close to the best one, thereby producing a weighted average of clustering results. Two (weighted) averaging approaches are considered: averaging component membership probabilities and averaging models. In both cases, Occam's window is used to determine closeness to the best model and weights are computed within a Bayesian model averaging paradigm. In some cases, we need to merge components before averaging; we introduce a method for merging mixture components based on the adjusted Rand index. The effectiveness of our model-based clustering averaging approaches is illustrated using a family of Gaussian mixture models on real and simulated data.

E1211: **Maximum likelihood estimation in a hybrid logistic model**
Presenter: **Kensuke Okada**, Senshu University, Japan
Co-authors: Shin-ichi Mayekawa

In existing logistic regression models, the logit of the response probability is related to a liner combination of the predictor variables. This model is called compensatory, because a high score in one predictor "compensates" for the low score in others. There can be another approach, in which the response probability is represented as a product of logistic functions of each predictor variable. This model is called noncompensatory. We propose a new hybrid logistic regression model, in which the response probability is given by the weighted sum of both compensatory and noncompensatory response probabilities. A maximum likelihood estimation for the hybrid model is also proposed. The log likelihood function is maximized using Newton-type non-linear maximization. The choice of initial values is important in this type of algorithm. The initial values of both compensatory and noncompensatory parameters are obtained by separate fitting of each single model. The initial value of the weight parameter is obtained by one-dimensional optimization given the fixed initial values of compensatory and noncompensatory parameters. The simulation study revealed that the proposed method can be successfully used when one has large enough sample size.

E1017: **A flexible ratio regression approach for zero-truncated capture-recapture counts**
Presenter: **Dankmar Boehning**, University of Southampton, United Kingdom

Capture-Recapture methods are frequently used to estimate the size of population of interest which can be only partially observed. Each member of the population carries a count of identifications by some identifying mechanism - the number of times it has been identified during the observational period. Only positive counts are observed and inference needs to be based on this observed, truncated, count distribution. For valid prediction of the number of unobserved units, it is crucial to use an appropriate count data model. According to previous works in the field, we consider ratios of neighbouring count probabilities; these can be estimated by ratios of corresponding observed frequencies, independent of whether they arise from zero-truncated or untruncated count distributions. The guiding principle here is that it is often easier to find an appropriate regression model than working directly with modelling the distributional form of the count distribution. We show how these ratios can be modeled by a regression approach, with a suitable link function. It is shown that this process leads not only to a valid count distribution, but also to a wide class of models. This is illustrated by analyzing various examples using fractional polynomials though other classes would be possible as well. Some simulation studies show the potential of this approach. Some other questions such as identifiability can be efficiently approached within the ratio regression framework.

ES134 Room L1 STATISTICAL INFERENCE I Chair: Jochen Einbeck

E913: **On quasi BIC for general LAQ model**
Presenter: **Hiroki Masuda**, Kyushu University, Japan
Co-authors: Shoichi Eguchi

The precise asymptotic behavior of the logarithmic marginal quasi-likelihood is derived when the underlying model is of general locally asymptotically quadratic (LAQ) type with bounded prior density. Under mild regularity conditions, our result enables us to explicitly describe the non-negligible part of the stochastic expansion as a functional of the quasi-maximum-likelihood estimator. In particular, it provides us with a very practical BIC-like statistics for assessing different LAQ-type quasi-likelihood models. The novelty is that we can deal with random asymptotic

Fisher information matrices and also possible model misspecification in a unified and verified manner; thus, non-ergodic models such as the case of estimation of random integrated volatility for Itô processes with or without jumps are within our scope. The proof is essentially based on the uniform tail-probability estimate of the underlying quasi-likelihood random fields.

E969: **Decompounding: a non-parametric Bayesian approach**
Presenter: **Shota Gugushvili**, Leiden University, Netherlands
Co-authors: Peter Spreij, Frank van der Meulen

Given a sample from a discretely observed compound Poisson process, we consider non-parametric estimation of the density f_0 of its jump sizes, as well as of its intensity λ_0. We take a Bayesian approach to the problem and specify the prior on f_0 as the Dirichlet location mixture of normals. An independent prior for λ_0 is assumed to be compactly supported and possess a positive density with respect to the Lebesgue measure. We show that under suitable additional assumptions the posterior contracts in the Hellinger metric around the pair (λ_0, f_0) at essentially (up to a logarithmic factor) the $\sqrt{n}$-rate, where n is the number of observations. In particular, this implies the existence of Bayesian point estimates converging (in the frequentist sense, in probability) to (λ_0, f_0) at the same rate.

E1062: **Construction of confidence intervals for high-dimensional single-index models**
Presenter: **Thomas Gueuning**, KU Leuven, Belgium
Co-authors: Gerda Claeskens

In the high-dimensional data framework, penalty terms such as LASSO or SCAD are usually added to the loss function and provide sparse estimates. The construction of confidence intervals for parameters is not as straightforward as in the classical low-dimensional data framework. Recently, a new method has been developed, the idea being to add a correction term to the penalized estimator. A desparsified estimator is then obtained for which asymptotic normality can be proven for generalized linear models and the construction of confidence intervals is then possible. We extend this method to partially linear single-index models in which a non-parametric term is present. We fit single-index models to high-dimensional data using penalty terms, we prove the asymptotic normality of the desparsified estimator and we construct confidence intervals for parameters. The simulation results show that the method performs well for high-dimensional single-index models.

E1183: **Testing for constancy in varying coefficient models**
Presenter: **Anneleen Verhasselt**, Hasselt Univeristy, Belgium
Co-authors: Mohamed Ahkim

Linear regression models are often too rigid for regression analysis. We consider varying coefficient models, which are an extension of the classical linear regression models in the sense that the regression coefficients are functions in certain variables (often time t). Varying coefficient models have been popular in longitudinal data and panel data studies, and have been applied in fields such as finance and health sciences. We estimate the coefficient functions with P-splines. An important question in a varying coefficient model is whether an estimated coefficient function is significantly different from a constant (or zero). We develop testing procedures based on the P-spline coefficients by making use of properties of B-spline basis expansions. The performances of the proposed testing methods are illustrated on simulated data and on real data.

ES137 Room G1 CONTRIBUTIONS TO NETWORK AND SPATIAL DATA ANALYSIS Chair: Vladimir Batagelj

E958: **Efficient Bayesian computation for exponential random graph models**
Presenter: **Alberto Caimo**, University of Lugano, Switzerland
Co-authors: Antonietta Mira

Powerful ideas recently appeared in the literature are adjusted and combined to design improved samplers for doubly intractable target distributions with a focus on Bayesian exponential random graph models. Different forms of adaptive Metropolis-Hastings proposals (vertical, horizontal and rectangular) are tested and merged with the delayed rejection (DR) strategy with the aim of reducing the variance of the resulting MCMC estimators for a given computational time. The DR is modified in order to integrate it within the approximate exchange algorithm (AEA) to avoid the computation of intractable normalising constant that appears in exponential random graph models. This gives rise to the AEA+DR: a new methodology to sample doubly intractable distributions that dominates the AEA in the Peskun ordering leading to MCMC estimators with a smaller asymptotic variance.

E1004: **Specifying and estimating relational event models with time-weighted statistics: an application to the interbank market**
Presenter: **Paola Zappa**, University of Lugano, Switzerland
Co-authors: Duy Q. Vu, Caterina Liberati, Alessandro Lomi

The aim is to deal with the modeling of continuous time relational event data containing information on sequences of time-stamped events encoding interactions between economic agents. We draw on newly derived relational event models and we extend them by defining time-weighting functions and specifying the corresponding time-weighted statistics that we use to account for coexistence and differentiated effects of short-term events and long-term "relations". By fine tuning the weighting-function parameters, we show how it may be possible to obtain more realistic and accurate representations of the data structure. Also, by comparing the time-weighted model with alternative specifications of time effects, we emphasize the parsimony afforded by the proposed approach. We illustrate advantages and empirical value of our model extension in a study of the effect of the recent financial crisis on banks' counterpart selection in interbank markets. Using an original dataset on tick by tick transactions in the EU interbank liquidity market during 2006-2009, we examine the dynamics of banks' lending activities and look at how they have changed during the observation period. We document differences in the propensity of instantaneous relational events to give rise to enduring network ties between trading partners in the interbank liquidity market during the financial crisis.

E1216: **Marginal modelling of spatial clustered binary data**
Presenter: **Manuela Cattelan**, University of Padova, Italy
Co-authors: Cristiano Varin

Spatial clustered data have been attracting increasing interest in recent years. Marginal modelling of spatial binary clustered data is discussed. Generalized estimating equations are the most popular method for marginal regression analysis of data constituted by independent clusters. We suggest a modification of generalized estimating equations to handle spatial dependence between clusters. Dependence is modelled by assuming that the cross-odds ratio is a smooth decaying function of the distance between observations. Inference is performed through the hybrid pairwise likelihood method, a two-step method where marginal parameters are estimated via optimal estimating equations while dependence parameters are estimated via pairwise likelihood. The methodology is illustrated through an application to a data set about the malaria prevalence in children recorded at villages in the Gambia.

E1260: **Combining observed-simulated extreme wind speeds for generating risk maps for electric grid disruptions in Portugal**
Presenter: **Kamil Turkman**, FCUL University of Lisbon, Portugal
Co-authors: Antonia Turkman, Patricia de Zea Bermudez, Paula Pereira, Pedro Miranda, Pedro Viterbo

Risk maps that indicate likely places of costly disruptions in electric grids are important decision support tools for administering the power grid. Although many factors contribute to the occurrence of damages, the most significant factor is extreme wind speed. However, generating such risk maps depends on reliable wind data at fairly high spatial and temporal resolutions. Wind data are typically available at a limited number of monitoring sites, often with a significant number of missing observations. On the other hand, data from a numerical-physical model are available

at regular grid cells level, obtained at high spatial and temporal resolutions. The problem is that simulated and observed extreme wind speeds, particularly at some stations, do not match well. Consequently the corresponding vulnerability maps based on simulated and observed daily maximum wind speeds may differ quite substantially. More accurate exposure assessment of wind on power grid disruptions can be made by combining observed and simulated data. The objective is to explore several methods to model the relationship between simulated and observed wind speeds at observation sites in order to extrapolate this relationship in space at grid cell resolution. We show how this strategy improves the quality of the vulnerability maps produced.

ES135 Room B1 CONTRIBUTIONS TO STATISTICS OF EXTREMES AND APPLICATIONS Chair: Jan Beirlant

E1136: **A note on a diagnostic for selecting the threshold in statistics of univariate extremes**

Presenter: **Ivette Gomes**, University of Lisbon, Portugal

The main objective of statistics of univariate extremes is the prediction of rare events, and thus the need for an adequate estimation of parameters related to natural disasters. The primary parameter is the extreme value index (EVI). One of the most recent and general approaches is the semi-parametric one, where it is merely assumed that the underlying model F is in the domain of attraction of a general EV distribution. The EVI estimation is then based on the top k order statistics in the sample or on the excesses over a high level u, and the choice of either k or u is crucial. The most common methods of threshold selection are based on the minimization of some kind of mean squared error estimator. Here, and for heavy right tails, i.e. a positive EVI, a threshold selection based on bias properties is considered. The methodology depends on a tuning parameter, and a choice of such a tuning parameter is provided. In order to achieve our objectives the asymptotic behaviour of an adequate linear combination of the scaled log-spacings is derived, the use of two alternative auxiliary statistics is suggested, and some overall comments are put forward.

E1172: **On the asymptotic behaviour of extreme geometric quantiles**

Presenter: **Stephane Girard**, Inria, France

Co-authors: Gilles Stupfler

A popular way to study the tail of a distribution is to consider its extreme quantiles. While this is a standard procedure for univariate distributions, it is harder for multivariate ones, primarily because there is no universally accepted definition of what a multivariate quantile should be. We focus on extreme geometric quantiles. We discuss their asymptotics, both in direction and magnitude, when the norm of the associated index vector tends to one. In particular, it appears that if a random vector X has a finite covariance matrix M, then the magnitude of its extreme geometric quantiles grows at a fixed rate and is asymptotically characterised by M. The case when X does not have a finite covariance matrix is tackled in a multivariate regular variation framework. The results are illustrated on simulated data.

E1274: **Clustering for regional time trend in the nonstationary heavy-tailed distribution**

Presenter: **Jong-June Jeon**, University of Seoul, Korea, South

A clustering method for regional time trend which is required for risk analysis is proposed. Applying regularization method to cluster regional trends of the nonstationary heavy-tailed distribution, we can estimate consistent clusters of time trends regardless of the distributions. The proposed method also gives a diagram of clustering time trends of sites, which is useful to manage the future risk.

E051: **Robust and bias-corrected estimation of extreme failure sets**

Presenter: **Armelle Guillou**, Strasbourg, France

Co-authors: Christophe Dutang, Yuri Goegebeur

In multivariate extreme value statistics, the estimation of probabilities of extreme failure sets is an important problem, with practical relevance for applications in several scientific disciplines. Some estimators have been introduced, though so far the typical bias issues that arise in application of extreme value methods and the non-robustness of such methods with respect to outliers were not addressed. We introduce a bias-corrected and robust estimator for small tail probabilities. The estimator is obtained from a second order model that is fitted to properly transformed bivariate observations by means of the minimum density power divergence technique. The asymptotic properties are derived under some mild regularity conditions and the finite sample performance is evaluated by a small simulation experiment. We illustrate the practical applicability of the method on a dataset from the actuarial context.

Monday 8.12.2014 10:35 - 12:15 Parallel Session P – CFE

CS112 Room E2 FINANCIAL APPLICATIONS I **Chair: Demetris Koursaros**

C256: **Does the probability of informed trading model fit empirical data?**
Presenter: **Quan Gan**, University of Sydney, Australia

The probability of informed trading (PIN) is a widely used information asymmetry measure. The PIN estimation procedure estimates parameters in structural models which are finite mixture of bivariate independent Poisson distributions. These models have two drawbacks. First, the mean and the variance of a Poisson component are the same, constraining the model's ability to fit marginal data. Second, the dependence structure on mixture of independent distributions does not capture the dependence in data. It is shown that structural PIN models do not adequately fit empirical data. The goodness-of-fit tests unanimously reject these structural models in the sample of NYSE stocks trading data.

C587: **Dynamics of firm learning from stock prices: Evidence from Europe**
Presenter: **Houdou Basse Mama**, ESCP Europe Business School, Germany

The dynamics of firm learning from stock prices are explored. Using an unbalanced panel of 2,312 European firms over the period from 1991 to 2011, I find strong evidence that firms respond to peer firms' stock price shocks when making investment decisions. A one standard deviation increase in peers' stock price innovations commands a 6.96% increase in corporate investment. This response is more pronounced for firms with lower stock price informativeness relative to peers. Furthermore, this sensitivity increases in firm size and depth of peer firms' analyst following as well as in heightened competition. Relatedly, the responsiveness of investment to peer stock price innovations increases in the strategic value firms attach to information embedded in peer stock price shocks. Indeed, higher peer market share, ROA and growth in sales apparently dampen the propensity of the average firm to pay attention to peer stock prices. These findings corroborate the proposition that not only information motives but also strategic motives propel firms to use the feedback received from stock prices. Finally, European firms tend to be more conservative than their U.S. counterparts in learning from stock prices.

C984: **Measuring the instability of China's financial system**
Presenter: **Lixin Sun**, Shandong University, China
Co-authors: Yuqin Huang

By employing several techniques including GARCH modelling, Equal-variance Weighting, VAR approach, econometric benchmarking, and H-P filter, the aim is to construct a financial stress index (CNFSI) and a financial conditions index (CNFCI) to measure the instability of China's financial system. The indices are based on the monthly data collected from China's inter-bank markets, stock markets, foreign exchange markets and debt markets. Using these indices, we identify the episodes of systemic financial stress, and then evaluate and compare the indices by predictive tests, total errors and noise/signal analysis. The empirical analyses suggest that the CNFSI performs better than the CNFCI in our study. Furthermore, we propose four leading indicators for monitoring China's financial instability, and provide an early warning system for China's macroprudential regulations.

C1188: **On method finding arbitrage opportunities from different markets**
Presenter: **Yasushi Ota**, Doshisha, Japan

The inverse problem for option pricing theory in financial markets is investigated. One of the most interesting problems discerned when applying the Black-Scholes model to financial derivatives, is reconciling the deviation between expected and observed values. Black and Scholes first discovered how to construct a dynamic portfolio of a derivative security and the underlying asset. Their approach provides a useful, simple method of pricing inclusive of financial derivatives, risk premium, and default probability estimation under the assumption that the risky asset is log-normally distributed. However, the theoretical prices of options with different strike prices as calculated by the Black–Scholes model differ from real market prices. First, we explain our new model, which is a type of an arbitrage model and formulate our new mathematical approach to an inverse problem in financial markets. Next, we apply microlocal analysis which is a mathematical method to prove a uniqueness of the solution to our inverse problem. While microlocal analysis is used for various models in physics and engineering, this is the first attempt to apply it to a model in financial markets. Moreover, we provide several numerical examples of recovering a real drift in currency markets.

C868: **Investor's demographics and portfolio objectives: an empirical study using factor analysis**
Presenter: **Saurabh Agarwal**, Indian Institute of Finance, United Kingdom

Demographics affect Investor's psychology thereby affecting portfolio choices made by investors. In a survey of 512 retail investors, it was found that Risk bearing capability affects choice among portfolio objectives most and Years to Retirement influences this choice least. "R" factor analysis has been undertaken to reduce the originally identified eight factors into a minimum number of factors influencing portfolio objectives. Timing of Portfolio, Security from Portfolio, Knowledge of Portfolio Selection and Life Cycle Portfolio were finally identified as four factors affecting Portfolio Objectives.

CS95 Room M2 CONTRIBUTIONS ON PANEL DATA **Chair: Martin Wagner**

C1068: **Local power of panel unit root tests allowing for structural breaks**
Presenter: **Yiannis Karavias**, University of Nottingham, United Kingdom
Co-authors: Elias Tzavalis

The asymptotic local power of least squares based fixed-T panel unit root tests allowing for a structural break in their individual effects and/or incidental trends of the AR(1) panel data model is studied. Limiting distributions of these tests are analytically derived under a sequence of local alternatives and analytic expressions show how their mean and variance is a function of the break date and the time dimension. The considered tests have local power which tends to unity if the panel data model includes only individual intercepts. For panel data models with incidental trends, the power of the tests becomes trivial. However, this problem does not always appear if the tests allow for serial correlation in the error term and completely vanishes in the presence of cross section correlation. Monte Carlo experiments demonstrate the usefulness of the asymptotic theory in small samples.

C1115: **Dimensionality and long-run homogeneity effects in panel vector error correction model**
Presenter: **Piotr Keblowski**, University of Lodz, Poland

The focus is on the impact of the system dimension on the finite-sample properties of the maximum likelihood estimator of long-run parameters in the multidimensional non-stationary panels. The panel VEC model is considered with different sets of restrictions on the system's structure. Therefore, it is consecutively allowed that (i) cross-sectional dependence in the error terms occurs, (ii) there is interaction of the short-run dynamics between cross-sections, (iii) there is interaction of the error-correction terms between cross-sections. The results for the individual time-series analyses, where cross-sections are assumed to be independent, are related to the results of the panel analysis. It is shown that there is a trade-off between the dimensionality effect which is well known from the standard time-series analysis and efficiency gains, which are due to cross-sectional dependencies. However, if there is a common cointegration rank and significant cross-sectional relationships, then the results of the MLE of long-run parameters in panels usually outperform the results for the standard time-series analyses, where cross-sections are assumed to be independent.

Moreover the performance of the MLE of long-run parameters in the panel VEC is enhanced if the cross-sections share the same long-run structure, i.e. the common cointegration vectors exist.

C1235: **A formula for predicting loan loss reserves under adverse GDP scenarios for EU15 banks**
Presenter: **Savas Papadopoulos**, Democritus University of Thrace, Greece

One of the most useful tools of macroprudential policy is stress testing for obtaining financial stability of a banking sector. We are trying to build a simple model based on balance-sheet and macro data to assess capital needs. We analyze annual data from medium and large EU15 banks for the crises periods after 2008. We fit a dynamic logarithmic model for loan loss reserves on negative log-differences of GDP. Standard techniques for dynamic panel data models cannot be applied due to highly unbalanced data. We apply standard methods for cross-sectional and panel data such as LAD, robust, FE, RE, SUR, etc, and we report in-and-out-of-sample goodness-of-fit measures. The model has different slopes and intercepts depending on the provision size and predicts very satisfactorily. The robust methods excel relative to the other methods. The derived formula might be very useful to supervisors and policymakers. Selected EU15 banks are stressed and the results are provided.

C1251: **A factor analytical approach to price discovery**
Presenter: **Simon Reese**, Lund University, Sweden
Co-authors: Joakim Westerlund, Paresh Narayan

Existing econometric approaches for studying price discovery presume that the number of markets is small, and their properties become suspect when this restriction is not met. They also require making identifying restrictions and are in many cases not suitable for statistical inference. These shortcomings are taken as a starting point to develop a factor analytical approach that makes use of the cross-sectional variation of the data, yet is very user-friendly in that it does not involve any identifying restrictions or obstacles to inference.

C215: **Asymptotically UMP tests for unit roots in cross-sectionally dependent panels**
Presenter: **Ramon van den Akker**, Tilburg University, Netherlands
Co-authors: I. Gaia Becheri

The (asymptotic) power envelope for testing for a unit root in a cross-sectionally dependent PANIC panel is derived. We demonstrate that two tests previously proposed are both asymptotically UMP in case the long-run variances of the idiosyncratic components are homogeneous. We propose a modification of these tests that is asymptotically UMP irrespective of homogeneity of the long-run variances. These results hold irrespective of the unobserved common components being I(0) or I(1).

CS101 Room B2 CONTRIBUTIONS ON TIME SERIES ECONOMETRICS III Chair: Sung Keuk Ahn

C046: **A general theory of rank testing**
Presenter: **Majid Al Sadoon**, Universitat Pompeu Fabra, Spain

An approach to rank testing that nests all existing rank tests and simplifies their asymptotics is proposed. The approach is based on the fact that implicit in every rank test there are estimators of the null spaces of the matrix in question. The approach yields many new insights about the behavior of rank testing statistics under the null as well as local and global alternatives in both the standard and the cointegration setting. The approach also suggests many new rank tests based on alternative estimates of the null spaces as well as the new fixed–b theory. A brief Monte Carlo study illustrates the results.

C041: **Fractional cointegration analysis of industrial metal prices**
Presenter: **Andreas Dechert**, University of Wuerzburg, Germany

The aim is to investigate the spot price relation between different industrial metals with fractional cointegration methods, which can be seen as a generalization of the classical cointegration approach. The motivation of their usage is that the application of traditional methods might be to restrictive as empirical data require a more flexible econometric approach. Despite classical cointegration methods like the ADF residual test amongst others indicate a cointegration relation among some of the metals, we find a slowly decaying structure of the autocorrelation function of the cointegration error terms. This can be explained by a long memory property and shows that shocks of the industrial metal market have long lasting effects on prices. For evaluation of this effect the fractional cointegration approach is chosen and estimated by current methods like the multivariate exact local Whittle estimator, which yields semiparametric estimates of integration orders and cointegration parameters in a unified approach, and the parametric fractional cointegrated vector autoregression model (FCVAR).

C1120: **A persistence-based Wold-type decomposition for stationary time series**
Presenter: **Federico Severino**, Bocconi University, Italy
Co-authors: Fulvio Ortu, Andrea Tamoni, Claudio Tebaldi

The Classical Wold Decomposition Theorem allows us to split a weakly stationary time series $\mathbf{x}$ into a non-deterministic component, driven by independent innovations, and a deterministic term. This decomposition is a special case of the more general Abstract Wold Theorem, well-known in functional analysis, which deals with isometric operators defined on Hilbert spaces. As the lag operator is isometric on the Hilbert space $\mathcal{H}_t(\mathbf{x})$ spanned by the sequence $\{x_{t-k}\}_{k\in\mathbb{N}_0}$, the Classical Wold Decomposition for time series obtains. When different isometries are involved, new Wold-type decompositions arise. In particular, when $\mathbf{x}$ is purely non-deterministic, the Haar operator is isometric on the Hilbert space $\mathcal{H}_t(\varepsilon)$, spanned by the classical Wold innovations of $\mathbf{x}$, and it provides a persistence-based decomposition. Indeed, the use of the Haar transform is rather popular in the analysis of macroeconomic and financial time series, with the aim of disentangling low frequency shocks, that survive in the long run, from temporary fluctuations. This methodology is embedded in a theoretical framework, where the Abstract Wold Theorem delivers a decomposition of the stochastic process $\mathbf{x}$, in which the concerned components explain different layers of persistence. This decomposition fruitfully applies to ARMA and ARFIMA processes.

C1142: **A CUSUM of squares test for cointegration based on OLS residuals with a model free limiting null distribution**
Presenter: **Julio Angel Afonso-Rodriguez**, University of La Laguna, Spain

A new and simple method is proposed to compute semiparametric CUSUM-type statistic based on the sequence of centered and squared OLS (Ordinary Least Squares) residuals from the estimation of a single-equation cointegrating regression model as the basis to test the null hypothesis of cointegration against no cointegration. The main novelty of this testing procedure is that, besides very simple and standard semiparametric corrections for serial correlation and endogeneity of the integrated regressors and the only use of OLS residuals, the non-standard limiting null distribution is invariant to the number and type of components appearing in the estimated regression, including both the number and order of the stochastic and deterministic trend components, respectively. This prominent result avoids the use of different quantiles to perform the test and thus allows us to capture the persistent nature of the regression errors. We derive such a limiting null distribution, establish its consistency rate under no cointegration and also present some numerical results to illustrate its finite-sample size and power properties, indicating a very good performance as a complementary testing procedure to the existing ones in this context. We also provide an empirical application to the US aggregate consumption function.

C1166: **Second-order least-squares estimation for regression models with ARMA errors: simulation results**
Presenter: **Dedi Rosadi**, Gadjah Mada University, Indonesia
Co-authors: Shelton Peiris

Previously, it has been shown that the second-order least squares estimator (SLSE) was more efficient than the ordinary least squares estimator (OLSE) to estimate regression model with iid (independent and identically distributed) and non-zero third moments errors. Other authors generalized the theory of SLSE to regression models with autoregressive errors. Under certain regularity conditions, we establish the consistency and asymptotic normality of the proposed estimator and compare its numerical performance with the corresponding OLSE and GLSE (Generalized Least Square Estimator). We extend the results into regression model with ARMA errors in particular, and provide simulation results.

CS04 Room N2 CONTRIBUTIONS ON VOLATILITY AND CORRELATION MODELLING Chair: Manfred Gilli

C902: **Conditional heteroskedasticity of return range processes**
Presenter: **Yan Sun**, Utah State University, United States
Co-authors: Jennifer Loveland, Isaac Blackhurst

Price range contains important information about the asset volatility, and has long been considered an important indicator for it. It is proposed to jointly model the [low, high] price range as a random interval and introduce an interval-valued GARCH (Int-GARCH) model for the corresponding [low, high] return range process. It can be viewed as an extension of the point-valued GARCH model that allows for interval-valued inputs. Yet its most important contribution lies in its capability to integrate realized measures and return-based model mechanism to produce information-rich estimation of the volatility. Model properties are presented under the general framework of random sets, and the parameters are estimated by a metric-based conditional least squares (CLS) method. The empirical analysis of the daily return range data of Dow Jones component stocks shows that the proposed Int-GARCH model enriches the low-frequency volatility model such as GARCH with high-frequency information, without having to suffer from the microstructure noises. It is further shown that Int-GARCH has improved model flexibility: the GARCH mechanism prevents it from being too sensitive to data like the model-free realized volatility (RV); the role of the intraday range also makes it less rigid than GARCH itself.

C1013: **Securities portfolio risk estimation and forecasting by the use of Bayesian self-organizing maps**
Presenter: **Mikhail Grishko**, Kazakh National University, Kazakhstan
Co-authors: Anuar Dyusembaev, Peter Andras

An approach is presented to estimating and forecasting the risk of a securities portfolio by using Bayesian self-organizing maps (BSOMs). We use BSOM to construct a multidimensional probability density function (PDF) for the portfolio's components. This allows us to reflect internal portfolio structure, namely its multi-component correlations, more precisely than with other methods based on modelling distribution of portfolio values. A securities portfolio PDF is represented as a Gaussian mixture with K kernels $(w_1,\ w_2,...,\ w_K)$ assigned to each portfolio component: $P(q(B_x)) = \sum_{i=1}^{K} p(q|w_i,\ \theta_i)P(w_i)$, where $p(qw_i,\ \theta_i)$ is a PDF for the w_i-th mixture kernel, $P(w_i)$is the prior probability of the data belonging to the w_i-th mixture kernel, $q(B_x)$ is the cost movement of portfolio $B_x = \{(b_1,b_2 ...,\ b_N)|b_i \geq 0,\ X_0(b) = x.\ \}$ and $b = (b_1,b_2,\ ...,\ b_N)$ is a vector of portions of components in the portfolio. The portfolio risk evaluation concludes by calculating the integral $p(z)$ for each given portfolio cost movement value z: $p(z) = \iint_{b \bullet q(B_x)=z;} P(q(B_x))dq(B_x)$, which is the projection (in a geometrical sense) of a given PDF onto the portfolio defining the hyperplane $b \bullet q(B_x) = z$. A test on real stock data demonstrates that the proposed model has better long-term risk forecasting abilities than classical models such as Histogram estimation, Markowitz model, Gaussian approximation and Kernel density estimation.

C1122: **Realized hedge ratio: predictability and hedging performance**
Presenter: **Vasiliki Skintzi**, University of Peloponnese, Greece
Co-authors: Chrysi Markopoulou, Apostolos-Paul Refenes

The aim is to explore the dynamic properties and predictability of the Realized Minimum Variance Hedge Ratio (RMVHR), constructed from five-minute spot and future returns of two stock indices and two exchange rates. A number of econometric models are employed to forecast directly the RMVHR and the out-of-sample performance is evaluated. Results from statistical measures suggest the presence of predictable dynamics in the evolution of the realized hedge ratio series. In terms of risk reduction, we conclude that realized hedge ratio forecasts dominate conventional methods that use daily data while the benefit is pronounced when economic gains are considered. The superior performance of RMVHR methods holds across different asset classes but is more conspicuous in the case of stock indices. Finally, this study assesses the effect of sampling frequency and transaction costs on hedging performance.

C1164: **Realized wavelet-based estimation of integrated covariance and co-jumps in the presence of noise**
Presenter: **Lukas Vacha**, FSV Charles University in Prague, Czech Republic
Co-authors: Jozef Barunik

Wavelet-based estimator of integrated covariation is proposed. Basing our estimator in two-scale covariance framework, we are able to utilize all available data and get unbiased estimator in the presence of noise as well. In addition, we focus on detection of co-jumps, and our estimator is able to distinguish between idiosyncratic jumps and co-jumps. The estimator is tested in a small sample numerical study, and is compared to other popular integrated covariation estimators under different simulation settings with changing noise as well as jump and co-jump levels. The results reveal that our wavelet-based estimator is able to estimate the realized measures with the greatest precision. Another notable contribution lies in the application of the presented theory. Our time-frequency estimators not only produce more efficient estimates, but also decompose the integrated covariation into arbitrarily chosen investment horizons.

C1187: **A statistical test for the Heston model**
Presenter: **Gianna Figa-Talamanca**, University of Perugia, Italy

Even though classical stochastic volatility models are outdated from a modelling viewpoint, a renewed attention has been devoted in the last few years to the Heston stochastic volatility model which is an affine model so that a quasi-closed formula is available for computing the price of European plain vanilla options. This property makes it possible to calibrate the model on a sample of market prices for derivative financial instruments. Calibration of parameters may be performed on market option prices as suggested or considering more complex derivatives, for which a quasi closed formula has been made available in recent years. Among these, volatility and variance swaps and options. Motivated by this new interest and by the huge increasing of suggestions for estimation and calibration techniques of stochastic volatility models, often leading to very different parameters values, a formal statistical test for the Heston model is suggested. More precisely, when a parameter set is estimated from a specific time series, assuming that the Heston model is the underlying data generation process, the test should detect whether the data auto-covariance structure is consistent with its theoretical counterpart. The test statistic is built upon the outcomes of a previous paper.

CS66 Room H2 CONTRIBUTIONS ON BANKING AND FINANCIAL MARKETS Chair: Jozef Barunik

C106: **Economic surprises and inflation expectations**
Presenter: **Magdalena Grothe**, European Central Bank, Germany
Co-authors: Soren Autrup

Price formation is analyzed in medium- to longer-term maturity segments of euro area and US inflation-linked and nominal bond markets around the releases of important economic indicators. We compare the pre-crisis and crisis periods, controlling for liquidity effects observed in financial markets. The results allow us to draw conclusions about the anchoring of inflation expectations in the two currency areas before and during the

crisis. We find a somewhat stronger anchoring of inflation expectations in the euro area than in the United States. During the crisis, the degree of anchoring of inflation expectations did not change in the euro area, but it decreased to some extent in the United States.

C749: **Macroprudential policy transmission in a small open economy with traditional and matter-of-fact financial frictions**
Presenter: **Fabia Carvalho**, Central Bank of Brazil, Brazil
Co-authors: Marcos Castro, Silvio Costa

The aim is to investigate the transmission channel of macroprudential instruments in a small open economy with important flows of funds and goods to and from the rest of the world, and with a banking sector that takes risk in retail and commercial loan concessions. Retail loan concessions are tightly related to banks' expectations with respect to borrowers' future labor income. Banks optimize their balance sheet composition under frictions intended to reproduce banks' strategic reactions to changes in funding costs, in risk perception and in the regulatory environment. We calibrate the model to the Brazilian economy and investigate the transmission channel of reserve requirements and capital requirements on the credit market and on the real economy. We also analyze the impact of unfavorable international conditions on the domestic credit market and other banking activities.

C1049: **Pricing in a complex financial market: instability from local measures and model uncertainty**
Presenter: **Giacomo Livan**, University College London, United Kingdom
Co-authors: Marco Bardoscia, Matteo Marsili, Tomaso Aste

Financial institutions need to price a huge number of instruments every day, and each of them is usually priced according to models calibrated on financial data. The key issue, however, is that different institutions price the same instruments using different models, calibrating them on different datasets. More formally, this is a situation where each instrument is priced according to a local market measure, which may be specific to that instrument and/or to that institution. It has been argued that such a practice may give rise to a system of inconsistent prices and, ultimately, to arbitrage opportunities. This issue is addressed within the simple framework of a one-period economy, where the differences in the pricing models are introduced as differences in the probability measures used to price different financial instruments: depending on the heterogeneity of such measures, the model undergoes a sharp transition separating an arbitrage free market state from a state where arbitrage opportunities do arise. Within the same framework, a notion of model uncertainty will be also introduced that naturally arises whenever agents perform pricing based on models that fail to take into account the entirety of the available instruments.

C1157: **Anomalous price dynamics at resistance line in stock and forex markets**
Presenter: **Tomoshiro Ochiai**, Otsuma Womens University, Japan
Co-authors: Jose Nacher

In recent years, the availability of large amounts of financial data stored by enterprises and institutions is offering new avenues to unveil market dynamics. The large-scale financial data analysis has already led to the identification of several characteristic stylized facts such as fat tail of log return in asset prices. We perform an extensive statistical data analysis of Japan's Nikkei-255 stock index average and several forex markets and show that both markets exhibit resistance and breaking-acceleration effects at the resistance and support lines (i.e. highest price and lowest price). However, stock and forex markets display a significant difference with regard to market open hours. In stock market, there are periods where no regular trades occur between the close of the market on one day and the next day's open while the forex market is operated as 24-h. When we restrict the time series data only to the opening time and drop the intermediate time data, there are almost no resistance and breaking-acceleration effects, and volatility is always constantly high at opening time. The quantitative analysis of financial markets suggests that revisions on the existing risk management strategies in stock market may need to be considered at the opening time.

C1221: **Measuring long- and short-run connectedness of financial markets**
Presenter: **Tomas Krehlik**, Academy of Sciences of the CR, Czech Republic
Co-authors: Jozef Barunik

A methodology is provided that allows one to understand how the information on financial markets is transmitted in long- and short-run. Currently used approaches construct the connectedness measure based on the forecast error vector decomposition of generalised vector auto-regression. Using the wavelet multi-resolution analysis and vector auto-regression estimation on the decomposed series, we are able to further decompose the connectedness measure according to the persistence and thus provide important insights for the literature studying information transmission mechanisms. The measure can be interpreted on various persistence levels and in sum closely tracks the currently used connectedness measures. We demonstrate the properties of the method through simulations using two-component generalized autoregressive conditional heteroscedasticity model under various long- and short-run composition of the process. Empirical application reveals that the connectedness of US equities is mainly driven by the long-run components.

CS40 Room A2 CONTRIBUTIONS TO FORECASTING Chair: Alessandra Amendola

C170: **A forecast rationality test that allows for loss function asymmetries**
Presenter: **Andrea Naghi**, The University of Warwick, United Kingdom

A new forecast rationality test is proposed, that allows for asymmetric preferences without assuming any particular functional form for the forecaster's loss function. The construction of the test is based on the simple idea that asymmetric preferences of forecasters imply an unconditional bias of the forecast error but not a conditional bias. The null hypothesis of forecast rationality under asymmetric loss (i.e. no conditional bias) is tested by constructing a Bierens conditional moment type test. We show through Monte Carlo simulations that: i) the currently available test in the literature that accounts for asymmetries is loss function sensitive, meaning that it is not robust to specific choices of loss function and it could lead to incorrect inferences if the forecaster's underlying loss function does not belong to a specific class of loss; ii) in the presence of nonlinear dependencies between the forecast error and the information set used to generate the forecasts, our test has better finite sample properties than the currently used test in the literature. We illustrate the implementation of our test through an empirical application using data from the Survey of Professional Forecasters issued by the Federal Reserve Bank of Philadelphia.

C970: **Forecasting copper prices with dynamic averaging and selection models**
Presenter: **Daniel Buncic**, University of St Gallen, Switzerland
Co-authors: Carlo Moretto

Data from the London Metal Exchange (LME) are used to forecast monthly copper returns using the recently proposed dynamic model averaging and selection (DMA/DMS) methodology which incorporates time varying parameters as well as time varying model averaging and selection into a unifying framework. Using a total of 18 predictor variables that include traditional fundamental indicators such as excess demand and inventories as well as indicators related to global risk appetite, momentum, the term spread, and various other financial series such as exchange rates and stock prices, we show that there exists a considerable predictive component in copper returns. Covering an out-of-sample period from May 2002 to June 2014 and employing standard statistical evaluation criteria we show that the out-of-sample R^2 relative to a random walk (RW) benchmark can be as high as 18 percent with the DMA framework, and as high as 9.6 percent when using a simpler time-varying parameter model. A visual assessment of the cumulative MSFEs shows further that a substantial part of the improvement in the forecast (relative to the RW model) is realised during the peak of the financial crisis period at the end of 2008.

C997: **Why do simple average forecast combinations perform so well?**
Presenter: **Laurent Pauwels**, University of Sydney, Australia
Co-authors: Felix Chan

In the theoretical and applied forecasting literature, forecast combination techniques are increasingly used. Despite the growing number of studies employing forecast combination, there are still some questions that remain unresolved. For example, why does a simple average with equal weight often outperform complicated weighting schemes in a mean-squared forecast errors sense? The attempt is to shed some light on two questions about forecast combinations. First, it revisits why combination of multiple forecasts outperforms an individual forecast and secondly it investigates why forecast combinations based on a simple average outperforms the more complex (optimal) weighing schemes. The main contribution lies in the formulation of these two enduring questions, which permits a simple and general exposition of the forecast combination puzzle and its related issues. As a result, a set of conditions is provided for which the simple average outperforms optimal weighting schemes based on mean squares error (MSE). Related results based on other forecast criteria, such as mean absolute deviation (MAD), will also be presented.

C1252: **Income replacement ratio for various households in national pension program in Japan**
Presenter: **Tadahiko Murata**, Kansai University, Japan
Co-authors: Nisuo Du

The aim is to investigate the income replacement ratio for various households in Japanese pension program using agent-based simulation. Although the government explains the future plan of Japanese pension system using "so called" typical household models, many citizens do not belong to such typical households. Therefore we employ an agent-based simulation that models each citizen in a society to estimate future scenarios for various households. Our simulation results show that the income replacement rate will not be enough for almost all types of households.

C1245: **Forecasting Colombian inflation rate: estimation using statistical and artificial intelligence approaches**
Presenter: **Carlos Hernan Fajardo Toro**, Universidad Ean, Colombia
Co-authors: Julio Cesar Alonso Cifuentes

In Colombia the CPI (Consumer Price Index) is calculated using monthly information collected in 24 cities and metropolitan areas. For each metropolitan area a different market basket of goods is employed for three different income levels: low income, middle income and high income. The first approach corresponds to a benchmark model: a SARIMA model for the total national CPI. The second approach uses a VAR model that includes as endogenous variables country-level and 24 city-level CPIs. The third approach is a "Bottom to top" model in which the national CPI is forecasted as a linear combination of the 24 city-level CPIs. We estimate a SARIMA model for each product of the nine groups of products that is included in each market basket. The last approach, based on artificial Intelligence tools, will use different kind and topologies of neural networks to obtain a forecast with different configurations based on the nature of the variables selected and compare the results to evaluate its efficiency with each other and in regarding the methods of the previous three approaches. In our analysis we used different metrics to evaluate the out of sample behavior of the different approaches.

CS43 Room O2 CONTRIBUTIONS ON QUANTILE REGRESSION, NON/SEMI-PARAMETRIC METHODS Chair: Isabel Casas

C160: **Exploring investors' expectation through quantile regression methods**
Presenter: **Yuan Tian**, Boston University, United States

Quantile regression offers a natural and flexible framework to the statistical analysis of nonlinear response models for conditional quantile functions gradually. By minimizing asymmetrically weighted absolute residuals, quantile regressions estimate full range of conditional mean functions and provide a complete statistical analysis of the stochastic relationships among random variables. We estimate conditional quantile functions of U.S. investor expectation; then, identifying the location shift and scale shift models, we are able to infer the investment behavior changes. The proposed methodology for time series survey data analysis proves to be robust and resilient to misspecifications. Overall, the results provided clear indications of the interest of quantile regression to the analysis of survey data in financial time series.

C1020: **Semiparametric nonlinear quantile model for financial returns**
Presenter: **Krenar Avdulaj**, Institute of Economic Studies-Charles University in Prague, Czech Republic
Co-authors: Jozef Barunik

In recent years an increasing number of papers apply quantile regression models to solve problems in financial literature. The contribution to the current literature is twofold. First, we propose to use realized measures in the nonlinear quantile regression framework to explain and forecast conditional quantiles of financial returns. Second, we apply the proposed model to a pool of the most liquid U.S. assets across different industries. The nonlinear quantile regression models are implied by copula specifications and allow us to capture possible nonlinearities, and asymmetries in conditional quantiles of financial returns. Using high frequency data covering most liquid U.S. stocks in seven sectors, we provide ample evidence of asymmetric conditional dependence and different level of dependence characteristics for each industry. The backtesting results of estimated Value-at-Risk favour our approach.

C250: **Option implied risk measures: a generalized empirical likelihood approach**
Presenter: **Xiao Xiao**, Erasmus University Rotterdam, Netherlands
Co-authors: Chen Zhou

A nonparametric method to estimate the implied volatility and risk neutral distribution of stock returns from option prices is proposed. The proposed method follows the generalized empirical likelihood approach, in particular, the empirical likelihood (EL) and empirical tiltering (ET) methods. Compared to parametric methods such as the Black-Scholes (BS) model, the proposed method is free of parametric assumptions. Compared to model-free methods, the proposed method does not require various options covering a large scope of exercise prices. Instead, our method can involve multiple liquidly traded options simultaneously. Simulation studies show that the ET estimates of implied volatilities are more accurate than the BS and EL approaches, if the options have longer duration and if the underlying risk neutral distribution exhibits heavy tails and non-zero skewness. When estimating the risk neutral distribution, the ET approach outperforms the model-free methods under heavy tails and non-zero skewness. In an empirical application, we estimate implied volatility and risk neutral density from options on the S&P500 index by the proposed ET approach.

C589: **A nonparametric viewpoint on time-indexed point processes. Testing for stationarity**
Presenter: **Ana-Maria Dumitru**, University of Surrey, United Kingdom

The work is part of a larger project building a battery of tests for evolutionary point processes. Most applications using point processes assume a certain parametric model for the process, without any prior knowledge about the characteristics of the process, such as stationarity or clustering. We propose a novel test for stationarity of the time-indexed point processes. The test, with the null of stationarity, is, to the best of our knowledge, the first attempt to test for stationarity of discrete-valued stochastic processes. It relies on one of the properties of stationary point processes involving the probability generating functional. We propose nonparametric estimators for the moment measures of the point process. Consequently, we derive consistent estimators for the probability generating function under both the null and the alternative hypotheses. We support these findings with a thorough simulation exercise confirming the adequacy of the test. Moreover, we apply the new test to the trade arrival data for some S&P500 stocks.

C1095: **Nonparametric estimation and forecasting of time series with deterministic trend and season and traffic fatalities in Germany**
Presenter: **Joachim Schnurbus**, University of Passau, Germany
Co-authors: Harry Haupt
A simple procedure for simultaneous smoothing in time series regressions with deterministic components, lagged endogeneous and exogeneous components is proposed. Using monthly data on traffic fatalities in Germany the forecasting performance is compared to several commonly used forecasting methods such as ARIMA and innovation state space models. Besides an analysis of the forecasting performance, we compare the proposed methods in terms of identifying and adapting to breaks in the trend.

CS39 Room F2 CONTRIBUTIONS TO DEPENDENCE MODELING AND COPULAS Chair: Ivan Kojadinovic

C1042: **Testing the simplifying assumption in vine copulas**
Presenter: **Malte Kurz**, Ludwig-Maximilians-University Munich, Germany
Co-authors: Fabian Spanhel
Vine copulas are becoming increasingly popular among researchers to model high-dimensional dependencies in various disciplines. In almost all applications, it is assumed that the data generating process satisfies the simplifying assumption, i.e., every conditional copula in a vine collapses to an unconditional copula. Recently, it has been shown that the violation of this assumption has serious implications for the construction and interpretation of vine copula models. In particular, all frequently used stepwise procedures in vine copula modeling are disputable if the simplifying assumption does not hold and results of conditional independence tests can be invalid. Therefore, it is of crucial importance to check the adequateness of the simplifying assumption in every application of vine copulas. We develop testing procedures that can be employed to test the simplifying assumption in a sequential manner and also propose a joint test for the whole vine copula. Moreover, we introduce a graphical tool to investigate the validity of the simplifying assumption. The finite sample performance of the tests is analyzed in an extensive simulation study. Finally, we apply our tests to real data and investigate whether the simplifying assumption is reasonable in popular fields of application.

C1184: **Extreme dependence in commodity trading strategies**
Presenter: **Matthew Ames**, University College London, United Kingdom
Co-authors: Gareth Peters, Guillaume Bagnarosa
The term structures of commodity futures have long been used to derive trading strategies seeking to profit from the roll yield. In its simplest form this constitutes buying or selling a futures contract based on whether the term structure is in backwardation or contango. More sophisticated strategies involving market neutral positions and momentum, which report superior risk-return profiles, have also been explored in the literature. Typically, traders seek to combine a number of (uncorrelated) term structure trades across different commodities in order to benefit from diversification. However, this approach ignores the large drawdowns that can occur when many strategies in the portfolio have large joint losses. We decompose the dependence structure of the multivariate returns within a portfolio of commodity curve strategies. We construct a multi-factor SDE model for long term and short term spot dynamics in an exponential affine family that admits closed form expressions for the futures curves with regard to the factors. Then we calibrate this model marginally and study the portfolio returns based on copula dependence. Optimal portfolios are constructed utilising this extreme tail dependence analysis. Risk-return profiles of such portfolios are shown to be superior to portfolios constructed using correlation alone.

C1253: **Modeling VIX index based on semi parametric Markov models with Frank copula**
Presenter: **Jegors Fjodorovs**, Riga Technical University, Latvia
Co-authors: Andrejs Matvejevs
The research studies the estimation of a semi parametric stationary Markov model based on Frank copula density function. Described techniques allows us to evaluate the parameters of the Frank copula, which has the best fit to previously selected model (estimators of the marginal distribution and the copula parameters are provided). These Frank copula-based models are characterized by nonparametric marginal distributions and parametric copula functions, while the copulas capture all the scale-free temporal dependence of the processes. In our copula dependence study we used MatLab, which helped to evaluate copula parameters and choose the best copula class, based on log likelihood estimations, for the selected financial market data. We show how to apply our technique to the financial index VIX - a market mechanism that measures the 30-day forward implied volatility of the underlying index, the S&P500. Also, using this MatLab we made VIX option index study - found the best copula fit under our condition, estimated nonlinear parameters and showed an evaluation steps for copula based semi parametric models. Taking into account the developed Frank's copula model we made imitations of the VIX index as a distribution of an expected value for a certain time span.

C093: **The GARCH-copula model as a hedge ratio of corporate bonds portfolio**
Presenter: **Renata Latocha**, Meriten Investment Management, Germany
The aim is to identify the model that minimizes the basis risk between a benchmark corporate bonds portfolio as a hedged item and a standardized CDS index portfolio as a hedging instrument. A hedge ratio is based upon the conditional tail correlation coefficient estimated using the copula function and GARCH variance. The results suggest that the "GARCH-copula hedge ratio" is capable of accounting for significant heteroscedasticity characteristics found to be typical for asset returns and is more efficient and more flexible than the widely-used OLS minimum-variance hedge ratio. The identified model retains predictive power for the portfolio of bonds issued by the selected companies. The values of estimated parameters of the corporate bonds prices functions indicate that the constructed portfolio protects and grows the wealth of investors through a stream of stable and positive rates of return. This abstract represents the author's personal opinions and does not necessarily reflect the views of the company she is affiliated to.

C1199: **A copula-VAR approach for the analysis of serial dependence in stock returns**
Presenter: **Giorgia Rivieccio**, Parthenope University, Italy
Co-authors: Giovanni De Luca
The description of the dynamic behavior of multiple time series represents an important point of departure to obtain accurate forecasts both in economic and financial analysis. The aim is the comparison of methods which exploit serial dependence in stock returns to improve out-of-sample portfolio performance. For multivariate time series, the popular and easy-to-use Vector AutoRegressive (VAR) model is compared to some copula models which allow for a non-linear and/or asymmetric dependence structure among the variables. After deriving the VAR-based and copula-based conditional expected returns, we construct different portfolios and compare them in terms of Sharpe ratio.

CS109 Room P2 CONTRIBUTIONS TO THE MACROECONOMY AND ASSET PRICES Chair: Baoline Chen

C657: **International yield curves and principal components selection techniques: an empirical assessment**
Presenter: **Luca Tiozzo Pezzoli**, University Paris I - Pantheon Sorbonne, France
Co-authors: Andrew F. Siegel, Fulvio Pegoraro
Using a common database, the aim is to provide a controlled empirical comparison of recently-proposed principal component (PC) methods for selecting a combination of common and local factors that characterize the joint dynamics of multi-country term structures. We build a database of daily Treasury yield curves for U.S., Germany, U.K. and Japan, using common criteria to filter coupon bond data, to ensure liquidity, and to interpolate the discount function. We then estimate each proposed PC method for all subgroups of these countries, using both yield levels and

yield differences at weekly frequency. We find, in general, that the proposed methods do not agree with one another on the preferred combination of common and/or local factors. We identify the explained variability decision criterion as an important source of this lack of agreement and recommend consideration of alternative statistical model selection techniques for the purpose of identifying common and local yield curve factors in international data.

C808: **Earnings vs Cash flows: the valuation perspective**
Presenter: **Megha Agarwal**, University of Delhi, India

An effort is done to compare the earnings based and cash flow based methods of valuation of an enterprise. The theoretically equivalent methods based on either earnings such as Residual Earnings Model (REM), Abnormal Earnings Growth Model (AEGM), Residual Operating Income Method (ReOIM), Abnormal Operating Income Growth Model (AOIGM) and its extensions multipliers such as Price/Earnings Ratio, Price/Book Value Ratio; or cash flow based models such as Dividend Valuation Method (DVM) and Free Cash Flow method (FCFM) all provide different estimates of valuation of the Indian giant corporate Reliance India Limited (RIL). An ex-post analysis of published accounting and financial data for four financial years from 2008-09 to 2011-12 has been conducted. A comparison of these valuation estimates with the actual market capitalization of the company shows that the complex accounting based model AOIGM provides closest forecasts. These different estimates may be derived due to inconsistencies in discount rate, growth rates and the other forecasted variables. Although inputs for earnings based models may be available to the investor and analysts through published statements, precise estimation of free cash flows may be better undertaken by the internal management. The estimation of value from more stable parameters as Residual operating income and RNOA could be considered superior to the valuations from more volatile return on equity.

C305: **Empirical similarity and Taylor rule: case-based decision making in the Federal Reserve Bank**
Presenter: **Yarema Okhrin**, Universitaet Augsburg, Germany

Monetary policy is frequently described by simple instrumental rules relating the interest rate to macroeconomic variables, commonly known as Taylor rule.We propose a novel approach to the empirical analysis of central banks' monetary policy decisions. Instead of estimating a Taylor rule or a monetary reaction function, we estimate a model based on case-based decision theory. Estimating such a model does not only make a difference from an econometric perspective but also in terms of economic theory. Case-based decision theory describes reasoning by analogy. The decision maker compares the current decision situation with past ones and chooses that action that worked best in similar cases in the past. Applied to monetary policy this means that the central bank chooses an interest rate that is similar to those set in the past, when economic conditions were most similar to the present. The specific research questions we want to answer are the following. First, we ask whether an empirical similarity approach rooted in cased-based decision theory fits the federal funds rate better than conventional regression models. Second, we explore which variables affected the Fed's interest rate decisions in the past and which of these variables received the largest weight when the Fed assessed the similarity of decision cases.

C1039: **Identifying monetary policy behavior in China: a Bayesian DSGE approach**
Presenter: **Bing Li**, Tsinghua University, China
Co-authors: Qing Liu

The aim is to identify the most data favored monetary policy rule for China within a dynamic stochastic general equilibrium (DSGE) model framework. In a canonical New-Keynesian DSGE model, we carry out a positive analysis by employing Bayesian approaches to estimate two main categories of monetary policy rules, the Taylor-type interest rate rule and the money growth rate rule, respectively. The estimation is based on China's quarterly data from 1996Q2 to 2009Q4. Our empirical results show that monetary policy behavior characterized by different rules can influence the key macroeconomic variables in very different ways. By Bayes factors, we also find that the DSGE model with the money growth rate rule can capture the features of China's data better. Therefore, our empirical findings suggest that the money growth rate rule is more favored by China's data and provides a better abstraction of the sophisticated monetary policy behavior in China.

C1239: **Valuing macroeconomic uncertainty in bond risk premia**
Presenter: **Kameliya Filipova**, University of St Gallen, Switzerland

A new framework is proposed for estimating and studying the macroeconomic determinants of bond risk premia. In contrast to most of the previous studies, which use the uncertainty coming from the macroeconomic volatility as the only source of premia variation, in our model the private sector confronts uncertainty also about the states of the economy. These states are determined endogenously and modeled by means of thresholds, which are directly linked to macroeconomic fundamentals. The private sector observe the state of the economy today and only based on the current relevant macroeconomic information and the optimal threshold structure, it infers the future state. Applying our model to US interest rate data we find pronounced nonlinear and state-dependent relationship between bond risk premia and the macroeconomy. Motivated by the empirical results, we incorporate the nonlinear bond risk premia into a term structure model suitable for pricing and study the yield curve's implications.

CS81 Room Q2 CONTRIBUTIONS TO QUANTITATIVE RISK MANAGEMENT Chair: Marco Bee

C942: **Option pricing using the continuous hidden threshold mixed skew-symmetric distribution**
Presenter: **Rachid Belhachemi**, Xian Jiaotong-Liverpool University, China
Co-authors: Pierre Rostan

While asymmetric mixture models improve on the option pricing error when compared to other pricing models, mispricing remains due to their lack of capturing the effect of the economic factors on prices level. The aim is to use continuous hidden threshold mixed normal models (HTSN) to price options.Compared to the discrete state models, our model has extra parameters that can be linked to economic dynamics and have economic interpretations. The model allows us to capture some stylized fact underlying option prices such as asymmetry and extreme realizations. The model can be estimated straightforwardly by maximum likelihood. We show that HTSN distribution accurately prices index options and captures adequately the smirk of implied volatility. The model also allows for a time-varying price of risk and non-normal innovations.

C1182: **On backtesting risk measurement models**
Presenter: **Hideatsu Tsukahara**, Seijo University, Japan

The focus is on the 'reality check' aspect of backtesting: monitoring the performance of a model for risk measurement. It is a tool for the model validation process which is indispensable for adequate financial risk management. In the case of Value-at-Risk (VaR), a popular procedure for backtesting depends on the number of VaR violations. Many people claim that it is easier to backtest VaR than Expected Shortfall (ES) and other risk measures because it can be backtested in a distribution-free manner with small samples. Now the concept called "elicitability" has recently attracted much attention. Roughly speaking, a statistical functional is called elicitable if it is a unique minimizer of some expected loss. While VaR is easily seen to be elicitable, it has been proved that ES and the distortion risk measures fail to satisfy this condition. While elicitability is useful for comparing and ranking models, there seems to be no clear connection with monitoring the performance of a model. We will illustrate this with simple examples. Moreover, we examine the concept of consistency of a statistical functional, proposed by Mark Davis, from the data-analytic point of view. Finally, we suggest some backtesting procedures for distortion risk measures, and check its effectiveness in a simulation study using ES and also proportional odds distortion risk measure.

C1107: **Density approximations and VaR computation for compound Poisson-lognormal distributions**
Presenter: **Marco Bee**, University of Trento, Italy

The Value at Risk (VaR) for the compound Poisson-lognormal distribution is usually computed via Crude Monte Carlo (CMC). However, as the coverage level increases, the relative error of this estimator grows exponentially. To overcome this difficulty, parametric approximations of the distribution are developed and used to compute VaR. As guidelines, two requirements are considered: the skewness-kurtosis space and the tail behavior. According to these criteria, the Generalized Beta distribution of the second kind (GB2) and a mixture of two lognormals provide a good fit. In certain cases the GB2 can be estimated by moment-matching, thus providing a simulation-free procedure for VaR computation. For coverage levels larger than 99%, Extreme Value Theory approaches based on the Maximum Likelihood and Probability Weighted Moments (PWM) estimation methods are proposed. Extensive simulation experiments find that the approximations guarantee non-negligible efficiency gains with respect to CMC. In particular, when the variance of the underlying lognormal is smaller than a certain threshold, the performance of the GB2 VaR estimated via moment-matching turns out to be excellent. For higher quantiles, the EVT-PWM VaR is significantly better, in terms of MSE, than both MLE-EVT VaR and CMC VaR.

C974: **Estimating risk-neutral density tails: a comparison**
Presenter: **Pirmin Meier**, University of St Gallen, Switzerland
Co-authors: Anselm Ivanovas

Having accurate estimates of the risk-neutral density tails is particularly important for risk management applications or to estimate higher moments of the risk-neutral distribution. Interestingly, this point has been addressed only in the very recent literatura. We review several recently introduced methods to estimate risk-neutral density tails from option data. In addition, we propose a new tail extension approach which combines the idea of price matching with the assumption of tails drawn from a generalized extreme value distribution. Based on a theoretical market model with known implied risk-neutral density, we conduct a performance analysis. The final outcomes of our study indicate that the best results regarding accuracy and stability are obtained either with implied volatility extrapolation by means of Gatheral's stochastic volatility inspired SVI model, or with our new price matching method. Moreover, we provide evidence that the still very popular constant implied volatility extrapolation approach often yields unsatisfactory results.

C1011: **Backtesting portfolio Value-at-Risk with estimated portfolio weights**
Presenter: **Pei Pei**, Central University of Finance and Economics, China

The aim is to theoretically and empirically analyze backtesting portfolio VaR with estimation risk in an intrinsically multivariate framework. For the first time in the literature, it takes into account the estimation of portfolio weights in forecasting portfolio VaR and its impact on backtesting. It shows that the estimation risk from estimating the portfolio weights as well as that from estimating the multivariate dynamic model of asset returns make the existing methods in a univariate framework inapplicable. And it proposes a general theory to quantify estimation risk applicable to the present problem and suggests practitioners a simple but effective way to carry out valid inference to overcome the effect of estimation risk in backtesting portfolio VaR. A simulation exercise illustrates our theoretical findings. In an application, a portfolio of three stocks is considered.

CS36 Room I2 CONTRIBUTIONS TO MACRO AND FORECASTING Chair: Anindya Banerjee

C980: **Reduction of forecast errors**
Presenter: **Till Weigt**, University of Muenster, Germany

Starting from k different forecasting models a new forecasting technique is developed. The basic idea is that, although the information contained in the past forecast losses of any one model cannot be used separately to improve systematically the future forecasts of that model, the correlations between the past forecast losses of all k models can be used to improve systematically the future forecasts of every model. Technically, a VAR(p) model is employed in which the current forecast loss of each model is regressed on the past forecast losses of all models under consideration simultaneously. The resulting forecasts of this VAR(p) model represent the future forecast losses which are then subtracted from the genuine future forecasts. It is shown that the expectation of the revised future forecast losses is smaller than the expectation of the genuine forecast losses asymptotically. Monte Carlo simulations confirm the asymptotic result for finite samples. An empirical study of different volatility models underpins the model's applicability out-of-sample. This new forecasting technique represents a methodically novel approach which belongs to the class of forecast combination models. However, conventional combination approaches calculate optimal combination weights to generate one superior forecast, while this approach improves each forecast on its own.

C1047: **Short-term forecasting with mixed-frequency data: a MIDASSO approach**
Presenter: **Boriss Siliverstovs**, KOF ETHZ, Switzerland

The targeted-regressor approach is extended for variables sampled at the same frequency to mixed-frequency data. Our MIDASSO approach is a combination of the unrestricted MIxed-frequency DAta-Sampling approach (U-MIDAS) and the LASSO-typepenalised regression called the elastic net. We illustrate our approach by using empirical example with the Purchasing Managers' Index computed for Switzerland. We address whether the fixed weighting scheme of the PMI components is supported by the data. We find that the relative weights of the PMI components are generally supported by the data, except for the fact that one component, found very informative for explaining GDP growth, is currently omitted from the PMI composition. We also compare the out-of-sample forecasting performance of the MIDASSO approach of GDP growth in Switzerland based on pre-selection of predictors with that of the approach without such variable pre-selection. The results of the forecasting exercise conform to previous results, reported for single-frequency data, that targeting most relevant predictors boosts the forecasting performance also when applied to mixed-frequency data.

C1165: **The conditional predictive ability of economic variables**
Presenter: **Eleonora Granziera**, Bank of Canada, Canada
Co-authors: Tatevik Sekhposyan

The relative performances of forecasting models are known to be rather unstable over time. However, it is not very well understood why the forecasting performance of economic models changes. We propose to address this question by evaluating the predictive ability of a wide range of economic variables for key U.S. macroeconomic aggregates: output growth and inflation. We take a conditional view on this issue, attempting to identify situations where particular kind of models perform better than simple benchmarks. Furthermore, we investigate whether using the conditioning information as a model selection criterion for model averaging could improve the accuracy of the predictions.

C094: **Forecasting using many predictors with neural network factor models**
Presenter: **Ali Habibnia**, LSE, United Kingdom

Modelling and forecasting financial returns aids understanding of market dynamics, but challenges include non-linearity, non-gaussianity, and comovenet of stock returns. A non-linear forecasting technique is proposed based on an improved factor model with two neural network extensions, which would be able to capture both non-linearity and non-normality of a high-dimensional dataset. This proposed model has been developed on the basis of statistical inference, and special emphasis is given to data-driven specification. It has been proved that a linear factor model is a special case of this neural network factor model.

C771: **Forecasting with a parsimonious subset VAR model**
Presenter: **Hyunchul Lee**, Chosun University, Korea, South
Co-authors: Chongcheul Cheong, Prashant Mool

The aim is to suggest using a unit t-value criterion in imposing restrictions on lags to formulate a subset vector autoregressive (VAR) model for the purpose of point forecasts. Among any other alternative models nested to the initial VAR model, this less restrictive modeling strategy produces the smallest log determinant of the residual covariance matrix adjusted by degrees of freedom. Each equation of the finally derived subset VAR model has a maximized adjusted by degrees of freedom in samples and consequently a minimized 1-step-ahead prediction error in out-of-samples. The applicability of this modeling strategy is excised to the case of a bivariate VAR model for output growth and inflation.

Monday 8.12.2014 10:35 - 12:15 Parallel Session P – ERCIM

ESI02 Room Sala Convegni STATISTICS OF EXTREMES AND APPLICATIONS Chair: Armelle Guillou

E150: **Multivariate peaks over thresholds modelling**
Presenter: **Holger Rootzen**, Chalmers, Sweden

Quite a number of different approaches to multivariate peaks over thresholds modelling has been proposed in the literature. Most of them are based on multivariate generalizations of the generalized Pareto distribution, but other approaches have also been used. The aim is to review and discuss some of these approaches. Further, work in progress on new parametric multivariate generalized Pareto models will be presented. These models have tractable likelihoods and permit use of the entire standard maximum likelihood machinery for estimation, testing, and model checking.

E257: **The extremogram: a measure of extremal dependence for univariate and multivariate time series**
Presenter: **Richard Davis**, Columbia University, United States
Co-authors: Thomas Mikosch, Ivor Cribben, Yuwei Zhao

The extremogram is a flexible quantitative tool that measures various types of extremal dependence in a stationary time series. In many respects, the extremogram can be viewed as an extreme-value analogue of the autocorrelation function (ACF) for a time series that can be used in various phases of model fitting and confirmation. We discuss estimation of the extremogram and demonstrate its use applied to financial time series models for both univariate and multivariate cases.

E410: **Max-stable processes on river networks**
Presenter: **Anthony Davison**, EPFL, Switzerland
Co-authors: Peiman Asadi, Sebastian Engelke

Max-stable processes provide natural probability models for extremes in spatial settings, and are increasingly widely used in applications. In some settings, however, the natural geometry is not Euclidean, and then standard max-stable models must be adapted. We describe the use of max-stable processes for modelling extreme floods in river networks, where dependence has two components, one corresponding to flow along the river and the other to rainfall, which may affect different sub-catchments simultaneously. The ideas are illustrated by an application to the Danube.

ES16 Room Q1 BAYESIAN SEMIPARAMETRIC INFERENCE Chair: Taeryon Choi

E103: **Bayesian asymptotics under the sup-L_1 distance in nonparametric conditional density estimation**
Presenter: **Pierpaolo De Blasi**, University of Torino and Collegio Carlo Alberto, Italy
Co-authors: Stephen Walker

Posterior consistency in conditional density estimation problems with respect to the sup-L_1 metric is studied. Currently, only the much weaker integrated L_1 metrics are used. We allow, for example, consistency for prediction at any designated conditional density. The prior we consider specifies a type of dependence between conditional densities which disappears after a certain number of observations have been observed. This is obtained through a sequence of increasingly finer partitions of the predictor space and by letting the conditional density be the same across all predictor values in a partition set and modeled independently by a nonparametric local prior. The rate at which the number of partition sets increases with the sample size determines when the dependence between pairs of conditional densities is set to zero and, ultimately, drives posterior consistency. We also provide some results on what type of convergence rates the posterior can achieve.

E478: **Shape constrained regression using Gaussian process projections**
Presenter: **Lizhen Lin**, University of Texas-Austin, United States
Co-authors: David Dunson

Shape constrained regression analysis has applications in dose-response modeling, environmental risk assessment, disease screening and many other areas. Incorporating the shape constraints can improve estimation efficiency and stability and avoid implausible results. We talk about nonparametric methods for estimating shape constrained (mainly monotone constrained) regression functions. We mainly talk about a novel Bayesian method focusing on monotone curve and surface estimation using Gaussian process projections. Inference is based on projecting posterior samples from the Gaussian process. Theory is developed on continuity of the projection and rates of contraction. Our approach leads to simple computation with good performance in finite samples. The projection approach can be applied to other constrained function estimation problems including in multivariate settings.

E518: **Nonlinear mixed-effects GP functional regression models with applications to motion data**
Presenter: **Jian Qing Shi**, Newcastle University, United Kingdom
Co-authors: Yafeng Cheng, Javier Serradilla, J. A. Eyre

The aim is to discuss a new semiparametric mixed-effects model for functional regression analysis with scalar response and mixed scalar and functional covariates, combining a parametric functional regression model with a nonparametric Gaussian process regression (GPR) model. The parametric component can provide explanatory information between the response and the covariates, while the nonparametric component can model the nonlinearity flexibly. The model is applied to a medical research project carried out in Newcastle University (UK) on assessing upper limb function for patients after stroke using motion data generated automatically during play of a bespoke, professionally-written action video game (Circus Challenge Assessment Game, CCAG). This is the first time a serious game can achieve automatic in-game assessment to a clinical standard, solely using low cost, commodity hardware and professionally written action video games and demonstrates the potential for remote monitoring of patients during home based rehabilitation programs.

E802: **The semiparametric Bernstein-von Mises theorem for location model with Dirichlet process mixture of normal densities**
Presenter: **Yongdai Kim**, Seoul National University, Korea, South
Co-authors: Minwoo Chae, Bas J. K. Kleijn

The Bernstein-von Mises theorem for symmetric location model is proved. For the prior of the location parameter and symmetric error density, the product prior is used while the symmetric error density is endowed with a Dirichlet process mixture prior of normal densities which is popularly used in nonparametric Bayesian density estimation problems. It is shown that if the nuisance parameter is integrated out from the likelihood with respect to the prior, the resulting integrated likelihood satisfies Le Cam's local asymptotic normality. Furthermore, the marginal posterior of the location parameter is shown to be consistent at parametric rate. These two results enable us to extend the proof of parametric Bernstein-von Mises theorem and yield the desired result.

ES43 Room E1 STATISTICAL ANALYSIS OF FMRI DATA Chair: Michele Guindani

E164: **A spatially adaptive CAR model with application to pre-surgical fMRI data**
Presenter: **Timothy Johnson**, University of Michigan, United States
Co-authors: Zhuqing Liu, Veronica Berrocal

Spatial smoothing is an essential step in the analysis of functional magnetic resonance imaging (fMRI). The standard method is to convolve the image data with an isotropic Gaussian kernel that applies a fixed amount of smoothing to the entire image. In pre-surgical brain image analysis,

where spatial accuracy is paramount, this is not a reasonable method – over smoothing some areas and under smoothing others. As a result, boundaries between activated and deactivated regions of the brain are blurred. To circumvent this problem we propose a novel, spatially adaptive, conditional autoregressive model that allows the amount of smoothing to vary across the brain, driven by the data itself. We compare our proposed model with two existing spatially adaptive models and with a Bayesian non-parametric Potts model that we previously proposed. Simulation studies show our model performs on par with the Bayesian non-parametric Potts model, yet is much more computationally efficient, and performs better than the other models. We demonstrate our model on a pre-surgical fMRI data set.

E452: **Time varying connectivity models for brain imaging data**
Presenter: **Ivor Cribben**, Alberta School of Business, Canada
Co-authors: Yi Yu

Recently, the use of graphical models for estimating functional connectivity or brain networks has become increasingly popular in the field of neuroscience. In most functional magnetic resonance imaging (fMRI) studies, the networks between brain regions are assumed to be stationary over time. However, there is now more evidence that the network or functional connectivity is changing over time even when the subjects are at rest. We introduce a novel method that dynamically clusters brain regions or voxels by their functional connectivity. This new method allows for situations where the number of brain regions is greater than the number of time points in the experimental time course ($n < p$). This method promises to offer deeper insight into the inner workings of the brain. Moreover, we propose a new nonparametric method for estimating time varying graphical structures, or networks between brain regions, that allows for smoothly changing graphs over time. We apply both new methods to simulated data and to fMRI data sets.

E615: **The graphical modeling and longitudinal analysis of fMRI data**
Presenter: **Mark Fiecas**, University of Warwick, United Kingdom
Co-authors: Ivor Cribben

Functional MRI (fMRI) studies could provide crucial information on the effects of rehabilitation on the neural mechanisms in patients with cerebral palsy. Using longitudinal fMRI acquired from patients and controls, we investigate how the functional network of the brain evolves over time. Existing statistical methods have been lacking in the analysis of brain connectivity analysis using longitudinal fMRI data. We propose a linear mixed model using an l1 penalization for estimating a longitudinal network from fMRI data. The mixed model approach accounts for individual variability and within-individual covariability in the connectivity parameters, and the l1 penalty will account for the high-dimensionality of the data, which occurs due to the large number of brain regions considered in the analysis. We illustrate the utility of our method for estimating an evolving network structure across time for an fMRI experiment that includes both phonation and speech phonation tasks for patients and controls.

E602: **Tracking whole-brain connectivity dynamics in the resting state**
Presenter: **Erik Erhardt**, University of New Mexico, United States
Co-authors: Elena Allen, Eswar Damaraju, Sergey Plis, Tom Eichele, Vince Calhoun

Spontaneous fluctuations are a hallmark of recordings of neural signals, emergent over time scales spanning milliseconds and tens of minutes. However, investigations of intrinsic brain organization based on resting-state functional magnetic resonance imaging have largely not taken into account the presence and potential of temporal variability, as most current approaches to examine functional connectivity (FC) implicitly assume that relationships are constant throughout the length of the recording. We describe an approach to assess whole-brain FC dynamics based on spatial independent component analysis, sliding time window correlation, and k-means clustering of windowed correlation matrices. The method is applied to resting-state data from a large sample ($n = 405$) of young adults. Our analysis of FC variability highlights particularly flexible connections between regions in lateral parietal and cingulate cortex, and argues against a labeling scheme where such regions are treated as separate and antagonistic entities. Additionally, clustering analysis reveals unanticipated FC states that in part diverge strongly from stationary connectivity patterns and challenge current descriptions of interactions between large-scale networks. Overall, we suggest that the study of timevarying aspects of FC can unveil flexibility in the functional coordination between different neural systems.

ES59 Room D1 COMPUTATIONAL APPROACHES IN FINANCIAL ECONOMICS Chair: Dietmar Maringer

E564: **Understanding banking networks from optimal portfolio choices**
Presenter: **Philipp Johannes Kremer**, EBS Universitaet fuer Wirtschaft und Recht, Germany
Co-authors: Ben Craig, Marcel Gorenflo, Sandra Paterlini

Network analysis can provide useful insights on the degrees of interconnectedness, contagion and spillover effects of banks, firms and the economic systems as a whole. Using data from securities holding statistics of the German central bank, we aim to analyze how the interbank network evolves in time with respect to alternative optimal network configurations, which could be considered as the socially optimal network structure with respect to different indicators, such as loan diversification, riskiness, interconnectedness, and liquidity and information contagion. As data contain noise, to detect relevant and robust signals we combine methods from network analysis with regularization approaches.

E565: **Credit risk dependence in the European corporate sector**
Presenter: **Wenwei Li**, EBS Business School, Germany
Co-authors: Ulrich Hommel, Sandra Paterlini

Recent financial crises have emphasized the important role of the linkages among different market participants as a channel for shock transmission. We aim to capture the credit risk dependence network of European public firms, especially under extreme market conditions. By means of clustering analysis and tail dependence measures, we can estimate the network configuration and its evolution in time of a large pool of European companies. The result is meaningful for both regulators and practitioners, as it reveals the dependence structure of firms' credit risk from a systematic view, and helps us understand systemic risk transmission mechanism during financial crises.

E447: **Optimal diversification in portfolio selection**
Presenter: **Margherita Giuzio**, EBS Universitaet fur Wirtschaft und Recht, Germany
Co-authors: Sandra Paterlini

One of the main challenges in portfolio selection is to find a vector of asset weights able to optimize the diversification while achieving optimal risk-return trade-off. We introduce a constraint on the level of diversification by means of an ℓ_q norm on its weights in a mean-variance framework. We compare different estimation techniques and show the empirical results of our approach on both simulated and real-world settings.

E937: **Order placement in a continuous double auction agent based model**
Presenter: **Alexandru Mandes**, Justus Liebig University Giessen, Germany

Current implementations of intraday financial markets by means of agent based models rely only on stochastic placement strategies. In real markets, trade size and timing are not random, but rather take into account the existing market liquidity. We have modeled the agent's order placement decision as an optimisation problem which minimizes the risk adjusted execution cost, taking into consideration various micro-structure factors, as well as intrinsic agents characteristics. Next, we have integrated the order submission model into a zero-intelligence agent based model. Thus, we were able to assess the impact of replacing the original random order placement by the micro-trading strategy, with respect to two high-frequency stylised facts. Our model has successfully reproduced the power-law tail of the relative price distribution of off-spread limit orders, even if there is no explicit power-law component assumed in the agents' design. Regarding market orders, both the binned-average price impact as well as the

individual price impact functions exhibit a concave shape – a trace of rational selective trading. On the opposite, in the absence of any intelligent trading decision, the expected market price impact shape is convex – confirmed by the results obtained with the alternative zero-intelligence agent based model.

ES71 Room M1 APPLICATIONS OF FUNCTIONAL DATA ANALYSIS Chair: Alicia Nieto-Reyes

E607: **Functional data analysis and neurodevelopmental brain mapping**
Presenter: **Philip Reiss**, New York University, United States
Co-authors: Lei Huang, Huaihou Chen, Thaddeus Tarpey

Magnetic resonance imaging of the human brain is an increasingly important field of application for functional data analysis (FDA). We focus on using brain imaging data to infer so-called neurodevelopmental trajectories: for example, how a quantity such as fractional anisotropy (a measure of white matter integrity) varies with age, at each point in the brain. This is an important class of problems in neuroscience, for which the conventional analytic framework treats each brain location separately. Alternatively, from an FDA perspective, one might view the problem as function-on-scalar regression, with age as the scalar predictor of interest, and a sequence of brain locations (or more ambitiously, the entire brain) as the domain of the functional responses. Neither paradigm leads to a fully satisfactory solution, but a new approach synthesizing aspects of both is shown to produce improved results with high computational efficiency. We also demonstrate how an FDA variant of k-means clustering offers a novel way to visualize the results in the form of a brain map.

E640: **A functional mixed model for 3D-acceleration signals of trotting horses**
Presenter: **Anders Tolver**, University of Copenhagen, Denmark
Co-authors: Helle Sorensen, Maj Halling Thomsen

Evaluation of the movement during trot in circles is commonly used in clinical lameness examinations of horses. Lameness as well as the size of the circle is known to affect the symmetry of the movement. Functional observations of accelerations were obtained for a group of sound horses during trot on a straight line and on circles of varying diameter in both directions. In order to describe the different sources of variation in the acceleration signals both the horse-to-horse variation and the possibly non-linear impact of the diameter of the circle must be taken into account. Further, the physical nature of the biomechanical system for quadropedal trot imposes many restrictions on the class of possible idealized acceleration signals that must be built into any reasonable statistical model for the data. Some practical questions of particular interest are: i) do horses have a preferred direction for trot in circles that is reflected in the symmetry of the acceleration signals, ii) at what circle size is the movement comparable to trot on a straight line?

E757: **A functional mixed model approach to house prince index modelling**
Presenter: **Juhyun Park**, Lancaster University, United Kingdom
Co-authors: Clement Lee

Partially observed sparsely measured functional data are considered where some of the individual characteristics are missing. A motivating example comes from modelling house price. House price index can be interpreted as an expected price change for a given period. One of the main difficulties in constructing such index stems from the fact that only limited information is available: each address or zip code is recorded, but not individual characteristics. To overcome this difficulty, a standard approach known as repeat sales regression estimates the index based on the difference of the price pairs that have been sold repeatedly. Under the assumption that the individual characteristics are time-independent fixed effects, the time-varying effect of the house price is estimated as the coefficient of the dummy variables of the time period based on weighted least squares method. Several variations of this approach have been proposed to deal with issues related to heteroscedastic errors, temporal or spatial dependence, but mostly these are parametric in nature. We propose a functional extension of the repeat sales regression model to account for smoothly varying time effect as well as temporal or spatial dependence. We use basis function representation of the smooth function components, but apply mixed model estimation method to the differenced data. We use the land registry sales data for England and Wales to demonstrate our methodology.

E836: **Dynamics of DNA minicircles in motion via Fourier analysis of functional time series**
Presenter: **Shahin Tavakoli**, EPFL, Switzerland
Co-authors: Victor Panaretos

The problem of studying the dynamics of DNA minicircles that are vibrating in solution is considered. At a large scale, DNA minicircles are modelled as elastic rods, and the problem of understanding their dynamics can be recasted into the problem of estimating the second order structure of a stationary functional time series (FTS). We tackle this problem by a frequency domain approach, where we estimate the spectral density operators (or spectra) of the DNA minicircle. We then carry out hypothesis tests to compare the spectra of two specific DNA minicircles, and localise their differences both in frequencies and on the DNA minicircles.

ES72 Room B1 BIND SOURCE SEPARATION: STATISTICAL PRINCIPLES Chair: Hannu Oja

E596: **Blind source separation for spatial compositional data**
Presenter: **Klaus Nordhausen**, University of Turku, Finland
Co-authors: Hannu Oja, Peter Filzmoser, Clemens Reimann

In regional geochemistry rock, sediment, soil, plant or water samples collected in a certain region are analysed for concentrations of chemical elements. The observations are thus usually high dimensional, spatially dependent and of compositional nature. A novel blind source separation approach for spatially dependent data is suggested. For the analysis, it is assumed that the multivariate observations are linear combinations or mixtures of latent components and that the spatial processes for these latent components are second order stationary and uncorrelated. In the presented approach the components can then be recovered by simultaneously diagonalizing the covariance matrix and a local covariance matrix. This method can be easily applied also in the context of compositional data after appropriate data transformations. The components obtained in this analysis are then uncorrelated and easily interpretable, and can be used for dimension reduction and for visual presentation of different features of the data.

E712: **Independent component analysis via distance covariance**
Presenter: **David Matteson**, Cornell University, United States
Co-authors: Ruey S. Tsay

A novel statistical framework for independent component analysis (ICA) of multivariate data is introduced. We propose a methodology for estimating and testing the existence of mutually independent components which are a linear transformation of a given dataset, and a versatile resampling-based procedure for inference. Independent components are estimated by combining a nonparametric probability integral transformation with a generalized nonparametric whitening method that simultaneously minimizes all forms of dependence among the components. U-statistics of certain Euclidean distances between sample elements are combined in succession to construct a statistic for testing the existence of mutually independent components. The proposed measures and tests are based on both necessary and sufficient conditions for mutual independence. When independent components exist, one may apply univariate analysis to study or model each component separately. Univariate models may then be combined to obtain a multivariate model for the original observations. We prove the consistency of our estimator under minimal regularity conditions without assuming the existence of independent components a priori, and all assumptions are placed on the observations directly, not on the latent components. Additional theoretical results are provided under additional assumptions. We demonstrate the improvements of the proposed method over

competing methods in simulation studies. We apply the proposed ICA approach to two real examples and contrast it with principal component analysis.

E849: **Separation of uncorrelated stationary time series using autocovariance matrices**
Presenter: **Sara Taskinen**, University of Jyvaskyla, Finland
Co-authors: Jari Miettinen, Klaus Nordhausen, Hannu Oja

It is assumed that the observed p time series are linear combinations of p latent uncorrelated weakly stationary time series. The aim is then to find an estimate for an unmixing matrix, which transforms the observed time series back to uncorrelated latent time series. In the so-called SOBI method (Second Order Blind Identification), joint diagonalization of the covariance matrix and autocovariance matrices with several lags is used to estimate the unmixing matrix. The rows of an unmixing matrix can be derived either one by one (deflation-based approach) or simultaneously (symmetric approach). The statistical properties of these two SOBI estimates are investigated. We give the limiting distributions under general conditions and compare asymptotical efficiencies under general multivariate linear processes. The theory is illustrated by some finite-simple simulation studies as well as a real EEG data example.

E1131: **Deflation-based FastICA with adaptive choices of nonlinearities**
Presenter: **Jari Miettinen**, University of Jyvaskyla, Finland
Co-authors: Klaus Nordhausen, Hannu Oja, Sara Taskinen

Deflation-based FastICA is a popular method for independent component analysis. In the standard deflation-based approach the row vectors of the unmixing matrix are extracted one after another always using the same nonlinearity. In practice the user has to choose the nonlinearity and the efficiency and robustness of the estimation procedure then strongly depends on this choice as well as on the order in which the components are extracted. We propose a novel adaptive two-stage deflation-based FastICA algorithm that (i) allows one to use different nonlinearities for different components and (ii) optimizes the order in which the components are extracted. Based on a consistent preliminary unmixing matrix estimate and our theoretical results, the algorithm selects in an optimal way the order and the nonlinearities for each component from a finite set of candidates specified by the user. The resulting ICA estimate is affine equivariant with a known asymptotic distribution. The excellent performance of the new procedure is shown with asymptotic efficiency and finite-sample simulation studies.

ES73 Room P1 STATISTICAL MODELS FOR THE HEALTH CARE ASSESSMENT Chair: Anna Paganoni

E441: **Study of heart failure hospitalizations using administrative data - first results from HFdata project**
Presenter: **Giulia Garavaglia**, Politecnico di Milano, Italy
Co-authors: Cristina Mazzali

In the "HFdata Project" administrative databases of the Lombardy region (Italy; 9 millions of inhabitants) were used to study patients hospitalized for heart failure (HF) from an epidemiological and economical point of view. Data on hospitalization were checked for completeness, correctness, internal and external validity. Thanks to and together with clinicians, observation units and criteria for patients' selection were defined, an algorithm for co-morbidities detection was implemented and clinically meaningful subgroups of patients were defined on the basis of diagnoses. First results concern epidemiological analysis on incidence of HF, characteristics of patients and their outcomes, subgroups description, general and specific resource burden. Additional analysis on cost drivers of in-hospital pathways have been performed in order to study trends, compositions and disease evolution.

E795: **Exploiting the use of large administrative databases in epidemiology: multi-state models for times to hospitalizations and death in Heart Failure**
Presenter: **Francesca Ieva**, Universita degli Studi di Milano, Italy
Co-authors: Christopher Jackson, Linda Sharples

To improve the understanding of the prognosis of patients affected by Heart Failure (HF) and to enable health-care providers to assess and manage resources, the use of large administrative databases for epidemiological purposes is exploited. In particular, we wish to jointly model disease progression, mortality and their relation with patient characteristics in the context of HF hospitalization process. In chronic diseases like HF, the disease course and associated clinical event histories for the patient population vary widely. In order to estimate some important quantities that enable prediction of healthcare resources consumption, it is important to exploit the information carried out by the patients' clinical history. We show how episodes of hospitalisation for disease-related events, obtained from administrative data, can be used as a surrogate for disease status. We propose flexible multi-state models for serial hospital admissions and death in HF patients, that are able to accommodate important features of disease progression. The models were applied to a dataset from the administrative data bank of the Lombardia region in Northern Italy.

E626: **Multivariate permutation test to compare survival curves for matched data**
Presenter: **Stefania Galimberti**, University of Milano-Bicocca, Italy
Co-authors: Maria Grazia Valsecchi

In the absence of randomization, the comparison of an experimental treatment with the standard may be done based on a matched design. When there is a limited set of cases receiving the experimental treatment, matching of a proper set of controls in a non-fixed proportion is convenient. In order to deal with the highly stratified survival data generated by multiple matching, we extend the multivariate permutation testing approach, since standard nonparametric methods are inefficient when the number of strata increases and strata are small and have low power in the absence of proportional hazards. The aim is to provide a global test for the comparison of two survival curves for highly stratified matched data by combining a finite number of dependent permutation tests performed at different time-points. This approach combines nonparametrically the partial tests that are based on the survival differences estimated at various time-points. Conditions for the applicability of this approach are very mild. We demonstrate the validity of the proposed method with simulations, and we illustrate its application to a motivating example derived from an observational study for the comparison of bone marrow transplantation and chemotherapy in the treatment of paediatric leukaemia.

E753: **Modeling medical malpractice claims: an example from Italy**
Presenter: **Marco Bonetti**, Bocconi University, Italy
Co-authors: Pasquale Cirillo

Medical malpractice involves patient injury or death attributed to negligent behavior by a medical practitioner. Modeling claims due to medical malpractice thus becomes very important from a health service, legal and regulatory, and an insurance point of view. A better understanding of such a phenomenon can have positive effects for hospitals and clinics in terms of quality assurance, service improvement, and cost reduction. On the other hand, it is helpful to insurance companies to allow for a reliable pricing of the policies, and to allow the implementation of an overall better risk management approach to loss. We describe a large analysis of medical malpractice in Italy that involves statistical models both for the number of claims and for their monetary amounts. We apply a Poisson process regression model to describe the number of events over time for different regions and medical departments. We then model the associated reported claim amounts with a mixed conditional model that separately handles the zero amounts and the positive amounts, also through regression models. We describe the methods and provide some of the main results that have been obtained, including point estimates, back-testing results, and forecasts.

ES64 Room G2 THEORETICAL ECONOMETRIC ADVANCES IN FINANCIAL RISK MEASUREMENT Chair: Melanie Schienle

E440: **Estimating the spot covariation of asset prices - statistical theory and empirical evidence**
Presenter: **Peter Malec**, Humboldt University Berlin, Germany
Co-authors: Markus Bibinger, Nikolaus Hautsch, Markus Reiss

A new type of estimator of the spot covariance matrix of a multidimensional semi-martingale asset price process which is subject to noise is proposed. The estimator is constructed based on a local average of block-wise constant spot covariance estimates. The latter originate from the local method of moments (LMM) building on locally constant approximations of the underlying process. We extend the LMM estimator to allow for autocorrelated noise and propose a consistent estimator of the order of serial dependence. We prove the consistency and asymptotic normality of the proposed spot covariance estimator and show that it benefits from the rate-optimality of the underlying LMM approach. Based on detailed simulations, we provide empirical guidance on the optimal implementation of the estimator and apply it to high-frequency data based on a cross-section of NASDAQ blue chip stocks. Employing the estimator to estimate spot covariances, correlations and betas in normal but also extreme-event periods yields novel insights into intraday covariance and correlation dynamics. We show that intraday (co-)variations (i) follow underlying periodicity patterns, (ii) reveal substantial intraday variability associated with (co-)variation risk, (iii) are strongly serially correlated, and (iv) can increase strongly and nearly instantaneously if new information arrives.

E483: **Anticipating long-term stock market volatility**
Presenter: **Christian Conrad**, Heidelberg University, Germany
Co-authors: Karin Loch

The relationship between long-term U.S. stock market risks and the macroeconomic environment is investigated using a two component GARCH-MIDAS model. Our results show that macroeconomic variables are important determinants of the secular component of stock market volatility. Among the various macro-variables in our dataset the term spread, housing starts, corporate profits, and the unemployment rate have the highest predictive ability for long-term stock market volatility. While the term spread and housing starts are leading variables with respect to stock market volatility, for industrial production and the unemployment rate expectations data from the Survey of Professional Forecasters regarding the future development are most informative.

E590: **Estimation of expected utility portfolios for large dimensional data**
Presenter: **Nestor Parolya**, Leibniz University Hannover, Germany
Co-authors: Taras Bodnar, Yarema Okhrin

The mean-variance (MV) portfolio in the high-dimensional case is estimated using the results from the random matrix theory. We construct a shrinkage-type estimator which is distribution-free and it is optimal in the sense of maximizing the expected utility, i.e. mean-variance objective function. Its asymptotic properties are investigated when the number of assets p depends of the sample size n such that $\frac{p}{n} \to c \in (0, +\infty)$ as n tends to infinity. The results are obtained under weak assumptions imposed on the distribution of the asset returns, namely the existence of the fourth moments is only required. Thereafter we provide a numerical study where the small- and large-sample behavior of the derived estimator are investigated. The resulting estimator shows significant improvements and it is robust to the deviations from normality.

E747: **Local stationary multiplicative modelling**
Presenter: **Christopher Walsh**, University of Vienna, Austria
Co-authors: Michael Vogt

A semiparametric multiplicative volatility model with a nonparametric part and a parametric GARCH component is studied. The nonparametric part is modelled as a product of a deterministic time trend component and of further components that depend on stochastic regressors. We propose a two-step procedure to estimate the model. The nonparametric component is estimated by a smooth backfitting procedure applied to a transformation of the model. The GARCH parameters are estimated in a second step via quasi-maximum likelihood. We show consistency and asymptotic normality of our estimators. Our results are obtained using mixing properties and local stationarity. Finally, we illustrate our method using financial data.

ES86 Room A1 OUTLIER AND ANOMALY DETECTION IN MODERN DATA SETTINGS Chair: Roland Fried

E189: **Outlier detection: efficient algorithms and definitions for specific settings**
Presenter: **Fabrizio Angiulli**, University of Calabria, Italy

Distance-based outlier detection approaches have gained popularity over the years and are today recognized among the most effective unsupervised techniques for identifying outliers within the data mining community. However, the first proposed algorithms ran in time quadratic in the size or exponential in the dimensionality of the data. Thus, efforts for developing scalable algorithms have been subsequently made. In this scenario, we point out relationships between distance-based outliers and statistical definitions and illustrate the design of efficient algorithms that can manage very large multidimensional datasets. Distance-based definitions overcome limitations of model based approaches and are quite general, in that the only thing they require is the availability of a distance function, but specific data settings may require more refined definitions. Hence, we also discuss approaches tailored for some specific settings, such as where disregarding uncertainty may lead to wrong or inaccurate conclusions, or where taking into account local notions of outlier may improve accuracy, or where the notion of distance may become meaningless.

E182: **Outlier detection methods applied to industrial context**
Presenter: **Valentina Colla**, Scuola Superiore Sant Anna, Italy
Co-authors: Silvia Cateni, Marco Vannucci

Outliers are anomalous data that occur due to erroneous measurements, anomalous process conditions or other causes. These incorrect values can unduly affect the results of an analysis and lead to incorrect conclusions or, vice-versa, can represent a rare and very relevant event to point out. In some applications outliers are more interesting than other data (e.g. network intrusion, medical diagnosis or fraud detection). Outliers detection is thus an important branch of data mining in many applications such as industrial processes, transportation, ecology, public safety. Classical outlier detection methods can be categorized into four approaches: statistical-based, distance-based, density-based and clustering-based. Traditional methods are often unsuitable to deal with real-world databases therefore in the recent years many contributions are proposed to overcome them and improve the quality of data. In the industrial context the search for anomalous data is important, as such data are collected through various sensors and devices. Outliers often are a consequence of malfunctioning of such sensors or of some deviation from normal operating conditions of the considered process. Recent efficient outliers detection methods based on artificial intelligence will be treated and compared in order to prove that in many cases they can outperform the widely adopted traditional methods.

E142: **Subspace search for outlier detection and description**
Presenter: **Emmanuel Muller**, Karlsruhe Institute of Technology - University of Antwerp, Germany

In many of today's applications, objects are described by a large variety of different attributes. However, only a few relevant attributes provide the meaningful information for outlier detection. The residual attributes are irrelevant and traditional outlier mining approaches fail to detect outliers in such high dimensional databases. To address this problem, subspace search focuses on a selection of projections. The objective is to find multiple subspaces, which show a significant deviation between outliers and regular objects. Thus, subspace search allows: (1) A clear distinction between clustered objects and outliers. (2) A description of outlier reasons by the selected subspaces. However, it lacks flexibility in handling different

outlier models. The proposal covers a flexible subspace selection scheme allowing instantiations with different outlier models. It utilizes the differences of outlier scores in random subspaces to perform a combinatorial refinement of relevant subspaces. The refinement allows an individual selection of subspaces for each outlier, which is tailored to the underlying outlier model. Hence, it enables the description of each outlier by its specific outlier properties. Empirical evaluation shows the flexibility of subspace search w.r.t. various outlier models and its ability to describe real world outliers.

E1091: **An R package for the analysis of outliers and interventions in count time series**
Presenter: **Roland Fried**, TU Dortmund University, Germany
Co-authors: Konstantinos Fokianos, Tobias Liboschik

Despite the increasing interest in models for time series of counts, many of the proposed methods are not yet made available in statistical software. Hence such models cannot be easily applied by other researchers or commercial users. We aim at filling this gap by publishing a package called `tscount` for the popular free and open source software environment `R`. We present the first version of the package that contains likelihood-based methods for the framework of count time series following generalised linear models. It does not only provide a flexible fitting function, but also functions for model assessment, outlier detection and many `S3` methods known from other model fitting functions in `R`. The usage of the package is demonstrated by real and simulated data examples.

ES100 Room H1 THE THEORY OF GRAPHICAL MARKOV MODELS: RECENT DEVELOPMENTS Chair: Giovanni Marchetti

E072: **Unifying graphs and separation criteria in graphical Markov models**
Presenter: **Kayvan Sadeghi**, Carnegie Mellon University, United States

The purpose is to unify all known graphs and separation criteria used in graphical models under the definition of mixed chain graphs (MCGs) with "the separation criterion". We provide several pairwise Markov properties that are equivalent under the compositional graphoid axioms. In addition, we show that these pairwise properties are equivalent to the global Markov property associated with the separation criteria for compositional graphoids. We show how the results collapse to the known Markov properties and theories for subclasses of MCGs.

E443: **Log-mean linear link function for discrete regression graph models**
Presenter: **Monia Lupparelli**, University of Bologna, Italy
Co-authors: Alberto Roverato

Regression graphs represent a valuable tool for investigating the independence structure in multivariate data involving variables which can be partitioned in groups, called blocks. Variables in each block are considered on equal standing and blocks are partially ordered depending on time or subject-matter considerations. Then, these models can be seen as sequences of univariate or multivariate regressions. Parameterizations of discrete regression graphs have been previously studied. In particular, one work adopts the multivariate logistic link function for modelling the sequences of regressions where the regression coefficients are marginal log-linear interactions. Inspired on the parameterization recently developed, we propose an alternative approach based on the log-mean linear link function for regression modelling where the regression coefficients are log-linear combinations of relative risks. Firstly, we show that discrete regression graph models can be simply specified by setting zero log-mean linear regression coefficients. Secondly, we compare this novel approach with other existent approaches; in particular we illustrate that, depending on the chosen parameterization, different sub-models of interest can be specified by means of further zero non-independence constraints.

E733: **Linear covariance models and their likelihood geometry**
Presenter: **Piotr Zwiernik**, University of California Berkeley, United States
Co-authors: Caroline Uhler

Gaussian models with linear constraints on the inverse of the covariance matrix are well understood. Since the likelihood function is concave, its maximization is straightforward. On the other hand models with linear constraints on the covariance matrix are more complicated and little is known about the behavior of the likelihood function. It has been observed that in certain small cases these models behave well unless a model is badly misspecified. We shed some more light on the geometry of the likelihood function of general linear covariance models, showing that with high probability standard numerical maximization procedures will easily reach the global maximum. As an example we consider Brownian motion tree models.

E766: **Causal models with hidden variables**
Presenter: **Robin Evans**, University of Oxford, United Kingdom

Directed acyclic graph models (DAG models, also called Bayesian networks) are widely used in the context of causal inference, since their structure admits to a calculus of intervention. However these models are not closed under marginalization, in the sense that they cannot faithfully represent the constraints imposed on margins of distributions which jointly follow a DAG. Other classes of graphs have been introduced to deal with this, including ancestral graphs and ADMGs, but we show that these are not sufficiently rich to capture the relevant models. We present mDAGs, a new class of graphical model which is appropriate for representing causal models when some of the variables are unobserved. Results on the Markov equivalence of these models show that when interpreted causally, mDAGs are the minimal class of graphs which can be sensibly used. Understanding such equivalences is critical for the use of causal structure learning methods, and we elucidate the state of the art on this.

ES104 Room L1 PARTICLE FILTERING Chair: Geir Olve Storvik

E348: **Particle filtering under interaction constraints**
Presenter: **Nick Whiteley**, University of Bristol, United Kingdom
Co-authors: Anthony Lee, Kari Heine

The potential benefits of parallel and distributed computing motivates study of the interaction structure of particle filters. Can we do away with resampling, or at least re-structure it in such a way as to be more naturally suited to non-serial implementation? The aim is to introduce some new algorithms in this context and discuss some of their theoretical properties.

E363: **Parameter estimation in state space models using particle approximations of the score vector**
Presenter: **Christopher Nemeth**, Lancaster University, United Kingdom

The score vector for a state-space model is, for most models, intractable. Recent work has shown that sequential Monte Carlo methods can be used to produce a particle approximation of the score vector. However, current methods either suffer from a computational cost that is quadratic in the number of particles, or produce estimates whose variance increases quadratically with the amount of data. We propose an efficient approach for estimating the score vector which has a computational cost that is linear in the number of particles, and displays only linearly increasing variance. The method is derived using a combination of kernel density estimation, to avoid the particle degeneracy that causes the quadratically increasing variance, and Rao-Blackwellisation. Estimates of the score vector can be used within a gradient ascent algorithm to perform both recursive and batch maximum likelihood estimation. Furthermore, we show that it is possible to use the estimate of the score vector within the proposal mechanism of the particle marginal Metropolis Hastings algorithm to reduce the autocorrelation of the posterior samples and improve the mixing of the sampler.

E398: **Exact approximation within particle filters**
Presenter: **Adam Johansen**, University of Warwick, United Kingdom

A number of recent developments in particle filtering are summarised focussing on the use of exact approximation techniques within particle filtering to facilitate the implementation of good approximations of idealised, intractable algorithms. Examples include "approximate Rao-Blackwellisation" and Monte Carlo approximation of block sampling strategies.

E241: **Estimation of static parameters using particle filters and a block independence approximation**
Presenter: **Geir Olve Storvik**, University of Oslo, Norway
Co-authors: Reinaldo Marques

Estimation of static parameters within state space models using particle filters remains a problematic issue. We present a novel method based on approximating the marginal posterior distribution of the latent state process by a block-independent version. Combining this with a sufficient statistic based approach for simulation of the static parameters, an online Bayesian particle filter algorithm can be constructed. We demonstrate through simulation experiments that the typical degeneracy problem of such algorithms can be avoided and efficient estimation of the parameters can be obtained.

ES107 Room N1 DESIGNING EXPERIMENTS WITH UNAVOIDABLE AND HEALTHY CORRELATION Chair: Jesus Lopez-Fidalgo

E112: **Optimal designs for random blocks model using corrected criteria**
Presenter: **Juan M. Rodriguez-Diaz**, University of Salamanca, Spain
Co-authors: Maria J. Rivas-Lopez, Sandra Martin-Chaves

The random blocks model defines a covariance structure in the data. GLS estimators of the parameters are used, and their covariance matrix is usually computed using the inverse of the GLS information matrix. This approach, however, underestimates the true variance of the parameters. Optimal designs taking into account the 'corrected' variance will be computed and compared with those using the traditional approach for different main-effect models.

E572: **Optimal design for repeated measures**
Presenter: **Rainer Schwabe**, Otto-von-Guericke University Magdeburg, Germany
Co-authors: Ulrike Grasshoff

In experimental situations often a substantial variability of the experimental units is present. This typically occurs in bio-sciences but may also appear in engineering experiments caused by varying quality conditions of the material. When repeated observations are obtained for each experimental unit, the variability and the resulting dependence of the observations within an experimental unit have to be taken into account in the statistical analysis. Since experimental units generally stem from a larger entity, it is commonly assumed that the impact of the experimental units is properly described by normally distributed random coefficients. For the particular case of random block effects the construction of optimal experimental designs has been widely discussed in the literature. Less has been done for random effects associated with the experimental conditions. The situation is treated that the experimental conditions remain fixed throughout all observations within an experimental unit. Then the design problem can be seen to be similar to that for certain univariate heteroscedastic observations. For these, quite surprisingly, the optimal designs are substantially affected by the increasing variability of the random coefficients associated with the experimental conditions. This behavior can be explained by the geometry of the design locus.

E115: **Trajectory selection for reconstruction and optimization of performance function in UMV**
Presenter: **Eva Riccomagno**, Universita degli Studi di Genova, Italy
Co-authors: Massimo Caccia, Eleonora Saggini

The definition and evaluation methods of performance indices during path-following and/or path-tracking tasks for unmanned marine vehicles is considered. Here, the performance is related to the tracking error (given by the reference path minus the observed/registered path) and, for sake of example, is chosen to be the integrals of the squared tracking error over a steady state time period, a settling period and the path approach and turn period, separately. For each time period, the design problem consists in the choice of a suitable subset of reference trajectories at which to evaluate the tracking error in order to reconstruct the performance function or to maximise it. Results from application of a Bayesian DOE strategy and also of Kriging are compared. The contribution is related to the current discussion on the establishment of new technological regulations for the introduction of autonomous robots in civilian applications.

E979: **An approach on design of experiments by using copulas**
Presenter: **Elisa Perrone**, Johannes Kepler University of Linz, Austria

Optimal experimental design is largely applied in many scientific fields. However, one of the neglected aspects of this theory is the research of the optimal design by considering properties like the level of the tail dependence as an intrinsic part of the phenomenon to be described. In this sense, copulas can be a very flexible tool. A new framework of design theory adapted to copula models is given, with also a generalization of the classical equivalence theorem. In addition, some examples are reported to show how copulas can be introduced in some classical design models to change the dependence as well as to enlarge and control the range of the association between the random variables by thinking at the copula parameter as one of the parameters to be estimated. Finally, some brief comments about how the use of copulas can also be seen as a way to check the robustness of the optimal designs obtained.

ES109 Room I1 ASTROSTATISTICS Chair: Simon Wilson

E693: **Estimation of cosmic microwave background radiation in the presence of point sources**
Presenter: **Ercan Engin Kuruoglu**, ISTI-CNR, Italy
Co-authors: Yuan Chen, Hing Cheung So, Diego Herraz Munoz

Recent satellite missions, Planck and WMAP, provide us measurements of celestial radiation in various microwave channels. The most important signal in these maps is the relic signal originating from the Big-bang, the Cosmic Microwave Background (CMB) radiation, which is contaminated with galactic and extragalactic sources as well as the satellite antenna noise. A numerical Bayesian approach, namely Markov chain Monte Carlo (MCMC) is employed to uncover the CMB from the microwave measurements in the areas outside the galactic plane. The CMB is modelled with a Gaussian distribution, while the point sources and antenna noise are modelled as following a Cauchy law and another Gaussian law, respectively. Since the contaminants together constitute a mixture noise with a Voigt profile PDF, the solution is not analytically tractable and MCMC presents itself as a very suitable and feasible methodology for estimation of CMB. The proposed method avoids detection and elimination of point sources before estimating the CMB, a procedure that would lead to information loss at the point source locations. Simulation results are presented on both synthetic and real data.

E881: **Spin scale-discretized wavelets on the sphere for the analysis of CMB polarization**
Presenter: **Jason McEwen**, University College London, United Kingdom
Co-authors: Martin Buettner, Boris Leistedt, Hiranya Peiris, Pierre Vandergheynst, Yves Wiaux

The aim is to construct a new spin scale-discretized wavelet transform on the sphere that supports a directional and steerable wavelet analysis of spin signals. Scale-discretized wavelets allow in practice the exact synthesis of a signal from its wavelet coefficients. We present new exact and efficient algorithms to compute the spin scale-discretized wavelet transform of band-limited signals on the sphere up to high band-limits. Finally,

we highlight the application of spin scale-discretized wavelets to analyse the polarization of the cosmic microwave background (CMB), a spin ± 2 signal observed on the celestial sphere.

E883: **EXONEST: the Bayesian exoplanetary explorer**
Presenter: **Kevin Knuth**, University at Albany - SUNY, United States
Co-authors: Ben Placek

The fields of astronomy and astrophysics are currently engaged in an unprecedented era of discovery as recent missions have revealed thousands of exoplanets. While, most of these exoplanets are detected by identifying transiting events, exoplanets often exhibit additional photometric effects that can be used to detect and characterize non-transiting exoplanets as well as improve characterization of transiting exoplanets. The EXONEST Exoplanetary Explorer is a Bayesian exoplanet inference engine based on nested sampling (MultiNest), and designed to analyze archived Kepler and CoRoT data. EXONEST accommodates plug-and-play models of exoplanet-associated photometric effects for the purpose of exoplanet detection, characterization, and scientific hypothesis testing. The current suite of models allows for both circular and eccentric orbits in conjunction with photometric effects, such as primary and secondary transits, reflected light, thermal emissions, ellipsoidal variations, Doppler beaming, and super-rotation. We discuss the EXONEST inference engine design, mention its successes, and introduce our plans to make the EXONEST Exoplanetary Explorer into an open source project with the capability to employ third party plug-and-play models of exoplanet-related photometric effects.

E1215: **Bayesian ICA-based source separation of Cosmic Microwave Background by a discrete functional approximation**
Presenter: **Jason Wyse**, Trinity College Dublin, Ireland
Co-authors: Simon Wilson

A functional approximation to implement Bayesian source separation analysis is introduced and applied to separation of the Cosmic Microwave Background (CMB) using WMAP data. The approximation allows for tractable full-sky map reconstructions at the scale of both WMAP and Planck data and models the spatial smoothness of sources through a Gaussian Markov random field prior. The proposed approach draws on the integrated nested Laplace approximation (INLA) methodology which has recently gained prominence in the statistics literature. The approximation is orders of magnitude faster than the alternative MCMC approaches to the CMB reconstruction problem. The performance and limitations of the approximation are also discussed.

ES111 Room C1 SAE AND INEQUALITIES MEASUREMENT Chair: Monica Pratesi

E653: **Domain prediction for discrete (count) and continuous outcomes using micro-simulation via quantiles**
Presenter: **Nikos Tzavidis**, University of Southampton, United Kingdom
Co-authors: Beate Weidenhammer, Timo Schmid, Nicola Salvati

Domain prediction for count outcomes typically relies on the use of generalised linear random effects models that make explicit distributional assumptions for example, assume a Poisson or Negative Binomial distribution. We propose a methodology for domain prediction that potentially allows for a more flexible mean-variance relationship. Using jittering, we transform the discrete outcome into a continuous one the quantiles of which are modelled by using a linear quantile mixed model, with domain random effects, that assumes an asymmetric Laplace distribution. Utilising the one to one relationship between the quantiles of the jittered and those of the discrete outcome, we obtain an estimator of the distribution function of the discrete outcome. A general set of domain-specific parameters that extends beyond averages is then estimated by sampling from the estimated distribution function. A modified version of this methodology can be used for domain prediction with continuous outcomes when there is evidence of misspecification of the assumptions of the nested error regression model. Mean Squared Error estimation is discussed. The potential of applying this methodology for estimating a range of deprivation and inequality indicators is explored.

E662: **SAS routines for variance estimation of poverty measures at regional level**
Presenter: **Francesca Gagliardi**, University of Siena, Italy
Co-authors: Gianni Betti, Achille Lemmi, Vijay Verma

The extent to which income inequality and poverty vary within countries across different regions is very relevant for policy decisions and monitoring. However, sub-national measures are scarce, given the complexity of producing indicators at the regional level from the available data and the methodological issues related to cross-countries comparability. Poverty and inequality measures have been produced on the basis of the so-called cumulation method using EU-SILC data. The quantification of efficiency gains from averaging across multiple years is not straightforward in surveys, such as EU-SILC, that are based on rotational panel. We have developed and tested two different methods to produce variance estimates for three-year averaged indicators in EU-SILC. A first, direct approach defines a common structure of strata and PSUs for the three waves of the sample, and applies standard JRR replications to the union of the three cross-sectional samples. An alternative (indirect) method has been developed to approximate the correlation across the cross-sectional waves using information from the longitudinal data of EU-SILC, which enables linking individuals and households across years when this is not possible in the cross-sectional datasets. The issue is to develop efficient SAS routines for cumulation and standard errors estimation using JRR methodology.

E858: **Small area estimates for consumption expenditure in Italy**
Presenter: **Luca Secondi**, University of Tuscia, Italy
Co-authors: Stefano Marchetti, Tiziana Laureti

The household consumption survey carried out yearly by the Italian Statistical Institute is designed to obtain reliable estimates of consumption expenditure at regional level (NUTS-2). However, local policy makers and stakeholders often require statistics on household living conditions at a lower geographical resolution/domain, but estimating these statistics directly from the household consumption survey at these levels often leads to inaccurate estimates. We try to overcome this problem, by the use of Small Area Estimation methods. In particular, we use a model-based estimator to obtain estimates of the consumption expenditure at provincial level (LAU-1). In order to obtain the estimates we use data coming from the 2012 household consumption survey together with archive population data. Results show that model based estimates have smaller errors than the corresponding direct estimates, i.e. estimates based on the sample data from the domain. Finally, intra-country comparisons of consumption expenditure are carried out by using Purchasing Power Parities. This adjustment enables us to take into account differences in the relative price levels across the country.

E904: **Occupational mobility between generations: a theoretical model with an application to Italy**
Presenter: **Irene Brunetti**, University of Pisa, Italy
Co-authors: Davide Fiaschi

A theoretical model is proposed to identify the crucial determinants of occupational mobility between generations. Inspired by the literature on Markov matrices we analyse the occupational choice of children conditioned to their parent's occupation. The model discloses that the incentive to belong to an occupational class (its social prestige and/or income) instead of another, the different opportunities generally related to family background and social environment of individuals, and the shifts in the occupational structure are the main determinants of occupational mobility. The application of the proposed model to a sample of Italian families shows that, between 1979 and 2008, the occupational status of children continues to be highly dependent on the occupational status of her/his father. In particular the youngest cohort (1966-1976) shows a decrease of occupational mobility, and both shifts in the occupational structure and the decrease of opportunity for the Working and Lower Middle class appear as the main sources of this result, despite the increase of the incentives. Furthermore, we analyse how occupational mobility varies across areas

within Italy.

ES112 Room F1 RECENT DEVELOPMENTS IN LATENT VARIABLE MODELS Chair: Angela Montanari

E456: **A latent variable model for clustering data with structure**
Presenter: **Geoffrey McLachlan**, University of Queensland, Australia

A mixture model with latent variable structured components to cluster feature vectors with known structure is considered. Latent variables in the components are used to allow for correlation between repeated observations on the same entity and also to allow for correlations between entities in the same cluster. We describe its use in the context of the clustering of gene profiles to improve the accuracy of the ranking and power of associated tests in the multiple testing of genes for no differential expression. This approach can be applied to multiple testing problems in general.

E458: **Modelling multiple ChIP-seq experiments via a Markov random field model with spatio-temporal dependencies**
Presenter: **Saverio Ranciati**, Universita di Bologna, Italy
Co-authors: Cinzia Viroli, Ernst Wit

The increasing availability of ChIP-seq output demands for advanced statistical tools to analyze the results of these experiments. A general mixture model for discrete data is proposed, exploiting a latent variable structure described by a Markov random field that takes into account both spatial and temporal dependencies. The goal is to retrieve the latent state of genomic regions bound or not by a protein, considering the case where multiple time points are observed. Furthermore, this statistical procedure aims to capture features that are commonly found in these datasets, such as variability coming from different technological/biological replicates, specific antibodies used or the conditions/treatments of each performed experiment. The method is evaluated through a large simulation study with different scenarios and also applied to a real dataset.

E533: **Regularized covariance matrix estimation via composite minimization**
Presenter: **Matteo Farne**, University of Bologna, Italy

A method to regularize large-dimensional covariance matrices under the approximate factor model assumption is presented. Existing methods perform estimation by recovering principal components and making the residual covariance matrix sparse. In our setting the same task is achieved by recovering the low rank plus sparse decomposition via least squares minimization under nuclear norm plus l_1 norm penalization. The best known method to solve this problem is based on the composition of soft thresholding (for sparsity) plus singular value thresholding (for low rank) algorithms. In this context, consistency of estimators is derived under specific assumptions on the eigenvalues of the low rank component matrix while allowing for the low rank and sparse matrices incoherence. Algorithm derivation and convergence analysis are provided, and the new procedure is compared to the existing ones under similar assumptions. The performance of our minimizer is described in a wide simulation study, where several low rank plus sparse settings are simulated according to different parameters.

E628: **Parsimonious hidden Markov models for clustering multivariate time series**
Presenter: **Antonello Maruotti**, University of Southampton, United Kingdom
Co-authors: Jan Bulla, Francesco Lagona, Marco Picone, Francesca Martella

The modelling of air pollutant concentrations has a long and varied history. We specify a multivariate hidden Markov model by describing concentrations data in terms of latent environmental regimes, i.e. specific distributions that the data take under latent environmental conditions. This approach is particularly convenient because the correlation structure of the data can be decomposed according to a finite number of easily interpretable distributions. We aim at defining a general class of parsimonious hidden Markov models, under the general mixture of factor analyzers framework, imposing different constraints on state-specific covariance matrices and allowing for a time-varying model-based clustering. Indeed, we develop a class of multivariate Gaussian HMM with a parsimonious covariance structure by considering a factor decomposition for the covariance matrix. The loading and noise terms of the covariance matrix can be constrained to be equal or unequal across latent states and the noise term can be further restricted to give a collection of eight parsimonious covariance structures. A further important goal of this research is to fit a model to environmental data which can provide a rich description of the changes in environmental conditions through time.

ES154 Room O1 MULTIVARIATE STATISTICS II Chair: Guang Cheng

E1074: **Quantiles of multivariate response variables: conditional concordance of quantile regression residuals**
Presenter: **Silvia Columbu**, University of Cagliari, Italy
Co-authors: Matteo Bottai

Unlike the mean, quantiles of a multivariate response variable are not uniquely defined. We propose a non-parametric method to study the dependence of the quantiles of a multivariate response conditional on a set of covariates. We define a statistic that measures the conditional probability of concordance of the signs of the residuals of the conditional quantiles of each univariate response. The probability of concordance is bounded from below by the value of the largest possible negative dependence and from above by that of the largest possible positive dependence. The value corresponding to the case of independence is contained in the interior of that interval. The conditional probability of concordance is modeled as a logistic regression with a logit link modified to constrain the predicted probabilities to lie within the feasible range. The estimated probabilities can be tested against the values of the largest possible dependence and independence. The method permits us to capture important aspects of the dependence of multivariate responses and assess possible effects of covariates on such dependence. The interpretation and potential of the proposed method are illustrated in a medical application.

E1140: **Designs for multi-wave experiments with history-matching**
Presenter: **Peter Curtis**, Queen Mary University of London, United Kingdom
Co-authors: Hugo Maruri-Aguilar

The interest is in sequential design for computer experiments in the context of history matching. History-matching aims to reduce the parameter space of a computer simulation over a sequential set of waves. We emulate an implausibility surface using a smooth supersaturated model, that is, a spline-like polynomial surface. This tractable emulator allows us to estimate level sets and minima with optimisation techniques that utilise gradients and Hessians such as Newton's Method or Levenberg-Marquardt. Designs are generated within these non-implausible regions by convex hull methods.

E1129: **Error propagation of compositional data transformations**
Presenter: **Mehmet Can Mert**, Vienna University of Technology, Austria
Co-authors: Peter Filzmoser, Karel Hron

The centered (clr) and isometric log-ratio (ilr) transformations are multivariate transformations allowing us to analyze compositional data with standard statistical methods in the Euclidean space. The presence of measurement errors, rounded zeros or values below detection limit may cause distortions on these transformations due to the non-linear and multivariate nature of the clr and ilr transformations. Thus, it is essential to determine the effect of the errors before these transformations are carried out. A calculus-based approach for propagation of the errors of compositional data on the log-ratio transformations (clr, ilr) is proposed. The concept of Taylor series expansions is used to propagate the errors in simplex, the sample space of compositional data. Based on Monte Carlo simulations, the effect of the errors of the compositions on the log-ratio transformed components is assessed.

E1244: **Statistical analysis of interval compositional data**

Presenter: **Paula Brito**, Universidade do Porto, Portugal

Co-authors: Peter Filzmoser, Karel Hron

Variability of individual multivariate observations can be effectively captured by data representations and methods of symbolic data analysis. In particular, individual numerical data may be aggregated in the form of intervals. Problems with statistical analysis of interval data using standard interval arithmetics can be avoided by representing them using interval midpoints and ranges. Nevertheless, further aspects need to be taken into account when aggregation is performed on compositional data, observations carrying relative information (e.g. proportions), whose geometric features are driven by the Aitchison geometry on the simplex. For example, when the relative structure of household expenditures is of primary interest, contributions of single variables (foodstuff, housing, etc.) can be merged according to regions into intervals. Consequently, the representation of interval compositional data using midpoints and ranges allows decomposing the original data set into two compositional data sets, which may then be analyzed in the framework of the logratio methodology, i.e., orthonormal coordinates are assigned to the original compositional data. As there is no canonical basis in the Aitchison geometry, special care needs to be taken to choose interpretable coordinates. We introduce basic methods of exploratory statistical data analysis for interval compositional data. Theoretical results are illustrated by real-world applications.

Monday 8.12.2014 14:50 - 16:30 Parallel Session Q – CFE

CS33 Room D2 STATISTICAL ANALYSIS OF CLIMATE TIME SERIES **Chair: Tommaso Proietti**

C1192: **Bootstrap inference on non-linear trends in climate time series data**
Presenter: **Jean-Pierre Urbain**, Maastricht University SBE, Netherlands
Co-authors: Marina Friedrich, Stephan Smeekes

The presence of possible non-linear deterministic trends in long historical climate time series such as temperatures is investigated. In contrast to the large literature considering the presence of structural breaks in deterministic trend functions, we adopt a flexible nonparametric specification of the trend function that is estimated by local polynomial methods. Most of the existing studies that follow a similar route rely on asymptotic confidence bands that are known to have rather poor coverage probability properties and are very sensitive to various issues such as the bandwidth selection and serial correlation for example. Simultaneous confidence bands for these trend functions are constructed using sieve bootstrap methods that are robust to serial correlation and have superior small sample properties. These allow us to shed some new lights on the issue of trends in climate data.

C1209: **Seasonal and cyclic changes in temperature data**
Presenter: **Tommaso Proietti**, University of Roma Tor Vergata, Italy
Co-authors: Eric Tobias Hillebrand

The aim is to assess whether and how climate change is reflected in the variation of the seasonal patterns of monthly temperature data. The position of the seasonal cycle affects strongly natural ecosystems and human activities such as agriculture, forestry, water management, transportation and tourism. Changes in the annual cycles are caused by changes in the luminosity of the sun and greenhouse gas, that is, atmospheric CO2 concentration. Location shifts in the annual temperature cycle are related to atmospheric CO2 concentrations, with possible anthropogenic causes. Both the amplitude and the phase of the annual cycle are subject to changes that are affected by global warming. Starting from the seminal work paper by Thomson, a number of studies have documented trends in the phase of the annual cycle towards later seasons that can be characterised as a structural break, with respect to the preceding period. This implies an earlier onset of the spring season at various European locations. Significant changes in the amplitude are also observed after 1954. The literature so far has concentrated on the measurement of this phenomenon by various methods, among which complex demodulation and wavelet decompositions are prominent. We offer new insight by considering models of seasonality that allow for varying seasonal drifts and heteroscedasticity. This enables us to assess the statistical relevance of the maintained changes and we validate the inferences by various methods, including out of sample predictive evaluation.

C1032: **Measuring the distance between global temperatures time series**
Presenter: **Umberto Triacca**, University of L Aquila, Italy

A novel distance between time series is proposed. This distance is based on a metric between pairs of sets of autoregressive moving average processes. It is used in order to measure the distance among global land and ocean and hemispheric temperature time series. Using this tool, we obtain an estimate of the structural dissimilarity among these climatological time series

C1089: **Data revisions and the statistical relation of sea-level and temperature**
Presenter: **Eric Hillebrand**, Aarhus University and CREATES, Denmark

Several studies have related temperature time series to sea-level time series in order to find statistical evidence for a causal influence. Since both temperature increase and sea-level rise are regionally varying phenomena, many studies have used globally aggregated temperature and sea-level time series. These time series are subject to continuous revisions and improvements, and we make the observation that these data revisions substantially influence the statistical relation of global mean temperatures and sea-level as inferred from the record. We repeat the analyses in previous works using revised data downloaded in 2013 and show that both the estimated coefficients of influence of temperature on sea-level and the resulting long-term projections of sea-level rise are sensitive to the revisions.

CS114 Room M2 CONTRIBUTIONS IN FINANCIAL ECONOMETRICS III **Chair: Christos Savva**

C038: **Time-varying risk premium in large cross-sectional equity datasets**
Presenter: **Elisa Ossola**, University of Lugano, Switzerland
Co-authors: Patrick Gagliardini, Olivier Scaillet

An econometric methodology is developed to infer the path of risk premia from a large unbalanced panel of individual stock returns. We estimate the time-varying risk premia implied by conditional linear asset pricing models where the conditioning includes both instruments common to all assets and asset specific instruments. The estimator uses simple weighted two-pass cross-sectional regressions, and we show its consistency and asymptotic normality under increasing cross-sectional and time series dimensions. We address consistent estimation of the asymptotic variance by hard thresholding, and testing for asset pricing restrictions induced by the no-arbitrage assumption. We derive the restrictions given by a continuum of assets in a multi-period economy under an approximate factor structure robust to asset repackaging. The empirical analysis on returns for about ten thousands US stocks from July 1964 to December 2009 shows that risk premia are large and volatile in crisis periods. They exhibit large positive and negative strays from time-invariant estimates, follow the macroeconomic cycles, and do not match risk premia estimates on standard sets of portfolios. The asset pricing restrictions are rejected for a conditional four-factor model capturing market, size, value and momentum effects.

C048: **A compound multifractal model for high-frequency asset returns**
Presenter: **Eric Aldrich**, University of California Santa Cruz, United States
Co-authors: Indra Heckenback, Gregory Laughlin

A model of high-frequency equity returns in clock time is built by separately modeling the dynamics of trade-time returns and trade arrivals. We first characterize the distributional behavior of high-frequency asset returns both in clock time and trade time and show that when controlling for pre-scheduled market news events, trade-time returns are well characterized by a Gaussian distribution at very fine time scales. We then develop a structured and parsimonious model of clock-time returns by subordinating a trade-time Gaussian distribution with a trade arrival process that is associated with a Markov-Switching Multifractal Duration model. Overdispersion in the distribution of inter-trade time durations is able to explain leptokurtosis and volatility clustering in clock-time returns, even when trade-time returns are Gaussian.

C1225: **Asymmetric connectedness of stocks: how does bad and good volatility spill over the U.S. stock market?**
Presenter: **Jozef Barunik**, Charles University in Prague, Czech Republic
Co-authors: Evzen Kocenda, Lukas Vacha

It is suggested how to quantify asymmetries in volatility spillovers that emerge due to bad and good volatility. Using data covering most liquid U.S. stocks in seven sectors, we provide sample evidence of the asymmetric connectedness of stocks at the disaggregate level. Moreover, bad and good volatility is transmitted at different magnitudes and the asymmetries sizably change over time in different sectors. While negative spillovers are often of substantial magnitudes, they do not strictly dominate positive spillovers. We find that the overall intra-market connectedness of U.S. stocks increased substantially during the recent financial crisis.

C1306: **On the similarity of time series dynamics: a criterion for model validation**
Presenter: **Francesco Lamperti**, Sant Anna School of Advanced Studies, Italy

Simulations are becoming increasingly popular both in economics and finance, however they suffer intrinsically from validation problems. Moreover, the choice of a suitable indicator quantifying the distance between the model and the data is pivotal to model selection. How to validate and discriminate between models are still open problems calling for further investigation. I present an information theoretic criterion to measure how close models' synthetic output replicates the dynamics of observable time series without the need to resort to any likelihood function or to impose stationarity requirements. This indicator is sufficiently general to be applied to any kind of model able to simulate or predict time series data, from simple univariate models to more complex objects including agent-based or dynamic stochastic general equilibrium models. More specifically, I use a simple function of the L-divergence computed at different block lengths in order to select the model that is better able to reproduce the distributions of time changes in the data. Using a known data generating process, I show how this indicator can be used to validate and discriminate between different univariate models providing a precise measure of the distance between each model and the data.

CS38 Room N2 MACROECONOMETRICS Chair: Marek Jarocinski

C169: **Density forecasting using Bayesian global vector autoregressions with common stochastic volatility**
Presenter: **Florian Huber**, WU Vienna, Austria

The aim is to put forward a Bayesian Global Vector Autoregressive Model with Common Stochastic Volatility (B-GVAR-CSV). We assume that country specific volatility is driven by a single latent stochastic process. Apart from computational advantages, this is also justified on the ground that the volatility of most macroeconomic quantities considered in our application tends to follow a similar pattern. Furthermore, Minnesota priors are used to introduce shrinkage to cure the curse of dimensionality. Finally, this model is then used to produce predictive densities for a set of macroeconomic aggregates. The dataset employed consists of quarterly data spanning from 1995:Q1 to 2012:Q4 and includes 45 economies plus the Euro Area. Our results indicate that allowing for stochastic volatility influences accuracy along two dimensions: First, it helps to increase the overall predictive fit of our model. This result can be seen for most of the variables considered in our forecasting exercise. Second, it helps to make the model more resilient with respect to outliers and economic crises. This implies that when evaluated over time, the log predictive scores tend to show significantly less variation as compared to homoscedastic models.

C386: **International great inflation and common monetary policy**
Presenter: **Jacek Suda**, Banque de France - National Bank of Poland, Poland
Co-authors: Anastasia Zervou

The Great Inflation of the 1970s was an international phenomenon. We study whether monetary authorities in the G7 countries were changing their responses to inflation in a similar manner during and following the Great Inflation era. Our results suggest that the common to the G7 countries inflation pattern during the Great Inflation period is associated with a common pattern in the monetary policy response to inflation. Specifically, first, we find that until the early 1980s monetary authorities in the G7 countries responded mildly to inflation and they systematically fought it throughout the 1980s. Second, we find that the estimated Taylor-rule coefficients on inflation are cointegrated, implying the existence of a long run relationship in the responses to inflation, during and right after the Great Inflation period. Third, we conduct a principal component analysis on the residuals of the estimated Taylor rules and conclude that the shocks' structure cannot account enough for the monetary policies' comovements. Finally, we find that the response to inflation weakens during the 2000s.

C817: **Variational Bayes inference for large vector autoregressions**
Presenter: **Tomasz Wozniak**, University of Melbourne, Australia
Co-authors: Reza Hajargasht

Variational Bayes provides an approximation to the joint posterior distribution of parameters of a model. The approximate posterior is usually accurate and of a tractable form. We show that when applied to large Bayesian Vector Autoregressions, proven to have excellent performance for forecasting of economic variables, Variational Bayes allows for fast and accurate computations of posterior distributions. The algorithms for the Variational Bayes estimation of VAR models with a variety of prior distributions, including hierarchical prior structures are derived. A procedure of choosing the optimal hyper-parameters of the prior distributions with respect to a Variational Bayes measure of the fit in sample is also proposed. Finally, a new estimator of the marginal data density based on the output from both MCMC and Variational Bayes estimation is shown to have good properties.

C374: **Financial regimes and uncertainty shocks**
Presenter: **Piergiorgio Alessandri**, Banca de Italia, Italy
Co-authors: Haroon Mumtaz

Financial markets are central to the transmission of uncertainty shocks. This paper documents a new aspect of the interaction between the two by showing that uncertainty shocks have radically different macroeconomic implications depending on the state financial markets are in when they occur. Using monthly US data, we estimate a nonlinear VAR where economic uncertainty is proxied by the (unobserved) volatility of the structural shocks, and a regime change occurs whenever credit conditions cross a critical threshold. An exogenous increase in uncertainty has recessionary effects in both good and bad credit regimes, but its impact on output is estimated to be five times larger when the economy is experiencing financial distress. Accounting for this nonlinearity, uncertainty accounts for about 1% of the peak fall in industrial production observed in the 2007-2009 recession.

CS52 Room P2 TEMPORAL DISAGGREGATION AND BENCHMARKING TECHNIQUES Chair: Gianluigi Mazzi

C036: **The statistical reconciliation of time series of accounts after a benchmark revision**
Presenter: **Baoline Chen**, Bureau of Economic Analysis, United States
Co-authors: Tommaso Di Fonzo, Marco Marini

The 2003-2007 U.S. annual input-output accounts, GDP-by-industry accounts and expenditure-based GDP are reconciled with the 2002 and 2007 quinquennial benchmarks and all contemporaneous constraints of the input-output accounts for the in-between years. The series are adjusted according to statistical procedures able to deal with large systems of accounts subject to both temporal and contemporaneous constraints. Our objective is to adjust the preliminary levels of the series such that they (i) are consistent with the quinquennial benchmarks available, (ii) fulfil all the accounting relationships for any given year, and (iii) show movements that are as close as possible to the preliminary information. To this end we use a simultaneous least-squares procedure based on the proportional first difference (PFD) criterion, a movement preservation principle. According to our past experiences, we evaluate the possible adoption of (i) a pure proportional adjustment (PROP) for series with breaks and high volatility that deteriorate the meaningfulness of growth rates and (ii) a priori constraints for groups of variables according to their different reliability, where this can reasonably be assumed.

C248: **Data reconciliation at Statistics Netherlands**
Presenter: **Nino Mushkudiani**, Statistics Netherlands, Netherlands
Co-authors: Jacco Daalmans, Reinier Bikker

Rapid developments of software and hardware tools of the last decades make it possible to create new, large data applications of methods that were not feasible before. An example of such an application is the implementation of the Benchmarking technique for Supply and Use tables at Statistics Netherlands (SN). These tables, consisting of over 10 000 time series, are reconciled using a state-of-the-art, commercial quadratic programming

(QP) solver XPRESS. We developed an extended multivariate Denton method for this application. Another application of reconciliation techniques for a large data set at SN is Census. We used the data reconciliation technique in combination with the repeated weighting method to make the large number of high dimensional tables consistent. Several production processes at SN use benchmarking techniques for reconciling high frequency time series with more reliable low frequency time series. Currently we observe the growing need for combining data sources of different frequencies and different properties in order to obtain combined figures. We believe that data reconciliation and benchmarking techniques are indispensable for solving these problems. At the moment we are researching the possibilities of using benchmarking techniques for different statistical processes, for example for short term statistics and labour force statistics.

C510: **Is growth-rate preservation really the best benchmarking method?**
Presenter: **Jacco Daalmans**, Statistics Netherlands, Netherlands
Co-authors: Nino Mushkudiani, Reinier Bikker, Tommaso Di Fonzo

Benchmarking monthly or quarterly series to annual data is a common practice in many National Statistical Institutes. The benchmarking problem arises when time series data for the same target variable are measured at different frequencies and there is a need to remove discrepancies between the sums of the sub-annual values and their annual benchmarks. Several benchmarking methods are available in the literature. The Growth Rates Preservation (GRP) benchmarking procedure is often considered the best method. It is often claimed that this procedure is grounded on an ideal movement preservation principle. However, we show that there are important drawbacks to GRP, relevant for practical applications, that are unknown in the literature. New benchmarking models will be introduced that do not suffer from GRP's side-effects.

C550: **Evaluation of temporal disaggregation methods**
Presenter: **Christoph Sax**, University of Basel, Switzerland

Using Monte-Carlo simulations, different standard methods for temporal disaggregation are evaluated. First, time series are simulated that are consistent with the simplifying theoretical assumptions of the methods. The methods are evaluated by their annual and sub-annual forecast accuracy, by their in-sample accuracy, and by their coefficient estimation accuracy. Second, several realistic SARIMA processes are simulated and used for evaluation, both with cointegrated and non-cointegrated time series. It is found that models perform best when the series are in accordance with the theoretical assumptions of the model. Particularly, Chow-Lin performs best for co-integrated series, Litterman or Fernandez for non-cointegrated series. If the degree of co-integration is not known, the Chow-Lin method with the max-log algorithm for finding the autoregressive parameter leads to good results.

CS70 Room F2 MIDAS MODELS: APPLICATIONS IN ECONOMICS AND FINANCE Chair: Eduardo Rossi

C649: **Nonlinear volatility dynamics: a smooth transition HAR approach**
Presenter: **Eduardo Rossi**, University of Pavia, Italy
Co-authors: Alexios Ghalanos

The role of macroeconomic variables in predicting the monthly realized variance of the S&P 500 is examined. Differently from existing studies on the topic, we focus on the nonlinear relationship between macroeconomic and financial variables and the future volatility. To this end we introduce the Smooth Transition (with exogenous) Heterogeneous AutoRegressive Moving Average (ST(X)-HARMA) model. In its unconstrained version, the predetermined variables enter linearly both in the dynamic of each state and in the threshold function which determines the transition between different states. The dynamics in each state follows a Heterogeneous AutoRegressive Moving Average model. The results of the in-sample analysis confirm that the role of macroeconomic and financial variables employed in the analysis is significant in determining the transition between states of high and low volatility. No evidence is found in favor of a nonlinear influence of macro and financial variables on the conditional mean in each state. The out-of-sample forecasting results confirm that overall the STAR models provide improvements over alternative linear models.

C727: **Capital flows and interest rates: a mixed frequency approach**
Presenter: **Emanuele Bacchiocchi**, University of Milan, Italy
Co-authors: Andrea Bastianin, Alessandro Missale, Eduardo Rossi

Many observers have pointed to current account imbalances as one of the factors triggering the global financial crisis. Looking at aggregate data, a global financial cycle clearly emerges in capital flows, asset prices and credit growth. Furthermore, this cycle is generally not aligned with the macroeconomic fundamentals of many of the actors of global financial markets. Excess credit creation, as well as asset price bubbles, under particular circumstances, can degenerate and act as a trigger for financial crises. Some authors have shown how US monetary policy plays a crucial role in the explanation of the financial cycle, driving the leverage of global banks and credit growth in the international financial system. Focusing on single countries, however, under floating exchange rates domestic monetary policy can fix domestic interest rates independently. The aim is to shed light on the relation between domestic interest rates and capital inflows for a set of countries facing different historical trends over the last decades. The econometric approach differs from other studies in the literature in that we exploit all the dynamic characteristics of the involved time series, that are collected with different frequencies. More specifically we propose and estimate a Mixed Data Sampled Structural Autoregressive (MIDAS-SVAR) model for domestic interest rates and gross capital flows.

C1002: **MIDAS systemic risk models**
Presenter: **Paolo Giudici**, University of Pavia, Italy
Co-authors: Paola Cerchiello

Financial network models are a useful tool to model interconnectedness and systemic risks in financial systems. We embed them in a stochastic framework, aimed at a more parsimonious and more realistic representation. To this end, we introduce conditional Gaussian graphical models, that can jointly model different sources of data: market prices, macroeconomic variables and microeconomic bank specific indicators. Conditional Gaussian graphical models can usefully decompose correlations between financial institutions into correlations between countries and correlations between institutions, within countries. While the former may be further explained by macroeconomic variables, the latter may be explained by idiosyncratic balance sheet ratio indicators. We have applied our proposed methods to the largest European banks, with the aim of identifying central institutions, more subject to contagion or, conversely, whose failure could result in further distress or breakdowns in the whole system. Our results show that, in the transmission of the default risk, there is a strong country effect, that reflects the weakness and the strength of the underlying economies. In addition, each country reveals specific idiosyncratic factors.

C1278: **A mixed-frequency model with stochastic volatility**
Presenter: **Clement Marsilli**, Banque de France, France
Co-authors: Laurent Ferrara, Massimiliano Marcellino

The recent econometric literature put a lot of emphasis on the role of financial variables in explaining economic growth fluctuations, especially during the Great Recession. In this respect, several mixed-frequencies have been introduced to account for this specific relationship. We extend the MIDAS regression model by allowing for stochastic volatility in the data. Parameter estimation is carried out within a Bayesian framework. As an application, we nowcast US real GDP growth rate using daily returns of stock prices by integrating stochastic volatility. We also incorporate a monthly macroeconomic variable in the model within a multi-frequency framework to account for business cycle fluctuations.

CS72 Room O2 ECONOMETRIC AND QUANTITATIVE METHODS APPLIED TO FINANCE Chair: Simona Sanfelici

C110: **Wavelet analysis and the credit spread puzzle**
Presenter: **Michaela Kiermeier**, University of Applied Sciences Darmstadt, Germany
Credit risk alone cannot explain the level of corporate credit spreads which is most prevalent for short maturity investment grade bonds. This phenomenon has entered the literature under the heading of Credit Spread Puzzle. We argue that the structural models of credit risk and standard econometric analyses implicitly assume that the relationship between credit spreads and their risk factors is time scale independent. In our approach we allow for inefficiencies such as noise trading, dispersed information, technical, feedback, fundamental, and rational trading. We use wavelet analysis to decompose the credit spread changes, and the risk factors, using the maximal overlap discrete wavelet transform (MODWT). Then we estimate the relationship on a scale-by-scale basis. The results show that the amount of credit spreads explained by risk factors is in fact high for certain time scales only. Monthly data concerning credit spreads and the risk factors is analysed. We find that the risk factors explain credit spreads best for time periods that range from 1.3 to 2.6 years.

C373: **Fourier estimation method with positive semi-definite estimators**
Presenter: **Jiro Akahori**, Ritsumeikan University, Japan
Co-authors: Nien-lin Liu, Maria Elvira Mancino, Yukie Yasuda
The aim is to present a modification of the Fourier estimation method of the spot volatility (matrix) process of a continuous Ito semimartingale where the estimators are always non-negative definite.

C385: **Measuring the leverage effect in a high frequency framework**
Presenter: **Imma Valentina Curato**, University of Ulm, Germany
Co-authors: Simona Sanfelici
Multi-factor stochastic volatility models of the financial time series can have important applications in portfolio management and pricing/hedging of financial instruments. Based on the semi-martingale paradigm, we focus on the study and the estimation of the leverage effect, defined as the covariance between the price and the volatility process and modeled as a stochastic process. Our estimation procedure is based only on a pre-estimation of the Fourier coefficients of the volatility process. This approach constitutes a novelty in comparison with the non-parametric leverage estimators proposed in the literature, generally based on a pre-estimation of the spot volatility, and it can be directly applied to estimate the leverage effect in the case of irregular trading observations and in the presence of microstructure noise contaminations, i.e. in a high frequency framework. The finite sample performance of the Fourier estimator of the leverage is tested in extensive numerical simulations.

C1149: **Hybrid generative-discriminative (HMM-SVM) machine-learning models for the forecasting of multivariate financial time series**
Presenter: **Andreas Koukorinis**, University College London, United Kingdom
Co-authors: Gareth Peters, Guido Germano
The dynamics and interaction of market microstructure variables (price, volume, duration) are crucial for the successful development of high-frequency predictive algorithms. Particularly, incorporating multiple sources of information from the market microstructure is important to forecast successfully and to implement algorithmic trading strategies.First, we use a hidden Markov model (HMM) for the intra-day price evolution and market regime. We demonstrate the existence of upper and lower dependence between the microstructure variables. We extended this to a hidden semi-Markov model (HSMM) and we use copula models for the joint dynamics of the observed transaction variables with the appropriate choices for marginal distributions. Next, we develop a hybrid regression algorithm based on discriminative learning, specifically a support vector machine (SVM), while also embedding generative models of the microstructure. We propose this combination to build our hybrid generative-discriminative approach.A set of experiments is carried out on the microstructure variables of 9 unique instruments for a period of a year. Our hybrid scheme outperforms standard HMM and HSMM.

CS86 Room B2 TOPICS IN TIME SERIES AND PANEL DATA ECONOMETRICS Chair: Martin Wagner

C481: **Monitoring a change from spurious regression to a cointegrating relationship**
Presenter: **Dominik Wied**, TU Dortmund, Germany
Co-authors: Martin Wagner
A monitoring procedure to detect a jump from spurious regression to cointegration is proposed. The procedure is based on OLS residuals obtained from regressions in a moving window. We derive asymptotic properties under the null hypothesis as well as under the alternative, allowing as usual in the cointegration literature for error serial correlation and regressor endogeneity, and investigate the finite sample behavior by Monte Carlo simulations.

C810: **Fully modified OLS estimation of spatially correlated cointegrated systems**
Presenter: **Leopold Soegner**, Institute for Advanced Studies, Austria
Co-authors: Martin Wagner
A system of spatially correlated cointegrating relationships is considered. Deterministic trend terms are also allowed for and it is shown that the convergence rate of the spatial correlation parameter is determined by the order of the trend polynomial. In addition to the correlation induced by the spatial autoregressive formulation we also allow for cross-unit correlation of the integrated regressors as well as the error terms. Cointegration amongst the regressors is not allowed, as is standard in the cointegrating regression literature. In order to obtain limiting distributions that allow for standard asymptotic inference, the fully modified OLS estimation principle is extended to the current situation. Finally, the methodology is applied to a credit risk data set.

C1064: **Testing for smooth transition cointegration with integrated or trending transition variable**
Presenter: **Oliver Stypka**, TU Dortmund, Germany
Co-authors: Martin Wagner
A test is developed for the null hypothesis of linear cointegration against the alternative of smooth transition cointegration when the transition variables is either an integrated or deterministically trending variable. In the first case we allow for correlation, but no cointegration, between the transition variable and the other explanatory variables. The test statistics are based on suitably modified FM-OLS or IM-OLS parameter estimates of the usual Taylor approximations to the smooth transition regression function. Note that hereby we provide parameter estimates for these functions with a mean zero Gaussian mixture limiting distribution, which may be of interest in itself. For the IM-OLS based test we also discuss fixed-b inference. The properties of the tests are evaluated with a simulation study.

C975: **Consistent estimation of seasonally cointegrated VARMA systems in state space representation**
Presenter: **Dietmar Bauer**, University Bielefeld, Germany
Many quarterly observed macroeconomic timeseries show trending behavior and seasonal fluctuations which cannot be accounted for using only deterministic trends and seasonal dummies but which can be eliminated using seasonal differencing motivating models showing seasonal (co)integration. These models typically are formulated in the VAR-framework. One complication in comparison to standard cointegration models here lies in the fact that seasonal cointegration involves complex cointegrating relations. Another problem lies in the difficulty of imposing restric-

tions on the short run-dynamics leading to unnecessarily high-dimensional parameter sets. These complications can be avoided if the timeseries of quarterly observations of s endogenously modeled variables is viewed as a multivariate annual timeseries of dimension 4s. In the VAR framework this only inflates the dimension of the parameter set without leading to immediate insights. In the state space framework, however, close links exist which allow for the specification of consistent estimates of the quarterly model on the basis of standard estimators for the annual model which is a standard I(1) model with additional structure. The aim is to present the relevant theory as well as two alternative consistent estimators based on subspace algorithms which are compared in a simulation study with respect to their small sample properties.

Monday 8.12.2014 14:50 - 16:30 Parallel Session Q – ERCIM

ES02 Room C1 BIG DATA ANALYSIS: PENALTY, PRETEST AND SHRINKAGE ESTIMATION II **Chair: S. Ejaz Ahmed**

E174: **Post model selection inference in high-dimensional problems**
Presenter: **Anand Vidyashankar**, George Mason University, United States

It is folklore that inference not accounting for model selection uncertainty can be biased. However, quantification of this bias and methods to mitigate them are not standard. First we provide a theoretical characterization of the inferential bias. In the process, we establish some unifications in the inferential bias as they occur in settings such as pretesting and regularization. Next, to mitigate the inferential bias, we describe two approaches for valid post-model selection inference in high-dimensional problems based on variants of bootstrap. We establish theoretical validity of the proposed methods, in finite and large samples, using newly developed concentration inequalities and high-dimensional central limit theorems. We illustrate our results using some simulations and real-data analysis.

E446: **Modeling time-varying effects for high-dimensional covariates: a new Gateaux-differential boosting approach**
Presenter: **Yi Li**, University of Michigan, United States
Co-authors: Kevin He, Ji Zhu

Survival models with time-varying effects provide a flexible framework for modeling the effects of covariates on event times. However, the difficulty of model construction increases dramatically as the number of variables grows. Existing constrained optimization and boosting methods suffer from computational complexity. We propose a new Gateaux differential-based boosting procedure for simultaneously selecting and automatically determining the functional form of covariates. The proposed method is flexible in that it extends the gradient boosting to functional differentials in general parameter space. In each boosting learning step of this procedure, only the best-fitting base-learner (and therefore the most informative covariate) is added to the predictor, which consequently encourages sparsity. In addition, the method controls smoothness, which is crucial for improving predictive performance. The performance of the proposed method is examined by simulations and by application to analyze the national kidney transplant data.

E579: **A tale of underfitted model: submodel selection and post-estimation**
Presenter: **Ejaz Ahmed**, Brock, Canada
Co-authors: Xiaoli Gao

Estimation of regression coefficients in a partially linear model with a diverging number of predictors is considered. In the arena of high-dimensional data analysis, it is often assumed that the model is sparse for simultaneous variable selection and estimation. In other words, it is assumed that the signals are well separated from zero in the model at hand. However, in many studies this assumption may not hold, resulting in producing biased parameters estimation. Our goal is to improve the estimation of the regression parameters in a realistic scenario when the model consists of weak signals as well as strong signals. We suggest an improved estimation strategy which incorporates the effect of weak signals and propose a high-dimensional shrinkage estimation strategy to improve the prediction performance of the candidate submodel. We develop the asymptotic properties of the shrinkage estimator and assess its relative performances with respect to the full model and submodel estimators. We show both analytically and numerically that the proposed high-dimensional shrinkage estimator (HDSE) performs better than the full model estimator. Interestingly, the simulation study reveals that suggested HDSE strategy significantly improves the prediction accuracy of any candidate submodel estimates generated from available LASSO-type variable selection methods. The relative performance of the proposed HDSE strategy is appraised by both simulation and real data analysis. Our data analysis strongly corroborates the analytical and simulated findings that the performance of the HDSE strategy in comparison with other strategies is favorable.

E864: **Generalized thresholding estimation of large covariance/correlation matrix for temporal data**
Presenter: **Bin Nan**, University of Michigan, United States
Co-authors: Hai Shu

The estimation of large covariance or correlation matrices using temporal samples with polynomially decaying correlation is addressed. Generalized from so-called long or intermediate memory univariate process, the temporal dependence assumption is imposed directly on the cross-covariance or cross-correlation matrix of the multivariate time series. We mainly focus on the asymptotic properties of generalized thresholding estimators including convergence rate, sparsistency and sign-consistency. As a byproduct, the consistency result of banding estimator is also provided. An intuitive cross-validation method is proposed for the thresholding parameter selection, which performs reasonably well in the simulations.The proposed thresholding estimators are applied to assessing the brain functional connectivity using fMRI data.

ES10 Room A1 ROBUST METHODS ON FUNCTIONAL DATA AND COMPLEX STRUCTURES **Chair: Graciela Boente**

E304: **An angle-based functional depth measure for shape outlier detection**
Presenter: **Andre Rehage**, TU Dortmund University, Germany
Co-authors: Sonja Kuhnt

In functional data settings, the information contained in the shape of the curves matters a great deal. The magnitude of functional data can usually be depicted by classical univariate measures; such that magnitude outliers are easily detected. Aiming at shape outlier detection a new functional depth measure is proposed based on the intersection angles which the centered curves form with each other. The measure is thereby called functional tangential angle depth (or short FUNTA depth) and is presented within a more global theoretical framework. In a simulation study its ability to detect shape outliers is investigated and compared to other popular functional depth measures. A bootstrap procedure is used to determine the cutoff value under which each depth value is declared as an outlier. Finally, the possibility of variable selection in functional linear models with functional response by means of a suitable functional depth measure is investigated.

E524: **Application of the random Tukey depth to restrictive clustering**
Presenter: **Alicia Nieto-Reyes**, Universidad de Cantabria, Spain
Co-authors: Rafael Duque, Domingo Gomez, Crescencio Bravo

A robust clustering procedure is performed for a multivariate dataset under several restrictions, such as the clusters are formed simultaneously by elements with homogeneous outputs for a prefixed subset of variables and with heterogeneous outputs for another prefixed subset of variables (of course, both subsets with empty intersection). We propose a procedure that makes use of statistical data depth; particularly, we employ the random Tukey depth, as the number of variables can be high. This problem is of interest in the field of computer-supported collaborative learnings, which studies how to take advantage of technology to improve systems that support collective learning processes in which groups of students collaborate to achieve common goals. Thus, the proposed methodology enables teachers to choose the criteria that best fit a specific learning setting, for example, forming groups in which the learners have a common set of skills (fluent communication, work speed, etc.) but they are different in other aspects (accuracy of the solutions, quality of the documentation, etc.).

E352: **Sparse robust regression estimators**
Presenter: **Gabriela Cohen Freue**, University of British Columbia, Canada
Co-authors: Matias Salibian-Barrera

In many current applications scientists can easily and cheaply measure a very large number of variables (for example, several thousands of gene expression levels) some of which are expected be useful to explain or predict a specific response variable of interest. These potential explanatory variables are most likely to contain redundant or irrelevant information, and in many cases, their quality and reliability may be suspect. A common scientific objective when analyzing this type of data is to identify a relatively small subset of explanatory variables that one can use to obtain good predictions for the response of interest. We developed a penalized robust regression estimator that can be used to identify a useful subset of explanatory variables while protecting the resulting estimator against possible aberrant observations in the data set. Using an Elastic Net penalty, the proposed estimator can be used to select variables, even in cases with more variables than observations or when many of the candidate explanatory variables are correlated. We present the new estimator and an algorithm to compute it. We also illustrate its performance in a simulation study and a real data set.

E463: **Test among populations based on the spatial sign operator**
Presenter: **Daniela Rodriguez**, Universidad de Buenos Aires - Conicet, Argentina
Co-authors: Graciela Boente, Mariela Sued

The sample spatial sign covariance operator with unknown location is studied. We show some asymptotic properties such as the consistency and asymptotic normality. We develop a test between populations based on the squared norm of the difference between the estimated spatial sign covariance operators of each population.

ES21 Room O1 BIOSTATISTICS AND BIOINFORMATICS Chair: Alexandros Beskos

E322: **Optimal screening and adaptive testing of sparse signals**
Presenter: **Wenguang Sun**, University of Southern California, United States

A common feature in large-scale scientific studies is that signals are sparse and it is desirable to significantly narrow down the focus to a much smaller subset in a sequential manner. We consider two related data screening problems: One is to find the smallest subset such that it virtually contains all signals and another is to find the largest subset such that it essentially contains only signals. These screening problems are closely connected to but distinct from the more conventional signal detection or multiple testing problems. We develop data-driven screening procedures which control the error rates with near optimality properties and study how to design the experiments efficiently to achieve the goals in data screening. An application to multistage high-throughput studies is given to illustrate the merits of the proposed screening methods.

E506: **Penalised regression methods for risk prediction in low-dimensional data with few events**
Presenter: **Menelaos Pavlou**, University College London, United Kingdom
Co-authors: Gareth Ambler, Shaun Seaman, Maria De Iorio, Rumana Omar

Prognostic regression models are used to predict the future course of clinical outcome for patients. For binary outcomes a logistic regression model is commonly used. When the number of events is small compared to the number of regression coefficients, model overfitting is a danger. An overfitted model tends to demonstrate poor predictive accuracy when applied to new data. We focus on low-dimensional data typically occurring in epidemiology and public health research. We review several penalised likelihood methods which may alleviate overfitting by shrinking the regression coefficients towards zero. In particular, we consider ridge and lasso and extensions of these (elastic net, adaptive Lasso and SCAD), their Bayesian analogues and Bayesian approaches based on spike and slab priors. We evaluate their predictive performance in comparison to standard MLE using simulation and real data analyses. MLE tends to produce overfitted models with poor predictive performance in scenarios with few events, while penalised methods offer significant improvement. Ridge regression performed well, except in scenarios with many noise predictors. Lasso tends to perform better than ridge in scenarios with noise predictors. Elastic net is an attractive compromise between the two and performed well in all scenarios. Adaptive Lasso and SCAD performed best in scenarios with many noise predictors. Bayesian approaches performed well in most scenarios and can be useful in risk prediction.

E1254: **Bayesian clinical classification from high-dimensional data: signatures versus variability**
Presenter: **Ton Coolen**, Kings College London, United Kingdom
Co-authors: A.C.C. Coolen, A. Shalabi, M. Inoue, J. Watkins, E. de Rinaldis

When data exhibit imbalance between a large number d of covariates and a small number n of samples, clinical outcome prediction is difficult due to overfitting and prohibitive computation demands. To address this we build generative and discriminative Bayesian outcome prediction protocols that can be applied to data of any dimension and any number of outcome classes. By solving the main Bayesian integrals and the optimal hyperparameters analytically, we need at most a small number of d-independent numerical integrations, leading to non-approximate algorithms with CPU demands that only scale with the dimensions of the data set as $O(nd)$. We compare our methods' performance on synthetic and real clinical (genomic) data sets to that of the benchmark mclustDA method of Fraley and Raftery. For small d they perform as well as mclustDA or better. For large dimension, e.g. 10,000 or more, mclustDA breaks down due to numerical limitations, whereas our Bayesian methods remain feasible and efficient. This also allows us to explore new phenomena typical of classification in high-dimensional spaces, such as the reduced discriminative effectiveness of signatures as opposed to the increased effectiveness of intra-class variability.

E570: **A simulation approach for change-points on phylogenetic trees**
Presenter: **Alexandros Beskos**, University College London, United Kingdom
Co-authors: Ajay Jasra, Adam Persing, David Balding, Maria De Iorio

Consider n sequences that each consists of m sites, where it is assumed that the data originate from a rooted binary tree of known topology with unknown ancestral node values. The topology of the tree is constant between sites, but the parameters of the tree can change along the length of the sequences. We consider Bayesian parameter inference from such a model with an unknown number of change-points on the sites; the posterior density is trans-dimensional. Computational inference from such a model is very challenging and two novel contributions are introduced to deal with this issue. Firstly, based upon the time machine principle, the top-most nodes of the binary tree are replaced with an approximation of the tree's stationary distribution; as more nodes are removed from the top of the graph, the cost of computing the likelihood is reduced linearly in n. The approach introduces a bias, which can be mathematically characterized. Secondly, a particle marginal Metropolis-Hastings (PMMH) algorithm is presented, which employs a sequential Monte Carlo (SMC) sampler and can use the first idea. This approach can deal with one of the bottle-necks of standard computational algorithms: the trans-dimensional nature of the posterior. Coupled with the time machine, the algorithm can perform biased inference from the model in a reliable fashion for moderate sized-data sets. The algorithm is successfully implemented on two numerical examples, and its potential to outperform competing biased methods is empirically demonstrated.

ES47 Room M1 FUNCTIONAL DATA ANALYSIS Chair: Alois Kneip

E119: **Conditional functional data analysis on random domains: Exploring Germany's Energiewende**
Presenter: **Dominik Liebl**, Universite libre de Bruxelles, Belgium
Co-authors: Alois Kneip

Hourly electricity spot prices are modelled as noisy discretization points of a time series of daily random price functions. In order to consider seasonal variations, we allow the price functions to depend on the daily mean air temperature, which leads to conditional random functions. A distinctive feature of our dataset is that the discretization points of each price function are observed only within a random subsection of the domain.

We consider this feature in our theoretical work and present in-probability convergence rates for the local linear estimators of the mean and covariance function as well as for the functional principal components. Furthermore, we introduce a new prediction approach, which allows us to recover the price functions beyond their random sub-domains. In our application, we take a look at Germany's Energiewende (energy transition) and compare the mean, covariance, and price functions of three consecutive years: one year before Germany's abrupt nuclear phaseout in mid-March, 2011 and two further consecutive years.

E495: **Optimal classification and nonparametric regression for functional data**
Presenter: **Alexander Meister**, University of Rostock, Germany

The purpose is to establish minimax convergence rates for classification of functional data and for nonparametric regression with functional design variables. The optimal rates are of logarithmic type under smoothness constraints on the functional density and the regression mapping, respectively. These asymptotic properties are attainable by conventional kernel procedures. The bandwidth selector is automatically adaptive. The functional data are considered as realisations of random variables which take their values in a general Polish metric space. We impose certain metric entropy constraints on this space; but no algebraic properties are required.

E591: **Registration and functional principal components**
Presenter: **Alois Kneip**, University of Bonn, Germany
Co-authors: Heiko Wagner

Functional data analysis deals with modeling samples of smooth functions $x_1,\ldots,x_n$. A basic tool of statistical analysis consists of the use of functional principal components. But an inherent problem in many applications is the possible existence of two types of variations: a phase variation (horizontally) due to time lags, and an amplitude variation. Determining mean and principal components then often does not lead to interpretable results, unless a registration procedure is applied which eliminates phase variation. In this context usually registration is considered as a pre-processing step. We propose a new algorithm that combines registration and principal component analysis. The approach is illustrated by some real data examples.

E722: **Functional linear instrumental regression**
Presenter: **Jan Johannes**, Universite catholique de Louvain, Belgium

The estimation of a slope function β is considered in functional linear instrumental regression, where in the presence of a functional instrument W the dependence of a scalar response Y on the variation of an endogenous explanatory random function X is modelled by $Y=\int_0^1\beta(t)X(t)dt+\sigma U$, $\sigma>0$, for some error term U. Taking into account that the functional regressor X and the error term U are correlated in many economical applications, the random function W and the error term U are assumed to be uncorrelated. Given an iid. n-sample of (Y,X,W) a lower bound of the maximal mean integrated squared error is derived for any estimator of β over certain ellipsoids of slope functions. This bound is essentially determined by the mapping properties of the cross-covariance operator associated to the functional regressor X and the best linear predictor W_o of X given the instrument W. Assuming first that W_o is known in advance a least squares estimator of β is introduced based on a dimension reduction technique and additional thresholding. It is shown that this estimator can attain the lower bound up to a constant under mild additional moment conditions. The best linear predictor of X given the instrument W is generally, however, not known. Therefore, in a second step W_o is replaced by an estimator and sufficient conditions are provided to ensure the minimax-optimality of the resulting two stage least squares estimator. The results are illustrated by considering Sobolev ellipsoids and finitely or infinitely smoothing cross-covariance operators.

ES62 Room D1 HIGH-FREQUENCY DATA STATISTICS: ADVANCES IN METHODOLOGY Chair: Hiroki Masuda

E152: **Simulation and inference of CARMA Levy models and the yuima package**
Presenter: **Stefano Iacus**, University of Milan, Italy
Co-authors: Lorenzo Mercuri

A new tool for the R package yuima is presented, available on CRAN, for the simulation and inference of a Continuous Autoregressive Moving Average (CARMA) model with some applications to real data. When dealing with the CARMA model, one of the advantages of the yuima package is the possibility of recovering the increments of the underlying noise and choosing the appropriate Levy model. The estimation of the parameters for the underlying Levy process makes yuima package appealing for modeling financial time series. Indeed, identifying the appropriate noise for a CARMA model allows us to capture asymmetry and heavy tails observed in the real data.

E203: **Common price and volatility jumps in noisy high-frequency data**
Presenter: **Markus Bibinger**, Humboldt-University Berlin, Germany
Co-authors: Lars Winkelmann

A test for common price and volatility jumps based on high-frequency data is presented. The test generalizes an approach for direct discrete observations of a semi-martingale to a latent observation model accounting for market microstructure noise. The construction relies on a non-parametric estimation of the instantaneous squared volatility process using a local method of moments spectral estimation approach. Price jumps are recovered using a truncation procedure which can be adapted to the local magnitude and intra-day shape of volatility. Central limit theorems for the estimators are given which facilitate the development of the test and open up new possibilities for inference on noisy high-frequency observations. We examine the finite-sample properties in Monte Carlo simulations and apply the method to explore price-volatility co-jumps in NASDAQ intra-day data.

E325: **Hybrid multi-step estimators for stochastic differential equations based on sampled data**
Presenter: **Kengo Kamatani**, Osaka University, Japan
Co-authors: Masayuki Uchida

An estimation problem of both drift and diffusion coefficient parameters for an ergodic diffusion process based on discrete observations is considered. Hybrid multi-step estimators are proposed and their asymptotic properties, including convergence of moments, are obtained.

E500: **Asymptotic expansion of functionals of high frequency data and their applications**
Presenter: **Nakahiro Yoshida**, University of Tokyo, Japan

The estimators used in high frequency financial data analysis, such as realized volatility, power variation, quasi-maximum likelihood estimator and Bayesian type estimator of volatility parameter, are in general asymptotically mixed normal. Limit theorems give a theoretical basis of statistical inference, however, higher-order approximation of the distribution is necessary to develop modern theoretical statistics, like higher-order statistical inference, prediction, information criteria, resampling methods and so on, as it was so in developments of ergodic statistics. We review an expansion method for a martingale having a mixed normal limit. The asymptotic expansion is described by an adjoint operation of a random symbol consisting of the adaptive random symbol and the anticipative random symbol. The former corresponds to the correction term appearing in the central limit case, and the latter is new and described by the Malliavin calculus. We discuss applications to the realized volatility and the power variation. Other applications are found in the quasi-likelihood analysis, prediction and information criterion for model selection.

ES81 Room L1 INFERENCE IN MODELS FOR CATEGORICAL DATA Chair: Tamas Rudas

E060: **Faithfulness of discrete distributions to graphs and hypergraphs**
Presenter: **Anna Klimova**, IST Austria, Austria

Co-authors: Tamas Rudas, Caroline Uhler

Faithfulness of discrete distributions is studied within several model classes. It is demonstrated first that there exist distributions that are not faithful to any directed or undirected graphical model, and thus these model classes are not closed with respect to the corresponding faithfulness relations. The class of hypergraphs whose hyperedges are maximal interactions of hierarchical log-linear models is considered next. The concept of parametric faithfulness to such a hypergraph is defined, and it is shown that this class is closed with respect to parametric faithfulness: for any distribution there exists a hypergraph to which the distribution is parametrically faithful. Parametric strong faithfulness ensures the existence of uniformly consistent estimators for the hypergraph parameters. The proportion of distributions that violate parametric strong-faithfulness depends on the parameterization and the measure of association that are used. Several examples when such proportions can be found in a closed form are given. Finally, an exact formulation of the proportion for the hypergraphs with decomposable sets of hyperedges is discussed.

E283: **A latent class model for ecological inference for the estimation of voters transitions**

Presenter: **Antonio Forcina**, University of Perugia, Italy

Co-authors: Roberto Colombi

Ecological inference tries to model the dependence between two variables at the level of individuals when observations are available only for sets of subjects clustered into local units. In the study of voters' mobility within a borough divided into polling stations, we observe the marginal distribution of voters in two elections and might want to infer the proportions of voters faithful to the party voted in the previous election, moved to a different party or abstained. All available models assume that the transition matrix for voters between two elections is the same in all polling stations. We propose a latent class model of voting transition where we assume that voters may be split into a given number of latent types and that, given the latent type, voting decisions in the two elections are independent. The model is determined by two matrices of probabilities whose product produces transition probabilities between elections: one gives the probability that a voter belongs to latent class U having voted party X at the first election and the second gives the probability that a voter who belongs to latent class U chooses party Y at the second election.

E303: **Maximum augmented empirical likelihood estimation of categorical marginal models for large sparse contingency tables**

Presenter: **Wicher Bergsma**, London School of Economics, United Kingdom

Co-authors: Andries van der Ark, Marcel Croon

Categorical marginal models (CMMs) are flexible tools for modelling dependent or clustered categorical data, when the dependencies themselves are not of interest. A major possible problem with maximum likelihood (ML) estimation of CMMs is that the size of the contingency table increases exponentially with the number of variables, so even for a moderate number of variables, say between 10 and 20, ML estimation can become computationally infeasible. An alternative method, which retains the optimal asymptotic efficiency of ML, is maximum empirical likelihood (MEL) estimation. However, we show that MEL tends to break down for large, sparse contingency tables. As a solution, we propose a new method, which we call maximum augmented empirical likelihood (MAEL) estimation and which involves augmentation of the empirical likelihood support with a number of well-chosen cells. Simulation results show good finite sample performance for very large contingency tables.

E513: **On directionally collapsible parameterizations of multivariate binary distributions**

Presenter: **Tamas Rudas**, Eotvos Lorand University, Hungary

Odds ratios and log-linear parameters are not collapsible, meaning that including a variable into the analysis or omitting one from it, may change the strength of association among the remaining variables. Even the direction of association may be reversed, a fact that is often discussed under the name of Simpson's paradox. A parameter of association is directionally collapsible, if this reversal cannot occur. It is shown that, subject to two simple assumptions, no parameter of association, which depends only on the conditional distributions, like the odds ratio does, can be directionally collapsible. On the other hand, every directionally collapsible parameter of association gives the same direction of association as a linear contrast of the cell probabilities does. Some implications for dealing with Simpson's paradox are discussed.

ES85 Room N1 SPATIAL DEPENDENCE FOR OBJECT DATA Chair: Piercesare Secchi

E384: **Spatial prediction for mildly non Euclidean data**

Presenter: **Davide Pigoli**, University of Cambridge, United Kingdom

Co-authors: Alessandra Menafoglio, Piercesare Secchi

Data belonging to Riemannian manifolds are arising in different applications, examples can be found in shape analysis, diffusion tensor imaging, covariance matrices and M-reps. In many cases, data are spatially distributed but it is not trivial to take into account spatial dependence because of the non-linear geometry of the manifold. In particular, it is not straightforward how to include a drift effect in the case of non-stationary field. A strategy is proposed to address the problem of spatial prediction for manifold valued data, with a particular focus on the case of positive definite symmetric matrices. Under the hypothesis that the dispersion of the observations on the manifold is not too large, data can be projected on a suitably chosen tangent space, where an additive model can be used to describe the relationship between response variable and covariates. Thus, classical kriging prediction can be generalized, dealing with the spatial dependence in this tangent space. The proposed kriging prediction is applied to the matrix field of covariances between temperature and precipitation in Quebec, Canada.

E696: **Functional clustering methods for dependent misaligned curves with applications to climate reconstruction**

Presenter: **Sara Sjostedt de Luna**, Umea University, Sweden

Co-authors: Konrad Abramowicz, Per Arnqvist, Piercesare Secchi, Simone Vantini, Valeria Vitelli

Functional clustering methods for dependent misaligned curves are proposed. Within a non-parametric framework, the methods utilize the dependence structure to cluster the functional data while adjusting for the misalignment of the curves. In particular we analyze a data set made of 6385 functional data derived from laminated sediment cores extracted from lake Kassjon and aim at reconstructing the evolution of seasonal climate in Northern Sweden in the last six thousand years. A new method of analysis, the Bagging Voronoi k-medoid alignment algorithm, is proposed which is able to jointly deal with the temporal dependence, the misalignment, and the presence of clusters that characterize these data.

E707: **On the measures of spatial dependence for Hilbert data: how much are you ready to pay for a kriging prediction?**

Presenter: **Alessandra Menafoglio**, Politecnico di Milano, Italy

Co-authors: Giovanni Petris

In the recent years, several environmental applications motivated a rapid growth of the geostatistical theory devoted to object data. Functional data and probability density functions are two examples of object data that geoscientists are increasingly interested in treating. We review some notions of spatial dependence which have been established in the literature on georeferenced Hilbert data, with a particular regard to the consequent concept of optimal spatial prediction, i.e., kriging. We propose a new general framework for object-oriented geostatistics grounded on the theory of spatial Gaussian random fields valued in infinite-dimensional separable Hilbert spaces. Together with the development of a new abstract model, we will discuss the properties of the finite-dimensional approximations, compared to the existing predictors based on trace-variography. We will analyze both the strengths and the open challenges of this new methodology, highlighting the critical issues to be addressed in real analyses, with a particular focus to the modelling and computational complexity of the proposed solutions.

E725: **"La citta che sale": a modern view of the city of Milan through mobile-network data**

Presenter: **Simone Vantini**, Politecnico di Milano, Italy

Co-authors: Piercesare Secchi, Paolo Zanini

The aim is to identify spatiotemporal patterns characterizing specific locations and/or specific periods possibly associated to different human activities taking place within the city of Milan in an unsupervised fashion. In particular we perform a Hierarchical ICA (HICA) of mobile network data referenced in space with spatial resolution of 250 m, and referenced in time with temporal resolution of 15 minutes. Each record is an intensity measure of the use of the mobile network in a specific site at a specific time. HICA, which is proposed, is based on a recursive hierarchical application of ICA on pairs of variables. The output of HICA is a multi-resolution, wavelet-inspired, non-orthogonal, and data-driven basis useful to perform sparse dimension reduction. Differently from ICA and similarly to wavelets, the basis provided by HICA is naturally ordered according to the dimension of each basis element support. Similarly to ICA, the basis provided by HICA, is not orthogonal and driven by the search for independent components. Moreover, we prove the sparsitency of HICA (i.e., if the variability is generated by independent sources acting on disjoint groups of variables, the probability that HICA correctly identifies these groups goes to one as the sample size increases). In the application, coherently with the geostatistical literature, instants of times are assumed to index variables while sites instances. The declination of HICA in this case allows us to impose temporal sparsity to the final representation. The analysis unveils interesting patterns interpretable in terms of working, residential, shopping, leisure, and commuting activities.

ES89 Room B1 ROBUST STATISTICS IN R Chair: Valentin Todorov

E200: **Robust standard errors for panel data: a general framework**
Presenter: **Giovanni Millo**, Assicurazioni Generali, Italy

A comprehensive, modular and flexible framework is described for estimation of robust standard errors in panel data. Heteroskedasticity and autocorrelation robust estimators are brought together with the SCC mixing-fields based estimator, the unconditional PCSE estimator and the recent double-clustering approach, trying to bring together the applied literatures in macroeconometrics, finance, political science and accounting by demonstrating the common features of these apparently different approaches. The covariance estimators are integrated in the R package 'plm' and allow robust specification and restriction testing over a number of different panel models.

E390: **Robust multiway analysis of compositional data in R**
Presenter: **Maria Anna Di Palma**, L Orientale, Italy
Co-authors: Valentin Todorov, Michele Gallo

Multiway data analysis addresses complex data structures represented as multiway data sets where data have more than two modes. The most popular methods for modeling multiway data are CANDECOMP/PARAFAC and TUCKER3. The standard algorithms for computing these models are based on alternating least squares (ALS) and thus are vulnerable to the presence of outlying data points. A single outlier could render the obtained estimates useless. Therefore robust methods are preferred. We present an R package, rrcov3way, implementing a set of functions for the analysis of multiway data sets, including PARAFAC and TUCKER3 as well as their robust alternatives. An additional feature to handle compositional data is also included through ilr transformation. Unified diagnostics, plotting functions, data examples and a manual in the form of vignette complete the package. In the presentation, basic usage of the package will be illustrated by analyzing real data from the UNIDO INDSTAT database. The database contains data on key industrial statistics indicators for the manufacturing sectors. A subset containing I countries, J sectors and K years for some indicators as value added and output will be analyzed.

E889: **Robust multivariate covariance estimation: a comparison**
Presenter: **Martin Maechler**, ETH Zurich, Switzerland
Co-authors: Maria Anna Di Palma, Valentin Todorov

Recently, several new (or newly popularized) estimators for multivariate scatter (covariance matrix when the 2nd moments exist) have been proposed; notably deterministic (Det) relatively fast versions to find local minima of the criterion for the multivariate MCD (Minimum Covariance Determinant), the S and the MM estimator (started from S). In addition, Falk's Comedian has gained renewed attraction. Where not available, we provide R implementations of these estimators and investigate their efficiency and robustness properties notably using Stahel's barrow wheel. An outlook, we will consider the behavior of these and related estimators under cellwise contamination. One goal is providing well implemented and tested versions of these new location and scatter estimates, available in Free Software R packages.

E1257: **Robust model estimation, through trimming and constraints, for mixtures of factor analyzers**
Presenter: **Francesca Greselin**, The University of Milano Bicocca, Italy
Co-authors: Salvatore Ingrassia, Luis Angel Garcia-Escudero, Alfonso Gordaliza, Agustin Mayo-Iscar

Mixtures of Gaussian factors are powerful tools for modeling an unobserved heterogeneous population, offering - at the same time - dimension reduction and model-based clustering. Unfortunately, the high prevalence of spurious solutions and the disturbing effects of outlying observations, along maximum likelihood estimation, open serious issues. We complement model estimation with restrictions for the component covariances and trimming, to provide robustness to violations of normality assumptions of the underlying latent factors. A detailed AECM algorithm, which enforces constraints on eigenvalues and tentatively discards outliers at each step, is also presented. Simulations and a real application are illustrated, and performances are compared to previous approaches showing aim and effectiveness of the proposed methodology. Moreover, the model estimation has been moved in a new setting where the mathematical and the statistical problem are well-posed.

ES91 Room P1 BAYESIAN NONPARAMETRICS AND MCMC Chair: Frank van der Meulen

E490: **Nonparametric heteroscedastic regression modeling, Bayesian regression trees and MCMC sampling**
Presenter: **Matthew Pratola**, The Ohio State University, United States
Co-authors: Hugh Chipman, Ed George, Robert McCulloch

Bayesian additive regression trees (BART) have become increasingly popular as flexible and scalable non-parametric models useful in many modern applied statistics regression problems. They bring many advantages to the practitioner dealing with large datasets and complex non-linear response surfaces, such as the matrix-free formulation and the lack of a requirement to specify a regression basis a priori. However, there are some known challenges to this modeling approach, such as poor mixing of the MCMC sampler and inappropriate uncertainty intervals when the assumed homoscedastic variance model is violated. We introduce a new Bayesian regression tree model that allows for possible heteroscedasticity in the variance model and devise novel MCMC samplers that appear to adequately explore the posterior tree space of this model.

E910: **Bayesian credible sets in the fixed design model for polished tail functions**
Presenter: **Suzanne Sniekers**, Leiden University, Netherlands
Co-authors: Aad van der Vaart

It is considered estimating the regression function f in the fixed design problem, where we have data $Y_i = f(x_i) + Z_i$ for $i \in \{1,\ldots,n\}$. Here (x_i) is a known sequence of points in the interval $[0,1]$, and (Z_i) is a sequence of unobservable i.i.d. standard normal random variables. The aim is to estimate the vector $\vec{f} = (f(x_i))_i$. We take a nonparametric Bayesian approach and use scaled Brownian motion W as a prior for f. In the Bayesian setup the observations are distributed according to the model $Y_i = W_{x_i} + Z_i$. We consider the posterior distribution of $\vec{W}$ given $Y_1,\ldots,Y_n$ and use this to construct a credible set for $\vec{f}$. The optimal scaling is dependent on the smoothness of the parameter function f. Generally this is unknown and must be estimated from the data. We consider the maximum likelihood estimator in the Bayesian model for this scaling and study its asymptotic

behaviour under the assumption that the data is distributed according to the frequentist model with some fixed function f. As a main result, we find that the asymptotic frequentist coverage of the credible set is one, if we assume that f satisfies a so-called *polished tail* condition.

E897: **Bayesian nonparametric estimation of Tsallis diversity indices under Gnedin-Pitman priors**
Presenter: **Annalisa Cerquetti**, University of Roma La Sapienza, Italy

Tsallis entropy is a generalized diversity index first derived in community ecology and then rediscovered in statistical mechanics. Bayesian nonparametric estimation of Shannon entropy and Simpson's diversity under uniform and symmetric Dirichlet priors has been already advocated as an alternative to maximum likelihood estimation based on frequency counts, which is negatively biased in the undersampled regime. We present a fully general Bayesian nonparametric estimation of the whole class of Tsallis diversity indices under Gnedin-Pitman priors, a large family of random discrete distributions recently deeply investigated in posterior predictive species richness and discovery probability estimation. We provide both prior and posterior analysis. The results, illustrated through examples and an application to a real dataset, show the procedure is easily implementable, flexible and overcomes limitations of previous frequentist and Bayesian solutions.

E409: **Bayesian estimation of discretely observed multi-dimensional diffusion processes using guided proposals**
Presenter: **Frank van der Meulen**, Delft University of Technology, Netherlands
Co-authors: Moritz Schauer

Bayesian estimation of parameters of a diffusion based on discrete time observations poses a difficult problem due to the lack of a closed form expression for the likelihood. Data-augmentation has been proposed for obtaining draws from the posterior distribution of the parameters. Within this approach, the discrete time observations are augmented with diffusion bridges connecting these observations. This poses two challenges: (i) efficiently generating diffusion bridges; (ii) if unknown parameters appear in the diffusion coefficient, then direct implementation of data-augmentation results in an induced Markov chain which is reducible. We show how both challenges can be addressed in continuous time (before discretisation) by using *guided proposals*. These are Markov processes with dynamics described by the stochastic differential equation of the diffusion process with an additional term added to the drift coefficient to guide the process to hit the right end point of the bridge. The form of these proposals naturally provides a mapping that decouples the dependence between the diffusion coefficient and diffusion bridge using the driving Brownian motion of the proposals. As the guiding term has a singularity at the right end point, care is needed when discretisation is applied for implementation purposes. We show that this problem can be dealt with by appropriately time changing and scaling of the guided proposal process.

ES98 Room E1 ADVANCES IN QUANTILE REGRESSION Chair: Stanislav Volgushev

E387: **Quantile-based spectral analysis: asymptotic theory and computation**
Presenter: **Tobias Kley**, Ruhr-University Bochum, Germany
Co-authors: Stanislav Volgushev, Holger Dette, Marc Hallin

An alternative method for the spectral analysis of time series is proposed. Instead of covariances, differences of joint copulas and the independence copula are used to quantify serial dependencies. The Fourier transformation of these quantities is considered and used to define a new spectrum: the *copula spectral density kernel*. It allows us to separate marginal and serial aspects of a time series and intrinsically provides more information about the conditional distribution than the classical location-scale model. Thus, the copula spectral density kernel is more informative than the spectral density obtained from the autocovariances. For an observed time series the copula spectral density kernel is estimated. The asymptotic properties, including the order of the bias and process convergence, of the $\ell^\infty([0,1]^2)$-valued estimator are established. The results are applicable without restrictive distributional assumptions such as the existence of finite moments and only a weak form of mixing, such as α-mixing or a geometric moment contraction condition, is required. The finite-sample properties of the new methodology and estimator are briefly investigated by simulations. A short data example is given to illustrate the potential for empirical applications.

E635: **IV quantile regression for group-level treatments, with an application to the effects of trade on the distribution of wages**
Presenter: **Denis Chetverikov**, UCLA, United States
Co-authors: Bradley Larsen, Christopher Palmer

A methodology for estimating the distributional effects of an endogenous treatment is presented that varies at the group level when there are group-level unobservables, a quantile extension of a previous work. Standard quantile regression techniques are inconsistent in this setting, even if the treatment is exogenous. Using the Bahadur representation of quantile estimators, we derive weak conditions on the growth of the number of observations per group that are sufficient for consistency and asymptotic normality. Simulations confirm superiority of this grouped instrumental variables quantile regression estimator to standard quantile regression. An empirical application finds that low-wage earners in the U.S. from 1990-2007 were significantly more affected by increased Chinese import competition than high-wage earners. We also illustrate the usefulness of the estimation approach with additional empirical examples from urban economics, labor, regulation, and empirical auctions.

E697: **Latent structures and quantiles of the treatment effect distribution**
Presenter: **Carlos Lamarche**, University of Kentucky, United States
Co-authors: Erich Battistin, Enrico Rettore

Conditions to identify quantiles of the distribution of gains conditional of values of the outcome variable in the base state are provided. The dependence between potential outcomes is modeled through a factor model. We improve on past research by allowing latent factors to affect both the level and the variance of potential outcomes, resulting in a more flexible distribution of gains. We show that the parameter of interest is a functional of features of the factor model that are non-parametrically identified. The analogue principle is used to obtain an estimator, for which we derive asymptotic behavior and finite sample properties via Monte Carlo simulations. We apply our method to evaluate the distributional effects of an Italian labor market policy that combines income support to eligible dismissed employees with wage benefits to employers who hire them. We find that the policy does not seem to have an impact on earnings three years after the program is implemented. The standard factor model specification is rejected in favor of the more flexible and general specification.

E839: **Quantile regression for extraordinarily large data**
Presenter: **Guang Cheng**, Purdue University, United States
Co-authors: Stanislav Volgushev

Quantile regression for massive data is considered by proposing a split-and-conquer approach. One aim is to figure out the most computationally efficient way to split the entire dataset while performing the best possible statistical inferences. Additionally, we comment on possible ways to handle heterogeneity in the data. The interesting interplay with the quantile index and comparison with mean regression will also be addressed.

ES118 Room Q1 CURRENT ISSUES IN BAYESIAN NONPARAMETRICS Chair: Yongdai Kim

E253: **On the general understanding of the empirical Bayes method**
Presenter: **Botond Szabo**, Budapest University of Technology, Hungary
Co-authors: Judith Rousseau

In Bayesian nonparametrics it is common to consider a family of prior distribution indexed by some hyper-parameters. The best choice of the prior out of this collection crucially depends on certain characteristics (e.g. smoothness, sparness,...) of the unknown function of interest, which

are usually not available. Therefore in practice it is common to apply data dependent choices for the hyper-parameters. Arguably, the marginal likelihood empirical Bayes method is one of the most well-known data-dependent Bayesian procedure. The performance of this method was investigated only in a few papers and mostly in specific and restricted models. The aim is to investigate the performance of this method in a general nonparametric framework. We provide a general theorem describing the frequentist behaviour of the empirical Bayes posterior distribution under "standard" assumptions. Then we apply the main theorem for various examples, recovering some of the existing results in the literature, side along with new models.

E477: **Generalized partially additive Bayesian spectral analysis regression models**
Presenter: **Taeryon Choi**, Korea University, Korea, South
Co-authors: Peter Lenk, Seongil Jo, Hyungjoon Kim

A Bayesian method for generalized partially additive regression is presented, using a spectral analysis of Gaussian process priors for the regression function. The smoothing prior distribution for the spectral coefficients incorporates hyper parameters that control the smoothness of the function and the tradeoff between the data and the prior distribution. We contrast our approach with existing Bayesian regression models for dealing with shape restrictions for the regression function and various noise distributions. The model includes covariate effects in the generalized partial linear model structure and flexible semiparametric Bayesian quantile regression model using Dirichlet mixtures. We derive the posterior distributions for these parameters and present numerical schemes to generate posterior samples. We illustrate the empirical performance of the proposed model based on synthetic data and real data applications in comparison with other existing methods. Asymptotic properties of the proposed models are also discussed.

E1262: **Multiscale Bayes in density estimation and Polya tree priors**
Presenter: **Ismael Castillo**, CNRS, France

Bayesian nonparametric estimation is discussed in the i.i.d. sampling model via a multiscale approach. As an application, the case of Polya tree priors is considered.

E1288: **Testability and consistency**
Presenter: **Bas Kleijn**, University of Amsterdam, Netherlands

Bayesian consistency theorems come in (at least) three distinct types, e.g. Doob's prior-almost-sure consistency on Polish spaces, Schwartz's Hellinger consistency with KL-priors and the 'tailfree' weak consistency of Dirichlet posteriors. We ask the question how these notions are related and argue that one characterises them conveniently using tests. We show that the existence of Bayesian tests is equivalent with Doob-like consistency of the posterior and show that Bayesian tests exist in much greater abundance than uniform tests. As examples we consider hypothesis testing problems like Cover's rational mean problem, tests for smoothness in Sobolev classes and tests for connectedness or cyclicality in (possibly infinite) networks. To achieve frequentist posterior consistency, we combine Bayesian tests with a prior condition that generalises Schwartz's KL-condition and accommodates weak consistency, e.g. involving the 'tailfree' property of the Dirichlet distribution and others.

ES144 Room I1 STATISTICAL METHODS FOR PROCESSES IN ENGINEERING AND INDUSTRY Chair: Christine Mueller

E097: **On ARL-unbiased c-charts for i.i.d. and INAR(1) Poisson counts**
Presenter: **Manuel Cabral Morais**, ULisboa, Portugal
Co-authors: Sofia Paulino, Sven Knoth

In Statistical Process Control (SPC) it is usual to assume that counts have a Poisson distribution. The non-negative, discrete and asymmetrical character of a control statistic with such a distribution and the value of its target mean may prevent the quality control practitioner to deal with a c-chart with: a pre-specify in-control average run length (ARL); a positive lower control limit; the ability to control not only increases but also decreases in the mean of those counts. Furthermore, as far as we have investigated the c-charts, proposed in the SPC literature, tend to be ARL-biased, in the sense that some out-of-control ARL values are larger than the in-control ARL, i.e., it takes longer to detect some shifts in process mean than to trigger a false alarm. We explore the notions of unbiased and uniformly most powerful unbiased (UMPU) and randomized tests to correct the bias of the ARL function. We use the programming language R to provide useful and informative illustrations of: ARL-unbiased c-charts for i.i.d. Poisson counts; 'quasi' ARL-unbiased c-charts for the mean of first-order integer-valued autoregressive (INAR(1)) Poisson counts.

E145: **Reliability prediction using regression models**
Presenter: **Maik Doering**, University of Hohenheim, Germany
Co-authors: Sebastian Bobrowski, Hong Chen, Uwe Jensen, Wolfgang Schinkoethe

In order to know the failure behavior of some mechatronic systems and components under specified operating conditions, plenty of endurance tests have to be run. However, such tests are often time-consuming and expensive. In practice manufacturers may have lots of failure data of similar products using the same technology basis under different operating conditions. Thus, one can try to derive predictions for newly developed components or new application environments through the existing data using regression models. Three regression models are considered: a parametric, a semi-parametric and a nonparametric approach. First, the parameters of the Weibull-distribution are modelled as linear functions of the covariates. Second, the Cox proportional hazards model, well-known in survival analysis, is applied. Finally, a kernel estimator is used to interpolate between empirical distribution functions. Further we discuss a model selection procedure. To illustrate this method of reliability prediction, the three classes of regression models are applied to real test data of motor experiments. Further the advantages and the disadvantages of the approaches are shown by Monte Carlo simulations.

E580: **Stochastic modeling of engineering materials for prediction of spatial mechanical characteristics**
Presenter: **Katharina Losch**, TU Kaiserslautern, Germany
Co-authors: Frank Balle, Katja Schladitz, Claudia Redenbach

Reinforcing metals with ceramic particles results in changes of material properties, leading for example to increased stiffness while preserving a light weight. This is also the case for SiC reinforced aluminium alloys, making them interesting for aero-space or automotive applications. In order to understand the material's behaviour, several deformation experiments were conducted within a scanning electron microscope (SEM). Our aim is to develop a 3D stochastic model of the material whose parameters can be estimated from 2D SEM images. Using this model, material properties can be simulated without expensive experiments improving the understanding of microstructure-property relationships. Ongoing work in the field is reported. We start by segmenting the particles using an ADMM-based method. Using the segmented images, particle sizes, shapes, and their spatial distribution are analyzed. As we want to infer on 3D characteristics, stereological methods have to be applied. The analysis and model fitting are challenging since the SiC particles show complicated and possibly non-convex shapes. Hence, standard models consisting of spherical, ellipsoidal or cylindrical grains are not suitable. Additionally, the distribution of the particles is clearly non-stationary and influenced by the manufacturing process.

E472: **Data depth for autoregression with application to crack growth**
Presenter: **Christine Mueller**, TU University Dortmund, Germany

A deterministic model of crack growth is given by the Paris-Erdogan equation. The stochastic version of this equation leads to a stochastic differential equation which can be approximated by an AR(1) model. Since crack growth has typically jumps which are innovation outliers, the

AR(1) model shall be analyzed by an outlier robust method. We propose to use data depth for this analysis and consider four types of data depth based on simplicial depth.

CS111 Room A2 CONTRIBUTIONS TO BAYESIAN ECONOMETRICS II **Chair: Joshua Chan**

C978: **Bayesian latent threshold dynamic models: identifying conventional and unconventional monetary policy shocks**
Presenter: **Jouchi Nakajima**, Bank of Japan, Japan
Co-authors: Takeshi Kimura

A new estimation framework is proposed for identifying monetary policy shocks in both conventional and unconventional policy regimes using a structural VAR model. Exploiting a latent threshold modeling strategy that induces time-varying shrinkage of the parameters, we explore a recursive identification switching with a time-varying overidentification for the interest rate zero lower bound. We empirically analyze Japan's monetary policy to illustrate the proposed approach for modeling regime-switching between conventional and unconventional monetary policy periods, and find that the proposed model is preferred over a nested standard time-varying parameter VAR model. The estimation results show that increasing bank reserves lowers long-term interest rates in the unconventional policy periods, and that the impulse responses of inflation and the output gap to a bank reserve shock appear to be positive but highly uncertain.

C1037: **Approximate Bayesian computation for Lorenz curves from grouped data**
Presenter: **Genya Kobayashi**, Chiba University, Japan
Co-authors: Kazuhiko Kakamu

There are mainly two approaches to calculate the Gini coefficients from the grouped data in a parametric framework. One is to assume a hypothetical statistical distribution for income. The other is to fit a specific functional form to the Lorenz curve. It is known that the Gini coefficient for the same hypothetical distribution is estimated more accurately from the Lorenz curve than from the distribution. However, for a flexible class of size distributions, such as generalized beta distribution, the likelihood function for the Lorenz curve for the grouped data is not analytically available and its evaluation is computationally expensive. Approximate Bayesian computation is utilised to avoid the evaluation of the likelihood function and to estimate the parameters of the Lorenz curve. The proposed approach is illustrated with the generalized beta distribution using simulated and real datasets. The empirical results suggest that the Gini coefficients can be estimated more accurately by fitting the Lorenz curve using the proposed method than by fitting the income distribution directly. It is shown that the posterior means of the Gini coefficients are not only included in the nonparametric lower and upper bounds, but also the 95% credible intervals are close to the bounds.

C1099: **Bayesian tempo-spatial estimation of the Japanese prefectural business cycle indicators**
Presenter: **Hiroki Nagashima**, Keio University, Japan
Co-authors: Teruo Nakatsuma

A new framework is proposed to evaluate the business cycle composite indicator that can incorporate regional interdependence. Many previous studies on the business cycle indicator for either nation or prefecture constructed it from macroeconomic time series in the corresponding nation or prefecture alone. As a result, interdependence in business cycle across nations or prefectures is often neglected. Although some studies were aware of the existence of the regional interdependence, it is not a common practice to explicitly incorporate it into the estimation of the business cycle indicator. Following the previous studies, we employ a Bayesian state-space modeling where the business cycle indicators, both nationwide and prefectural, are unobservable latent variables, but we make the state-space model include two types of interdependence: (1) the nationwide business cycle affects the economy in each prefecture and (2) the business cycles in neighboring prefectures are synchronized with each other to some degree. The first type of interdependence is modeled as a hierarchical structure among the nationwide and prefectural business indicators, while the second one is modeled as a spatial autoregressive (SAR) model of the prefectural indicators. We applied our new framework to the Japanese prefectural data and found supportive evidence for prefectural interdependence.

C1116: **Estimating an education production function for Switzerland, 1871-1911**
Presenter: **Camilla Mastromarco**, University of Salento - Lecce, Italy
Co-authors: Ulrich Woitek

The aim is to determine the difference between factors affecting educational outcome via the education production function and factors influencing the "efficiency" of resource use in education by using a stochastic frontier model. 19th Century Switzerland is an ideal testing ground for analyzing the impact of cultural diversity on education outcomes, because of the regional heterogeneity, and, at the same time, political stability. We build on previous works on 19th Century Prussia and the importance of protestantism for education quality. Previous works show that not only religion matters, but also other cultural determinants play a role, such as the degree of conservatism. We have measures for quite a number of district characteristics, but determinants such as "culture" are potentially not fully captured by religion and degree of conservatism - therefore, unobserved heterogeneity across districts might be wrongly attributed to inefficiency. Therefore, we estimate a panel version of a stochastic educational production frontier with both observed and unobserved heterogeneity, following the Bayesian approach. This approach allows for time- variation, and, having both observed and unobserved heterogeneity in the inefficiency term, it is also possible to use the estimated unobserved heterogeneity as a diagnostic tool to judge the performance of the available measures of observed district heterogeneity.

CS113 Room O2 COMPUTATIONAL ECONOMETRICS II **Chair: William J. Mccausland**

C684: **Semi-global solutions to DSGE models: perturbation around a deterministic path**
Presenter: **Viktors Ajevskis**, Bank of Latvia and Riga Technical University, Latvia

An approach based on a perturbation technique is presented to construct global solutions to dynamic stochastic general equilibrium models. The main idea is to expand a solution in a series of powers of a small parameter scaling the uncertainty in the economy around a solution to the deterministic model, i.e. the model where the volatility of the shocks vanishes. If a deterministic path is global in state variables, then so are the constructed solutions to the stochastic model, whereas these solutions are local in the scaling parameter. Under the assumption that a deterministic path is already known, the higher order terms in the expansion are obtained recursively by solving linear rational expectations models with time-varying parameters. A method is proposed which rests on backward recursion for solving this type of models.

C783: **Portfolio choice with parameter uncertainty: Bayesian analysis and robust optimisation comparison**
Presenter: **Ana Monteiro**, University of Coimbra, Portugal
Co-authors: Antonio Santos, Rui Pascoal

Parameter uncertainty has been a recurrent subject treated in the financial literature. The normative portfolio selection approach considers two main kinds of decision rules: utility maximization and mean-variance criteria. Assuming that the mean-variance criteria is a good approximation to the utility maximization paradigm, a major factor of concern is parameter uncertainty which, when is not taken into account, can lead to meaningless portfolios. A statistical approach, based on a Bayesian analysis, can be applied to parameter uncertainty. This can be compared with a robust optimization approach, where it is assumed that the value of the unknown parameters can change within a given region. Comparisons over these two approaches are performed. In order to evaluate their performance, we extend a measure presented in the literature that quantifies the effects of the estimation risk. The *area under the curve* (AUC) measure allows us to distinguish the approaches and select the one that implies lower losses.

C914: **A reinforcement learning approach to solving incomplete market models with aggregate uncertainty**
Presenter: **Vadym Lepetyuk**, Bank of Canada, Canada

Co-authors: Andrei Jirnyi

A method of solving heterogeneous agent models is developed in which individual decisions depend on the entire cross-sectional distribution of individual state variables, such as incomplete market models with liquidity constraints. The method is based on the principle of reinforcement learning, and does not require parametric assumptions on either the agents' information set, or on the functional form of the aggregate dynamics. It uses stochastic simulation with a kernel regression estimator both to approximate continuation values and to prioritize the updating states.

C989: **Non-ergodicity as partial identification**

Presenter: **Matteo Guido Richiardi**, University of Torino, Italy

Co-authors: Jakob Grazzini

The ergodic axiom lies at the core of economics. The fact that ergodicity is seldom testable in the data favors assuming it from the onset. However, non-ergodic, multiple equilibria models might offer valuable insights into the functioning of the real world and be more appropriate in many modeling circumstances. Far from being intractable, we show that non-ergodic models are amenable to estimation, the condition for identification being similar to those for ergodic models. We show that non-ergodicity can be usefully thought of as an additional source of uncertainty which can be quantified in empirical applications and leads to a distribution of estimates. We test our method by means of a Montecarlo experiment on an agent-based model of the labor market.

CS59 Room N2 CONTRIBUTIONS ON STOCHASTIC VOLATILITY Chair: Fulvio Corsi

C1016: **A generalized Schwartz model for energy spot prices - estimation using a particle MCMC method**

Presenter: **Wei Wei**, Aarhus University, Denmark

Co-authors: Anne Floor Brix, Asger Lund

A two-factor geometric spot price model with stochastic volatility and jumps is considered. The first factor models the normal variations of the price process and the other factor accounts for the presence of spikes. Instead of using various filtering techniques for splitting the two factors, as often found in the literature, we estimate the model in one step using a MCMC method with a particle filter. In our empirical analysis we fit the model to UK natural gas spot prices and investigate the importance of allowing for jumps and stochastic volatility. We find that the inclusion of stochastic volatility in the process used for modeling the normal price variations is crucial and that it strongly impacts the jump intensity in the spike process. Furthermore, our estimation method enables us to consider both a continuous and purely jump-driven specification of the volatility process, and thereby assess if the volatility specification also influences the spike process and the overall model fit.

C1071: **Forecasting global equity indices using large Bayesian VARs**

Presenter: **Tamas Krisztin**, WU Vienna University of Economics and Business, Austria

Co-authors: Florian Huber, Philipp Piribauer

A large Bayesian Vector Autoregressive (BVAR) model with common stochastic volatility is proposed to forecast global equity indices. Using a dataset consisting of monthly data on global stock indices the BVAR model inherently incorporates co-movements in the stock markets. The time-varying specification of the covariance structure moreover accounts for sudden shifts in the level of volatility. In an out-of-sample forecasting application we show that the BVAR model with stochastic volatility significantly outperforms the random walk both in terms of root mean squared errors as well as Bayesian log predictive scores. The BVAR model without stochastic volatility, on the other hand, underperforms relative to the random walk. In a portfolio allocation exercise we moreover show that our BVAR model with common stochastic volatility as an investment strategy produces significantly larger returns as compared to naive trading schemes.

C1108: **Returns or differences? Methods for risk functional form selection**

Presenter: **Sergei Morozov**, Ernst and Young, United States

Three categories methods are reviewed that help decide whether market risk factors should be represented as returns or differences (or some hybrid form). Methods in the first category rank alternative representations by their performance with respect to stationarity tests, in-sample or out-of-sample measures of fit, or by information-theoretic considerations. These methods must be handled with care and are often inconclusive. Second category of methods revolves around parametric models that nest both return and difference representations. Among such specifications, the one with elasticity of volatility may be preferable to those involving interpolation or mixture, while accommodating stochastic sources in volatility or its elasticity can also be useful. Third category nests return and difference representations in a non-parametric class and includes methods to estimate the volatility as a smooth function of the level, allowing for a possibility of functional representation switch depending on the level. The methods are illustrated using daily interest rate swap data, for which we find that the return formulation is preferable when rates are below roughly 2.8% while the difference representation works better for rates above this cutoff.

C1298: **Estimation of stochastic volatility and jumps using high-frequency data and Bayesian inference methods**

Presenter: **Milan Ficura**, University of Economics, Czech Republic

Two different approaches that can be used for the estimation of stochastic volatility and its decomposition into continuous and discontinuous (jump) components are compared. Namely the non-parametric approach based on high-frequency data (realized variance and multi-power-variation measures) and Bayesian inference approach (using the MCMC algorithm) in which the continuous stochastic volatility and the occurence of jumps are estimated as latent processes. A large focus is given to the analysis of the dynamics of the jump component of volatility and to the differences in its clustering behaviour (modelled using the self-exciting Hawkess process) when estimated using both of the aforementioned methods. The analysis is performed on the EUR/USD currency pair.

CS07 Room B2 CONTRIBUTIONS ON NON-LINEAR TIME-SERIES MODELS Chair: Alain Hecq

C960: **Nonparametric detection of discontinuity-points in varying coefficient regression models**

Presenter: **Bonsoo Koo**, Monash University, Australia

Estimators of location and size of multiple discontinuity-points in semi-parametric varying coefficient regression models are proposed. The discontinuity-points are explained by the values of a stochastic regressor since we incorporate thresholds into varying regression coefficients. We investigate asymptotic properties including the rate of convergence and the limiting distribution of the proposed estimators. We provide a testing procedure for a number of discontinuity-points given the sample period and we derive the limit distribution of the test statistic. A Monte Carlo study is presented to confirm the validity of our estimation and inference.

C1119: **A score driven approach for state-space models with time-varying parameters**

Presenter: **Fabrizio Venditti**, Bank of Italy, Italy

Co-authors: Davide Delle Monache, Ivan Petrella

The estimation of state-space models with time-varying parameters typically implies the use of simulation methods and, when volatility evolves stochastically, nonlinear filtering techniques. We model parameters' variation in a Gaussian state-space model where the parameters are driven by the score of the predictive likelihood. In this setup, conditionally on past data, the model remains Gaussian and the likelihood function can be evaluated using the Kalman filter. We derive the analytical expressions for the score and the information matrix needed to update the time varying system matrices. We show that this leads to a new set of recursions running in parallel with the standard Kalman filter recursions. The resulting augmented Kalman filter allows us to estimate simultaneously the unobserved state vector and the time-varying parameters by maximum

likelihood. The model is further extended to handle data at mixed frequencies. We show the usefulness of our approach for the constructions of an index of financial conditions, derived as the unobserved factor summarizing the co-movement between a panel of monthly financial indicators and quarterly macro variables such as GDP.

C1161: **Extremum estimators in practice: are approximate gradients ever useful and what can we do about it?**
Presenter: **Mateusz Dziubinski**, Aalborg University and CREATES, Denmark

The aim is to suggest and study an accurate, simple, and fast method of obtaining the gradient for the purpose of applying extremum estimation with numerical optimization algorithms in econometrics: Algorithmic Differentiation (AD), an automatic computational differentiation method, exact to the maximum extent allowed by the machine precision. It is applicable to a broad range of models. Non-linear objective functions necessitate the use of numerical optimization algorithms, many of which rely on the gradient (and often the Hessian) information. Furthermore, robust covariance matrix estimators also involve the use of derivative information. The lack of closed-form expressions often leads to the reliance on simple numerical approximations, in particular finite difference (FD) methods. We illustrate the relative performance of AD-based estimation with a numerical case study focusing on the generalized autoregressive conditional heteroskedasticity (GARCH) model. AD achieves as high as 12.5 times higher accuracy with competitive computational performance. Considering the reliability, we provide a benchmark with the goal of successful convergence to within 1% distance from the data generating process parameter values: we find that estimation using AD exhibits successful convergence in the vast majority of Monte Carlo (MC) experiments, whereas estimation using FD suffers from a considerably high convergence failure ratio. We also find that covariance matrix estimators are significantly affected by the choice of the gradient computation method: Comparing MC mean squared error and the MC means of the variance estimates of the parameter, AD achieves at least an order of magnitude higher accuracy.

C756: **Nonparametric testing for serial independence using the NRL statistic**
Presenter: **Witold Orzeszko**, Nicolaus Copernicus University, Poland

The new author's nonparametric test for detecting nonlinearity in time series is introduced. In the proposed test the NRL indicator, originally proposed as a measure of the effects of noise reduction, is applied as the test statistic. The bootstrap distributions are applied to determine the critical values of the test. The results of the Monte Carlo simulations performed in order to investigate its size and power properties are presented. Next, the new test is applied to detect the nonlinear dynamics in selected financial time series.

CS108 Room P2 CONTRIBUTIONS TO APPLICATIONS IN MACROECONOMICS AND TIME SERIES Chair: Luis F. Aguiar-Conraria

C431: **Interpreting economic policy uncertainty - real economic activity causality: the role of infrequent structural shifts and omitted variables**
Presenter: **Paraskevi Salamaliki**, University of Konstanz, Germany
Co-authors: Ioannis Venetis

A thorough examination of Granger causality inference is provided by focusing on the role of trend specification and the presence of infrequent structural shifts, as well as the role of omitted variables, in vector autoregressive (VAR) models. Our analysis extends previous work by considering the cases of structural shifts in the level and/or the slope of the trend function (growth rate changes) in trend stationary VAR models. The effects of infrequent shifts on the size of Granger causality tests in finite samples are examined through extensive simulations under different combinations of breaks and sets of dynamic parameters. We further provide a detailed empirical investigation of the role of economic policy uncertainty in real economic activity, which shows that the predictive ability of economic policy uncertainty for real economic activity significantly depends on the presence (or absence) of infrequent structural shifts in the time series employed and the absence of relevant variables from the information set. Finally, we show that employing a multi-horizon causality testing procedure provides a clearer picture on the dynamic interactions among the system variables and the direct or indirect nature of causal effects.

C966: **Regional business cycles across Europe**
Presenter: **Ana Gomez-Loscos**, Bank of Spain, Spain
Co-authors: M Dolores Gadea, Eduardo Bandres

Much effort has been devoted in the existing literature to studying the business cycle in Europe. Numerous studies have analyzed the business cycles and the synchronization between countries that make up the Monetary Union; while the attention placed by scholars to the regional framework has been scarce. Against this background, the aim is to draw a map of the regional business cycles in the euro area countries to help the identification of the role of the economic geography and the macro-factors. As previously shown for the Spanish case, the national cycle may hide different regional rhythms of economic activity, which may have implications for implementing economic policies. We use different dating techniques -non-parametric, parametric and finite mixtures. While we find evidence in favor of a unique cluster amongst the European countries, we determine four different groups of European regions according to their business cycles dating and their synchronization degree. The results have important implications for policy makers, both at the European level, in terms of convergence policies, and at a country level, in terms of domestic policies.

C1103: **Financial bubble detection using cross-sectional dispersion of price earnings ratios**
Presenter: **Takayuki Mizuno**, National Institute of Informatics, Japan
Co-authors: Takaaki Ohnishi, Tsutomu Watanabe

How can we detect economic bubbles? Greenspan claimed "it was very difficult to definitively identify a bubble until after the fact - that is, when its bursting confirmed its existence" in 2002. However, recently, we found that real estate bubbles are detectable making use of information on the cross-sectional dispersion of real estate prices. We apply this technique in order to detect financial bubbles. During bubble periods, prices tend to go up considerably for some stocks, but less so for others, so that market capitalization inequality across companies increases. A key characteristic of stock market bubbles is not the rapid market capitalization hike itself but a rise in market capitalization dispersion. One of the important determinants of market capitalization is the income of the company. We introduce PER in order to eliminate the contribution of the income of a company to its market capitalization. We next show that the PER distribution in NYSE and NASDAQ had a power-law tail during .COM and subprime bubble periods, while it was very close to a lognormal before and after the bubble period. These evidences suggest that the power-law tail observed during bubble periods arises due to the lack of price arbitrage across stocks.

C1141: **CO2 price dynamics in the carbon market of California**
Presenter: **Luis Aguiar-Conraria**, Universidade do Minho, Portugal
Co-authors: Rita Sousa, Maria Joana Soares

The carbon market in California is the latest emerging greenhouse gas (GHG) market, created under the Assembly Bill 32 (AB32). Whereas there has been extensive research on carbon prices, built mainly on data from Europe, we present a first analysis of the California carbon allowances prices, after they started to be traded in August 2011. We characterize CO2 prices interrelation with energy prices in California. We follow the previous studies and consider CO2; prices interrelation with energy prices (in our case, gas, oil and electricity).To characterize carbon markets, we rely on multivariate wavelet analysis (MWA) and work in the time-frequency domain, estimating how carbon price relationships behave at different frequencies and how they evolve over time. We choose to work with MWA mainly for two reasons. First, it has been shown that energy price dynamics are strongly nonstationary and so it is important to use methods that do not require stationarity. Second, we note that changes in power supply quantities, on a large scale, are neither easy nor quick. Therefore, it makes sense to consider the presence of long-term decisions, or at lower

frequencies, i.e., correlations in several temporal cycles. This can be easily done with wavelet analysis.

CS105 Room E2 FINANCIAL APPLICATIONS II **Chair: Andreas Savvides**

C1156: **A variance spillover analysis without covariances: what do we miss?**
Presenter: **Katja Gisler**, University of Saint Gallen, Switzerland
Co-authors: Matthias Fengler

The aim is to evaluate the relevance of covariances in the transmission mechanism of variance spillovers across the US stock, US bond and gold markets from July 2003 to December 2012. For that purpose, we perform a comparative spillover analysis between a model that considers covariances and a model that considers only variances. Our results emphasise the importance of covariances. Including covariances leads to an overall increase of the spillover level and detects the beginnings of the financial crisis and of the US debt ceiling crisis earlier than the spillover measure that considers only variances. Even for the low-dimensional system that we consider, one misses important variance spillover channels when covariances are excluded.

C1050: **Financialization of the commodity markets. Conclusions from the restricted VARX ADCC MVT GARCH**
Presenter: **Karol Szafranek**, Narodowy Bank Polski, Poland

The global economy is highly dependent on the commodity prices, which are, by and large, the outcome of the market-specific supply and demand fundamentals. As a result, driven by different determinants, financial assets and commodity prices should be negligibly correlated. However, systematically growing engagement of the noncommercial investors equipped with financial engineering innovations, generous inflow of capital resulting from the necessity for wider diversification of investment portfolios combined with the strengthening influence of purely financial and speculative motives have recently led to much stronger relation between the financial and commodity markets sparking a heated debate on the commodities markets financialization. The presented empirical analysis supports the claim that since 2005 commodity markets have been under heavier influence of macroeconomic, financial and speculative determinants. Results of the restricted VARX ADCC MVT GARCH model demonstrate that the inclusion of the commodity markets' growing sensitivity to macroeconomic conditions, financial markets turmoil and the impact of behavioral aspects alters the dynamic conditional correlation between commodities and the financial markets from 2005 to 2011 signaling the process of financialization. Additional conclusions are drawn regarding the structural stability of the market correlations as well as the stability of the parameter estimates.

C918: **A computational model of hidden costs in saving decisions**
Presenter: **Rodica Ianole**, University of Bucharest, Romania
Co-authors: Elena Druica

An alternative hypothesis to the standard economic models of saving is proposed, based on new behavioral and computational economics lines of research. We acknowledge the importance of traditional factors of influence for saving (income, interest rate, demographics etc), but we extend the explanatory spectrum by focusing on the conformity variable. We introduce the idea that if each individual from a specific group may initially have a low propensity to consume (and implicitly a high propensity to save), under group dynamics and direct/indirect group pressure all the individuals may change their behavior in order to be more similar to the next person, leading to a new group mainly characterized by a propensity to consume. At this stage of research, we validate the hypothesis by using NetLogo simulations, departing from a biological framework – the Altruist model. We find that saving behavior is bonded by a psychological cost which has the potential to switch the individuals' focus from a saving philosophy centered on benefits, to a philosophy centered on costs.

C1160: **IPO timing: an option to expand**
Presenter: **Hanane Dakhli**, Sorbonne, France

The existing literature dealing with the Initial Public Offerings (IPOs) shows that this market operations generally go along with the undervaluation of the candidate company and a significantly increase of market prices of listed companies belonging to the same industry. Given this observation, we propose a stochastic model to determine the optimal timing of IPOs by considering this investment decision as a real option to expand. Since investors evaluate the IPO candidates from observed market prices of peer firms, it may force to delay its IPO when market prices go down. Our results show that the launch of an IPO can be viewed as the exercise is an expensive real option. Moreover, the optimal exercise of this option perfectly coincides with the optimal timing of IPOs when there is a positive market price trend. Finally, our conclusions also lead to give an explanation to the clustering of IPOs.

CS51 Room M2 CONTRIBUTIONS IN FINANCIAL ECONOMETRICS IV **Chair: Roy Cerqueti**

C1186: **Option pricing via risk-neutral density forecasting**
Presenter: **Stanislav Khrapov**, New Economic School, Russia

A novel approach to option pricing is proposed. It exploits strong predictability in option-implied risk-neutral densities. To illustrate the idea we use a mixture of log-normal and a generalized beta as the candidates for the distribution of underlying stock price under the risk-neutral measure. Using the closed-form solutions for option prices we extract risk-neutral densities and forecast them one week ahead. This forecast allows us to compute one week ahead option prices and compare them to the observables. In the empirical exercise we show that the option pricing performance is on par with state-of-the-art stochastic volatility models.

C1219: **Relation between risk and return in international stock markets revisited**
Presenter: **Christos Savva**, Cyprus University of Technology, Cyprus

The aim is to provide further evidence of the effects of the conditional skewness and kurtosis on the relationship between the returns and conditional volatility in international stock markets with our model being able to endogenously measure the impact of the skewness and kurtosis. More specifically we adopt a model that utilizes the non-centered skewed generalized t distribution with heteroskedastic error term. Once we account for the effects of skewness-kurtosis the relationship between risk and return becomes positive and significant.

C1292: **Portfolio optimization with down side risk: an application on the Romanian stock market**
Presenter: **Cristiana Tudor**, Bucharest University of Economics, Romania
Co-authors: Andrei Anghel, Maria Tudor

The Modern Portfolio Theory assumes investors are looking for optimal portfolios, i.e. portfolios which provide the minimum variance for a given level of return. Nonetheless, minimizing the variance penalizes for both the downside and the upside of the distribution, which in turns implies that investors are also avoiding increasing prices or gains. To eliminate this anomaly, we propose a portfolio optimization technique which employs CVAR to estimate the size of the tail loss and develop it in R. We apply this technique on the Bucharest stock market and compute CVaR both parametrically and from historical prices.

C923: **Portfolio technical efficiency assessment with DEA: the case of the PSI-20 enterprises**
Presenter: **Nuno Ferreira**, ISCTE-IUL, Portugal
Co-authors: Manuela M Oliveira

Financial researchers have long been interested in the factors that explain differences in the enterprises financial performance. Following the general

concept of efficiency which relates the difference between observed and optimal values of inputs and outputs, the efficiency of the individual companies of the Portuguese stock market is assessed using DEA methodology. With the interest income, depreciation, cost of goods, employees and the net sales as factors applied in a truncated regression, the technical efficiency (TE) scores of the PSI-20 enterprises index are explored in order to verify the major contributors to efficiency variability. The results revealed that TE is higher for the enterprises in communications, media and energy economic sectors whereas the construction has the lowest TE scores. The "employees" factor contribute statistically to the stock market inefficiency, thus the bigger the enterprise the lesser the efficiency. The factors "cost of goods" and "interest income" are also destabilized factors to the efficiency though without statistical significance.

Monday 8.12.2014 16:55 - 18:15 Parallel Session R – ERCIM

ES140 Room A1 CONTRIBUTIONS TO ROBUSTNESS AND OUTLIER IDENTIFICATION Chair: Matthias Templ

E229: **Iterative identification of multiple time series outliers using singular spectrum analysis and outlier maps**
Presenter: **Jacques De Klerk**, North-West University, South Africa

Singular Spectrum Analysis (SSA) is a powerful non-parametric time series technique which is finding wide application in time series analysis. SSA is particularly powerful for time series exhibiting seasonal variation with/without trend components. SSA can be applied to time series found in market research, economics, meteorology and oceanology, to name but a few. SSA places a univariate time series into a multivariate framework by unfolding into a Hankel structured matrix. Outliers that might be present in time series can unduly influence model fitting, forecasting results and confidence intervals constructed using bootstrap methodology. A method is proposed for using iterative identification of multiple time series outliers using a stepwise approach by applying outlier maps constructed using robust principal component analysis to identify outliers and eliminating the effect of each outlier using SSA signal reconstruction methods, until all outliers are identified. Outlier maps represent multivariate data in a two dimensional plot consisting of projected orthogonal distances plotted against score distances, in order to identify outliers. The usefulness of the technique is illustrated using simulated time series and an example from practice. The method is practically compared to existing time series methods that can identify multiple outliers.

E1043: **Outlier identification and productivity analysis on the Farm Accountancy Data Network (EU-FADN – DG AGRI)**
Presenter: **Steffen Liebscher**, Martin-Luther-University Halle-Wittenberg, Germany
Co-authors: Thomas Kirschstein, Mathias Kloss

A key source of information about farm sector productivity in the EU is the Farm Accountancy Data Network (EU-FADN – DG AGRI), a farm level data set for 25 out of the 27 EU member states. It covers accounting as well as structural information about farms from all over Europe. However, these data may contain outliers biasing production function estimation. A way to deal with this problem is to identify and to remove the outliers prior to the estimation. In this field of application, only simple variable-wise techniques have been used so far. Therefore, a non-parametric outlier identification procedure based on minimum spanning trees, which takes the multivariate nature of the underlying data into account, is applied in order to detect the outliers. Results show that by using this method outliers can be detected which would stay hidden when using only standard outlier detection procedures. Taking into account these outliers in follow-on production function estimation leads to substantially different coefficient estimates as compared to those obtained from the original data or with different procedures. Consequences of these findings, which in turn are used as a basis for policy reforms, are presented.

E1289: **On the robustness of risk measures**
Presenter: **Mikhail Zhelonkin**, University of Geneva, Switzerland
Co-authors: Valerie Chavez-Demoulin

The topic of robustness in risk measurement has emerged relatively recently, but it has already attracted a lot of attention. The problem was studied using various approaches, and several methods aiming at robustifying the risk measures have been proposed. However, the general robustness theory is still missing. We focus on the parametric estimators of risk measures and use the Hampel's infinitesimal approach to derive the robustness properties. We study two most popular risk measures, namely Value-at-Risk and Expected Shortfall. We derive the influence functions for the general parametric estimators of these measures. We show that for various distributions the classical estimators, e.g. maximum likelihood, have unbounded influence functions, and are not robust. Using the expression of the influence function we propose a general strategy to construct the robust estimators and explore their properties. Finally, we demonstrate the performance of our new methodology with several illustrative examples.

ES146 Room E1 REGRESSION MODELS Chair: Adrian Bowman

E853: **Identifiability of regression models from noisy design by numerical algebraic fans**
Presenter: **Sonja Kuhnt**, TU Dortmund University, Germany
Co-authors: Eva Riccomagno, Nikolaus Rudak

An important issue in the design of experiments is the question of identifiability of models. In the considered application noisy non-standard designs occur as intermediate result in a two stage process of thermal spraying. Observations or predictions from a standard experimental design are themselves input variables for an eventual output. Presenting the design points as zeros of a system of polynomial equations is the entry key to using tools from algebraic statistics, especially numerical algebraic fans, for the study of model identifiability. Model building is thereby opened towards higher order interactions rather than restricting the class of considered models to main effects or two-way interactions only. The new approach is compared to classical model building strategies in an application to a thermal spraying process.

E941: **Linear regression model with compositional response**
Presenter: **Eva Fiserova**, Palacky University, Czech Republic
Co-authors: Karel Hron

When the response variables are compositional, i.e. multivariate observations carrying only relative information (proportions, percentages), a special treatment for their regression analysis is necessary. Compositional data are characterized by the simplex sample space with the Aitchison geometry that forms the Euclidean structure of the sample space. Using proper log-ratio transformations, the data are moved isometrically to the real Euclidean space where it is possible to use standard regression tools. A multivariate approach to regression modeling has several advantages in comparison to a series of univariate models. Specifically, multivariate models respect the association between outcomes, and thus, in general, procedures are more efficient. Further, they can evaluate the joint influence of predictors on all outcomes and avoid the issue of multiple testing. On the other hand, there are situations when the multivariate model can be decomposed into a series of simpler models, univariate or multivariate, depending on the issue. The aim is to present differences and equivalences between multivariate and univariate approaches to regression with compositional response variables. Theoretical results will be applied to a real-world example from geochemistry.

E1124: **A quantile regression model for failure time data with time-dependent covariates**
Presenter: **Yair Goldberg**, University of Haifa, Israel
Co-authors: Malka Gorfine, Yaacov Ritov

Since survival data are collected over time, often important covariates also change over time. Such covariates are referred to as time-dependent covariates. Quantile regression offers a flexible survival data modeling by allowing the covariates to vary with quantiles. A novel quantile regression model accommodating time-dependent covariates is presented, for analyzing survival data which are subject to right censoring. In addition, a doubly-robust estimator is introduced. The utility of the proposed methodology is demonstrated using the Stanford heart transplant dataset.

E1001: **Some results on Partial Least Squares regression, shrinkages and DoF**
Presenter: **Paolo Foschi**, University of Bologna, Italy

The statistical properties of the PLS regression estimator is considered. As already recognised by some authors, the shrinkage factors represent one of the most natural statistics for studying the PLS estimator. By means of a Krylov formulation, it is shown that the PLS estimator solves a weighted polynomial regression problem whose weights are the elements of the response vector. The shrinkages, then, corresponds to the associated

predictors. The shrinkages and the estimator range on a domain that consists of the union of a set of simplices. This allows us to express both the estimator and the shrinkage vector as an average of the extremes of that domain. The weights of that average are rational functions of the response vector. Some conclusions can be drawn from that characterisation. Firstly, it results that not only the shrinkages can exceed one, but that they can also become negative. Also, the derived expressions allow us to construct a counter example that invalidates the conjecture that the number of DoF of the PLS estimator always exceeds the number of parameters.

ES152 Room L1 STATISTICAL INFERENCE II **Chair: Gil Gonzalez-Rodriguez**

E638: **A Poisson-half-logistic model: properties, estimation and Bayes prediction under progressive type-II censoring**
Presenter: **Alaa Abdel-Hamid**, Beni-Suef, Egypt

A new two-parameter lifetime distribution with increasing-constant hazard rate is proposed. The genesis of the distribution is stated on a complementary risk problem base. Properties of the distribution are discussed. Based on progressively type-II censored samples, maximum likelihood estimates of the considered parameters are obtained and their properties are investigated via a simulation study. Point and interval predictions of future lifetimes are discussed. One-sample and two-sample Bayesian prediction schemes are also considered. The coverage probabilities of the prediction intervals are computed. Finally, an example based on simulated data and another one based on real set data are given to illustrate the obtained results.

E907: **Affine invariant divergences with applications to robust statistics**
Presenter: **Hironori Fujisawa**, Institute of Statistical Mathematics, Japan
Co-authors: Takafumi Kanamori

In statistical inference, divergences play an important role. An estimator of parameter can be obtained as the minimizer of divergence. We focus on an invariant divergence under affine transformation of data, and then we obtain an explicit class of divergences with empirical estimability. It is proved that this class is uniquely determined under some conditions, including affine invariance and empirical estimability. A definition of cross entropy is extended to deal with a broader class of divergence. We also investigate the relation to the Bregman divergence.

E945: **Difference-based weighted mixed Liu estimator in semiparametric partial linear models**
Presenter: **Esra Akdeniz Duran**, Istanbul Medeniyet University, Turkey

A commonly used semiparametric partial linear model with multicollinear parametric part and correlated errors is considered. We propose analyzing this model using a difference-based approach. The procedure estimates the linear component based on the differences of the observations and then estimates the nonparametric component by either a kernel or spline method using the residuals of the linear fit. We combine the difference based method with Liu-type estimation to handle multicollinear parametric part. Availability of some prior information regarding regression coefficients in the form of stochastic linear restrictions $r = R\beta + \varepsilon$ is assumed. When the sample information and prior information are to be assigned not necessarily equal weights, we introduce a new estimator by combining ideas underlying the weighted mixed regression estimation and difference-based Liu estimator under the assumption that the errors are not independent and identically distributed. We call this new estimator as generalized difference-based weighted mixed Liu estimator. The proposed estimator is analysed and compared in the sense of mean-squared error. The efficiency properties of the difference-based weighted mixed regression method are analyzed. Numerical performance of the procedure is studied using both simulated and real data.

E947: **Inference for univariate and bivariate interval censored data**
Presenter: **Baris Surucu**, Orta Dogu Teknik Universitesi, Turkey
Co-authors: Konul Bayramoglu

A lifetime of an individual or an item may not be exactly observed but may be known to lie between two successive times. This situation is described as interval censoring. In the first part, we focus on different models on interval censored data. First, we consider maximum likelihood estimation when interval boundaries are fixed. To do that, we make use of modified maximum likelihood estimation, previously proposed in the literature. Secondly, we assume random intervals and obtain the corresponding likelihood function. For the random interval case, we consider two different models. In the first model, copula approach is taken. In the second model, exponential failure times and random interval distributions are considered. We discuss the parameter estimation and obtain Fisher information matrices. In the second part of the study, we examine possible models for bivariate interval censoring. Subjects in laboratory tests may experience bivariate events, each of which may be interval-censored. To estimate the association between two survival variables, we assume that they follow a copula model. We present some numerical results to show the efficiencies of the models. We also give an algorithm to generate interval censored data for these models.

ES88 Room Q1 BAYESIAN METHODS II **Chair: Sara Wade**

E842: **Bayesian modelling ultra-high frequency financial data via particle filters**
Presenter: **Golnaz Shahtahmassebi**, Nottingham Trent University, United Kingdom
Co-authors: Rana Moyeed

A class of models, called dynamic zero inflated Poisson difference (DZPD) models, is introduced which enable us to model integer-valued variables taking both negative and positive values in a dynamic framework. Using sequential Monte Carlo methods, we develop a strategy to sequentially learn about the states and parameters of the DZPD model. Furthermore, we introduce the sequential Bayes factor (SBF) and sequential deviance information criterion (SDIC) and use them for the purpose of model comparison. We illustrate the application of the DZPD model on three sets of ultra-high frequency (UHF) financial data: Western Texas Intermediate (WTI) oil price change, E Mini S&P 500 and FTSE100 index change. The unique characteristics of such data, e.g. discrete structure of price change and unequally spaced time intervals, have introduced new theoretical and computational challenges. Thus, we show that the DZPD model is flexible enough to adapt to the structure of the data and is able to characterise the behaviour of price/index change well.

E1150: **Stochastic domination in predictive density estimation for ordered normal means under α-divergence loss**
Presenter: **Genso Watanabe**, Mejiro University, Japan
Co-authors: William E. Strawderman

Stochastic domination in predictive density estimation problems is considered when the underlying loss metric is α-divergence ($D(\alpha)$) loss introduced by Csiszàr. The underlying distributions considered are normal location-scale models, including the distribution of the observables, the distribution of the variable whose density is to be predicted, and the estimated predictive density which will be taken to be of the plug-in type. The scales may be known or unknown. We derive a general expression for the α-divergence loss in this set-up, and show that it is a concave monotone function of quadratic loss, and also of the variances (predicand, and plug-in). We demonstrate $D(\alpha)$ stochastic domination of certain plug-in predictive densities over others for the entire class of metrics simultaneously when "usual" stochastic domination holds in the related problem of estimating the mean with respect to quadratic loss. Presented examples of $D(\alpha)$ stochastic domination relate to the problem of estimating the predictive density of the variable with the larger mean.

E1195: **Bayesian inference from ranks with applications in genomics**
Presenter: **Valeria Vitelli**, University of Oslo, Norway
Co-authors: Oystein Sorensen, Arnoldo Frigessi, Elja Arjas

Modeling and analysis of ranks have received renewed interest in the era of big data, when recruited and volunteer assessors compare and rank objects to facilitate decision making in disparate areas, from politics to entertainment, from education to marketing. In the field of genomics, use of ranks has been advocated because of its platform independence, and has been successful for prediction models and for meta-analysis. The Mallows rank model is among the most successful approaches, but its use has been limited to a particular form based on the Kendall distance, for which the normalizing constant has a simple analytic expression. We develop computationally tractable methods for performing Bayesian inference in Mallows models with any right-invariant metric, thereby allowing for greatly extended flexibility. Our methods also allow estimation of consensus rankings for data in the form of top-t rankings, pairwise comparisons, and other linear orderings. In addition, clustering via mixtures allows finding substructure among assessors. Finally we construct a regression framework for ranks which vary over time. We illustrate and investigate our approach on simulated data, on performed experiments, and on benchmark data.

E1217: **A truncation algorithm for normalized RMI mixtures**
Presenter: **Raffaele Argiento**, CNR-Imati, Italy
Co-authors: Ilaria Bianchini, Alessandra Guglielmi

A well-known problem, when fitting a Bayesian mixture model whose mixing distribution is a normalized measure with independent increments (normalized RMI), is the computational effort needed to compute the posterior. In this case, the MCMC schemes are complicated by the presence of an infinite number of parameters. We propose to exploit a truncation method to approximate the mixing measure and simplify the computations. Since a normalized RMI is a random discrete measure where the weights are obtained by normalization of the jumps of a Poisson process, we discard those larger than a threshold ε. In this way the number of parameters becomes finite and an efficient blocked Gibbs sampler scheme to simulate from the posterior is built. Furthermore we assign a prior distribution for ε, so that data can "drive" the degree of approximation. We specialize our analysis considering a large class of normalized RMIs, named generalized normalized gamma processes. Finally, the performance of our algorithm either on an univariate or a multivariate dataset will be illustrated.

ES110 Room N1 CONTRIBUTIONS TO DEPENDENCE MODELS AND COPULAS Chair: Irene Gijbels

E092: **Multivariate normalized rank reflection symmetry of copula functions**
Presenter: **Lorenzo Frattarolo**, Universita Ca Foscari, Italy
Co-authors: Dominique Guegan, Monica Billio

A new test is proposed for normalized rank reflection symmetry of copula functions that is valid beyond the bivariate case. We provide a probabilistic interpretation of this symmetry and discuss some of its implications. The asymptotic theory for the test is based on a new result in empirical processes that allows for sub exponentially beta mixing data. A simulation study of the empirical size and power is conducted and examples of application are provided.

E153: **Simulating Levy-frailty copulas built from an α-stable Levy subordinator**
Presenter: **Lexuri Fernandez**, Technical University of Munich, Germany
Co-authors: Matthias Scherer

The efficiency of three different methodologies to simulate Levy-frailty copulas (a subclass of Marshall-Olkin copulas) built from α-stable Levy subordinators is investigated. We compare the following techniques: a method based on the recursive formula for general exchangeable Marshall-Olkin copulas, the simulation of the involved α-stable subordinator on a fine grid, and the simulation of the approximation of the α-stable subordinator by a compound Poisson process. We measure efficiency in terms of computational speed, considering different values for the dimension of the mentioned copulas and index α of the subordinator. As an interesting by-product, this provides an approximate simulation algorithm for the exponential functional of an α-stable Levy subordinator.

E983: **Generalized additive models for conditional dependence structures**
Presenter: **Thibault Vatter**, University of Lausanne, Switzerland
Co-authors: Valerie Chavez-Demoulin

A generalized additive modeling framework is developed for taking into account the effect of predictors on the dependence structure between two variables. We consider dependence or concordance measures that are solely functions of the copula, because they contain no marginal information: rank correlation coefficients or tail-dependence coefficients represent natural choices. We propose a maximum penalized log-likelihood estimator, derive its root-n-consistency and asymptotic normality, discuss details of the estimation procedure and the selection of the smoothing parameter. Finally, we present the results from a simulation study and apply the new methodology to two real datasets. In the first application, we study the expression of genes involved in the development of breast cancer. More specifically, we quantify a correlation increase between target genes as a function of predictor genes. In the second application, we study the dependence structure of intraday asset returns. We show that the intraday dependence pattern, due to the cyclical nature of market activity, is shaped similarly to the conditional second moment.

E1024: **Modelling the dependence in multivariate longitudinal data by pair copula decomposition**
Presenter: **Marta Nai Ruscone**, Cattolica del sacro cuore, Italy
Co-authors: Silvia Angela Osmetti

A new flexible way of modeling the dependence between the components of non-normal multivariate longitudinal-data is proposed by using the copula approach. The presence of longitudinal data is increasing in the scientific areas where several variables are measured over a sample of statistical units at different times, showing two types of dependence: between variables and across time. In order to account both type of dependence the proposed model considers two levels of analysis. First given a specific time, we model the relations of variables using copula. The use of the copula allows us to relax the assumption of normality. In the second level, each longitudinal series, corresponding to a given response over time, is modelled separately using a pair copula decomposition to relate the distributions of the variables describing the observation taken in different times. The use of the pair copula decomposition allows us to overcome the problem of the multivariate copulae used in the literature which suffer from rather inflexible structures in high dimension. The result is a new extreme flexible multivariate longitudinal model, which overcomes the problem of modelling simultaneous dependence between two or more non-normal time-series.

ES156 Room O1 CLUSTER-BASED METHODS Chair: Enea Bongiorno

E1026: **Non-Euclidean cluster analysis for covariance matrices with applications to diffusion tensor data**
Presenter: **Safa Elsheikh**, University of Brighton, United Kingdom
Co-authors: Diwei Zhou, Andrew Fish , Roma Chakrabarti

The focus is primarily on clustering covariance matrices, taking into account the non-Euclidean nature of the space of positive semi-definite matrices. Non-Euclidean metrics including the log-Euclidean, Riemannian, Euclidean root, Cholesky and Procrustes size-and-shape metrics, have been integrated into a cluster analysis algorithm. The main motivation is to cluster diffusion tensor data in medical image analysis according to the geometric structure of the tensors. How to choose the number of clusters in the real application is considered. Simulation studies are carried out to compare different non-Euclidean methods and to validate the cluster analysis algorithm. The methods are also applied to the segmentation of the corpus callosum region from a real human brain diffusion tensor image. Non-Euclidean metrics are preferable to cluster diffusion tensor data. The segmentation of the corpus callosum is effective.

E1098: **Clusterwise sparse principal component regression**
Presenter: **Bunpei Arishige**, Doshisha University, Japan
Co-authors: Hiroshi Yadohisa

Multiple regression analysis with numerous independent variables gives rise to several problems; chief amongst these is multicollinearity. This can make the coefficient estimates highly variable in response to small changes in the model. Furthermore, spurious correlation, caused by latent cluster structures in the data can lead to the misinterpretation of the results. To overcome these two problems concurrently, Nakamura and Maekawa proposed a clusterwise regression method using reduced rank regression. However, the calculated loading matrix used in Nakamura and Maekawa's method is fixed in every cluster, so that we have to put the same interpretation on each cluster. In addition to this, the loading matrix calculated by Nakamura and Maekawa's method cannot always be easily interpreted. We propose a new method called clusterwise sparse principal component regression, in which we use cardinality constraints for the loading matrix. In this method, we apply sparse principal component regression to each cluster, such that, we can interpret the loading matrix of each cluster, separately.

E1034: **Reduced k-means for multivalued quantitative symbolic variables**
Presenter: **Hiroyasu Abe**, Doshisha University, Japan
Co-authors: Kensuke Tanioka, Hiroshi Yadohisa

In recent years data tables have become very large. In such situations, data may not be easy to visualize and get interpreted by k-means clustering. A problem is that the scatter plot is rendered unclear by the large number of individuals and the lack of concepts associated with these individuals. However, the large sample problem can be solved using symbolic data analysis. In addition, another problem is the difficulty in determining the relationships among many variables. The large number of variables also obscures the interpretation of the cluster centers. Consequently, k-means clustering of many variables may yield insignificant results. These large variable problems can be resolved using tandem analysis, which combines k-means clustering with a dimensional reduction method such as principal component analysis. Tandem analysis, however, is problematic because the low dimensional space generated by the dimensional reduction method does not always have a clustered structure. Nevertheless, a better solution is offered by the reduced k-means (RKM) algorithm, which derives a low dimensional space with a definite cluster structure. The application of RKM to multivalued quantitative symbolic variables is proposed for visualizing the classification results of data containing several individuals and a large number of variables.

E1038: **Fixed factor analysis with clustering observations**
Presenter: **Kohei Uno**, Osaka-University, Japan
Co-authors: Kohei Adachi

In the fixed factor model for factor analysis (FA), the common factor scores given to observations are treated as fixed parameters, but they cannot be estimated jointly with the other parameters (factor loadings and unique variances), since the maximum likelihood (ML) diverges which follows from normality assumptions. In order to allow all parameters to be jointly estimated, we propose a constrained fixed factor model, in which the observations are supposed to be classified into a few clusters with each cluster characterized by an equivalent factor score. The ML procedure with this model is named fixed clustered FA (FCFA). An iterative algorithm for FCFA is developed which provides the ML estimates of factor loadings, unique variances, the membership of observations to clusters, and their factor scores. Real data examples demonstrate that FCFA attains the better classification of observations than the factorial and reduced K-means procedures which are based on principal component analysis (PCA) rather than FA. We discuss that this superiority of FCFA follows from differences of FA to PCA.

ES143 Room M1 CONTRIBUTIONS ON FUNCTIONAL DATA AND TIME SERIES Chair: Pedro Galeano

E274: **Simplicial principal component analysis for density functions in Bayes spaces**
Presenter: **Karel Hron**, Palacky University, Czech Republic
Co-authors: Alessandra Menafoglio, Matthias Templ, Klara Hruzova, Peter Filzmoser

Probability density functions are frequently used to characterize the distributional properties of large-scale database systems. As functional compositions, densities carry primarily relative information. As such, standard methods of functional data analysis (FDA) are not appropriate for their statistical processing. The specific features of density functions are accounted for in Bayes spaces, which result from the generalization to the infinite dimensional setting of the Aitchison geometry for compositional data. The aim is to introduce a concise methodology for functional principal component analysis of densities. We propose the simplicial functional principal component analysis (SFPCA), which is based on the geometry of the Bayes space $\mathcal{B}^2$ of functional compositions with Hilbert space properties. We perform SFPCA by exploiting the centred log-ratio transform, an isometric isomorphism between $\mathcal{B}^2$ and L^2 which enables one to resort to standard FDA tools. Advances of the proposed approach are demonstrated using a real-world example of population pyramids in Upper Austria.

E855: **Aspects of functional data analysis in short term prediction of non-stationary economic time series**
Presenter: **Daniel Kosiorowski**, Cracow University of Economics, Poland
Co-authors: Dominik Mielczarek, Jerzy Rydlewski, Malgorzata Snarska

Functional data analysis (FDA) offers several novel methods of time series prediction which conceptually differ from classical econometric approaches and hence give us a hope for a development of new areas of economic theory. One of the FDA approaches is to divide the observed time series into a sample of trajectories, and to predict a future trajectory by using functional regression ideas applied to a collections of pairs or triplets of the extracted trajectories and lagged trajectories. We critically investigate several recently proposed nonparametric methods of forecasting of non-stationary time series using FDA ideas, and propose a new robust method of short term prediction of economic time series which is very competitive to the methods known from the literature. Our forecasting method appeals to data depth ideas for functional data, is computationally feasible and may be applied in practice. We will demonstrate the effectiveness of our proposal using several economic time series and show pros and cons of using FDA ideas in comparison to more conventional methods of time series analysis. We will underline a novelty of the FDA approaches in studying economic phenomena using analytical as well as merit arguments.

E1200: **Application of T-spaces in modeling compositional time series**
Presenter: **Petra Kynclova**, Vienna University of Technology, Austria
Co-authors: Peter Filzmoser, Karel Hron

Multivariate time series describing relative contributions to a total (like proportional data) are called compositional time series, and they need to be transformed first to the usual Euclidean geometry before a time series model is fitted. It is shown how an appropriate transformation can be chosen, resulting in coordinates with respect to the Aitchison geometry of compositional data. Using vector autoregressive models, the standard approach based on raw data is compared with the compositional approach based on transformed data. The results from the compositional approach are consistent with the relative nature of the observations, while the analysis of the raw data leads to several inconsistencies and artifacts. The compositional approach is extended to the case when also the total of the compositional parts is of interest. Moreover, a concise methodology for an interpretation of the coordinates in the transformed space together with the corresponding statistical inference (like hypotheses testing) is provided.

E1233: **Modelling systematic risk asymmetry of the American Real Estate securities**
Presenter: **Paola De Santis**, Iccrea Banca SpA, Italy

Co-authors: Carlo Drago

The time varying systematic risk of the Americal Real Estate securities is explored. In particular, we estimate with relevant econometric tools the systematic risk of the American Real Estate securities and we show their changes in the year of the subprime crisis. In order to evaluate the systematic risk we have used the Fama-French three factor model which is a linear model aimed at studying the relationship between the extra-return of the REIT index, used as proxy of the American Real Estate securities, and the extra return of the S&P500 index, represented by the market portfolio return. To estimate the model we use different specifications (GARCH-ARCH-AR). We evidence the presence of changes in the model parameters, by rolling analysis and we identify the different structural breaks in the time series with the Quandt test. In order to find the model which best represents the phenomenon we use these structural breaks in order to divide the time interval into sub-homogeneous intervals and we estimate the Fama-French model in these time intervals. In this way we obtain different models in the different sub-period and different specifications that confirm the thesis of the time-varying systematic risk of the American Real Estate securities and its increase in the year of the subprime crisis.

ES136 Room B1 CONTRIBUTIONS TO HIGH-DIMENSIONAL DATA ANALYSIS Chair: Ana Colubi

E936: **A bootstrap approach to construct high-dimensional variable selection confidence sets**

Presenter: **Chao Zheng**, University of Melbourne, Australia

Co-authors: Davide Ferrari

Up to date, methods for selecting a single model have received considerable attention in many contexts, including regression analysis. However, in the presence of a large number of predictors – and especially when the number of predictors grows with the sample size – the hope of identifying a single best model is usually unrealistic due to the dominant effect of model selection variability. In the context of a high dimensional linear regression setting, we propose to select a relatively small set of models – called a variable selection confidence set (VSCS) – containing models that are not significantly different from the true model at a given level of confidence. When the signal-to-noise ratio is large, the VSCS will contain a relatively small number of equally useful models, each containing possibly different predictors. When the data contain only little information the VSCS will be large. To construct the VSCS, we propose a bootstrapped penalized likelihood approach in combination with hypothesis testing. As a by-product of the procedure, we obtain a well-performing feature selection method by principled averaging of the models in the VSCS.

E1207: **Persistence of terms in lasso**

Presenter: **Hugo Maruri**, QMUL, United Kingdom

Co-authors: Simon Lunagomez, Peter Curtis

The lasso methodology has recently atracted attention in the context of models with hierarchy restrictions. In these models, an interaction term is allowed only if both main effects are active (strong hierarchy) or if at least one main effect is active (weak hierarchy). Recent papers address nonlinear, Bayesian and convex relaxation approaches to hierarchical lasso. Our proposal aims to describe the evolution of terms as the parameter of regularization takes different values. As model terms become active or inactive, the structure of interactions changes. We propose a description of the model in terms of components and cycles, borrowing from recent developments in computational topology in the area of persistent homology. In our setting, components are groups of variables that interact, whereas cycles describe higher order interactions that are not currently included in the model. We briefly describe the elements from statistics and computational topology involved and then present and discuss preliminary results.

E1227: **Desparsified lasso for nodewise graph estimation using a shrinkage estimator**

Presenter: **Lourens Waldorp**, University of Amsterdam, Netherlands

Statistical inference for high-dimensional data is becoming a necessary tool for researchers. Large-scale networks are estimated and in many cases differences between networks are required to be tested. Several methodologies have been developed to tackle the issue of testing hypotheses for parameters in the high-dimensional setting. Some such methods are (a) multisample splitting by Meinshausen and others, (b) the significance test for the lasso by projecting the data on the active support by Lockhart and others, and (c) the desparsified lasso by van der Geer and others. We focus on using the desparsified lasso. The desparsified lasso reduces the bias of a 'normal' lasso and, consequently, makes derivation of the limiting distribution possible. An interesting result is the fact that the nodewise regression by Meinshausen and Buhlmann results in a clear bound such that the nice properties of the desparsified lasso are obtained. We study the properties of the desparsified lasso when using a shrinkage estimator instead of a nodewise regression relaxed inverse. A crude bound on the estimates of the shrinkage parameters, from the Ledoit and Wolf theorem, seems to lead to similar results of the desparsified lasso, although additional assumptions are required. With some simulations we show some finite sample results.

E838: **Improving the graphical lasso estimation for the precision matrix through roots of the sample covariance matrix**

Presenter: **Vahe Avagyan**, Universidad Carlos III de Madrid, Spain

Co-authors: Andres Alonso, Javier Nogales

The purpose is to focus on the estimation of a high-dimensional precision matrix. We propose a simple improvement of the graphical lasso framework (glasso) that is able to attain better statistical performance without sacrificing too much the computational cost. The proposed improvement is based on computing a root of the covariance matrix to reduce the spread of the associated eigenvalues, and maintains the original convergence rate. Through extensive numerical results, using both simulated and real datasets, we show the proposed modification outperforms the glasso procedure. Finally, our results show that the square-root improvement may be a reasonable choice in practice.

ES142 Room G1 CONTRIBUTIONS TO ORDINAL AND PREFERENCE DATA Chair: Tamas Rudas

E1220: **Bayesian mixture of Plackett-Luce models for partially ranked data**

Presenter: **Cristina Mollica**, Sapienza Universita di Roma, Italy

Co-authors: Luca Tardella

The Plackett-Luce model is one of the most popular and frequently applied parametric distributions to analyze partial top-rankings of a finite set of items. A Bayesian finite mixture of Plackett-Luce models is illustrated, that extends a Bayesian device recently introduced in the literature in order to account for unobserved sample heterogeneity. We describe an efficient way to incorporate the latent group structure in the data augmentation approach and how to interpret existing maximum likelihood procedures as special instances of the proposed Bayesian analysis. Bayesian inference is conducted with the combination of the Expectation-Maximization algorithm for maximum *a posteriori* estimation and the Gibbs sampling iterative procedure, with a focus on the identifiability problems that can affect the results of the MCMC technique. The novel Bayesian Plackett-Luce mixture is illustrated with an analysis of real preference partially ranked data, which discusses the application of several relabeling algorithms to solve the label-switching issue and the resulting posterior estimates.

E976: **Misspecification test for finite mixture logistic models for clustered binary and ordered responses**

Presenter: **Silvia Bacci**, University of Perugia, Italy

Co-authors: Francesco Bartolucci, Claudia Pigini

An alternative to using normally distributed random effects in modeling clustered (i.e., longitudinal or hierarchical) binary and ordinal responses is based on using a finite mixture. This approach gives rise to a flexible class of generalized linear mixed models for item responses, multilevel data, and longitudinal data. A test of misspecification for these models is proposed which is based on the comparison between the marginal and the conditional maximum likelihood estimates of the fixed effects as in the Hausman's test. The asymptotic distribution of the test statistic is derived;

this is of chi-squared type with a number of degrees of freedom equal to the number of fixed effects for the covariates which vary within the cluster. It turns out that the test is simple to perform and may be used also to select the number of support points of the finite mixture, when this number is unknown. The approach is illustrated by a series of simulations and three empirical examples covering the three main fields of application.

E1057: **Estimation for discrete response variables in dual frame surveys**
Presenter: **David Molina**, University of Granada, Spain
Co-authors: Antonio Arcos, Maria del Mar Rueda, Maria Giovanna Ranalli

Recently, dual frame surveys have gained much attention and became largely used by statistical agencies and private organizations to decrease sampling costs or to reduce frame undercoverage errors that could occur with the use of only a single sampling frame. It is usual when conducting a dual frame survey to find questionnaire items with discrete outcomes (especially in public opinion, marketing and official surveys). Items where respondents must select one in a series of non-ordered options can be modeled by a multinomial distribution while those other items with ordered categories can be modeled by using the ordinal logistic regression. Extension of dual frame estimation techniques to the case of estimation of proportions when the variable of interest has discrete outcomes is presented. Several estimators for the population proportions are calculated by using two different approaches: "single frame" and "dual frame". To check the efficiency of the proposed procedures in the presence of different sets of auxiliary variables, some Monte Carlo experiments were carried out.

ES147 Room F1 METHODOLOGICAL STATISTICS II Chair: Tommaso Proietti

E133: **Econometric inference theories and multiple use of the same data**
Presenter: **Benjamin Holcblat**, BI Norwegian Business School, Norway
Co-authors: Steffen Groenneberg

In fields that are mainly nonexperimental, such as economics and finance, it is unescapable to compute test statistics and confidence regions that are motivated by previous examinations of the same data. We formalize the inadequacy of the Bayesian and Neyman-Pearson inference theories for such a practice. We show that this is avoidable by presenting elements of a general econometric theory that is immune to multiple use of the same data modulo approximation error.

E1058: **Estimation in first order moving average models using the Hyvarinen scoring rule**
Presenter: **Valentina Mameli**, University of Cagliari, Italy
Co-authors: Monica Musio, Alexander Philip Dawid

Composite likelihoods are surrogates of likelihood useful for estimating parameters in complex high-dimensional statistical models. The increasing importance of this methodology is due to its computational feasibility in a variety of applications. However, for the first order moving average model, the pairwise likelihood method, which is a special case of composite likelihood, has very poor asymptotic efficiency as the moving average parameter tends to the boundary of the parameter space. We consider alternatives to the pairwise likelihood approach, based on the theory of proper scoring rules. We focus on the Hyvarinen scoring rule, which does not require calculation of the normalizing constant. We compare numerically the behaviours of univariate and multivariate Hyvarinen scoring rule estimators with those of the pairwise and full likelihood-based estimators, for the first order moving average model. Simulation studies reveal that the Hyvarinen scoring rule estimators outperform the pairwise estimator in terms of efficiency.

E1177: **Empirical characteristic function tests for GARCH innovation distribution using multipliers**
Presenter: **Maria Dolores Jimenez-Gamero**, University of Sevilla, Spain

Goodness-of-fit tests for the error distribution in GARCH models based on measuring deviations between the empirical characteristic function of the residuals and the characteristic function of the distribution in the null hypothesis have been proposed. Since the asymptotic distribution of these test statistics depends on unknown quantities, their null distribution is usually estimated through a parametric bootstrap. Although very easy to implement, the parametric bootstrap can become very computationally expensive as the sample size or the number of parameters increase. It is proposed to approximate the null distribution through a weighted bootstrap. The estimator is studied both theoretically and numerically. The asymptotic properties are similar to those of the parametric bootstrap but, from a computational point of view, it is more efficient.

E981: **Blind source separation for multivariate conditionally heteroscedastic time series**
Presenter: **Markus Matilainen**, University of Turku, Finland
Co-authors: Klaus Nordhausen, Hannu Oja

The Blind Source Separation (BSS) model for time series assumes that the components of the observed p-variate time series are linear combinations of p independent and stationary time series, and the idea is then to try to transform them back to latent time series (sources). In the so called AMUSE and SOBI solutions the aim is to diagonalize cross-autocovariance matrices with one or several lags, correspondingly, while the unmixing matrix estimate in the classical FOBI solution diagonalizes the covariance matrix and the (marginal) matrix of fourth moments. AMUSE and SOBI find the independent sources if the sources are ARMA processes, for example, but fail for conditionally heteroscedastic GARCH processes. On the other hand, FOBI seems to find solution for GARCH models but, unfortunately, ignores the serial dependence. We introduce a new generalized FOBI method, which jointly diagonalizes several lagged fourth moment matrices and seems to work nicely both for ARMA processes and for GARCH processes. The properties of the new estimate are discussed and illustrated with a simple simulation example. The performances of the SOBI, FOBI and gFOBI estimates are compared in a simulation study.

ES150 Room I1 APPLIED STATISTICS AND DATA ANALYSIS II Chair: Ivette Gomes

E1203: **Evaluation of sampling plans by bootstrapping**
Presenter: **Fernanda Otilia Figueiredo**, Faculdade de Economia da Universidade do Porto and CEAUL, Portugal
Co-authors: Adelaide Maria Figueiredo, Maria Ivette Gomes

In a real framework we merely have access to the measurement results of the quality characteristic on a historical data set of items previously inspected, being in most of the cases difficult to modelling the data, or instead, work with the underlying fitted distribution. To compare and evaluate the performance of specific sampling plans by variables in such cases we can use the bootstrap methodology combined with Monte Carlo simulations. On the basis of simulated and real reference data sets we analyse the performance of the proposed methodology.

E616: **Imprecise event trees framework for evaluation of Russian economy perspectives**
Presenter: **Maria Yudaeva**, Saint-Petersburg State University, Russia
Co-authors: Nikolai Hovanov, Dmitrii Kolesov

Event tree is a powerful tool for decision-making, forecasts, scenarios and negotiations evaluation. But it is difficult in practice to determine transitional probabilities for event tree nodes. Imprecise event trees are constructed on the basis of imprecise knowledge concerning transitional probabilities. Such knowledge allows constructing admissible sets of transitional probabilities for each node and randomization of uncertainty allows obtaining numeric evaluations of transitional probabilities. Finally, evaluations of probabilities of final alternatives and complex event can be calculated and taken into account for decision-making or forecast formation. An illustrative example of Russian economy perspectives evaluations is presented.

E1232: **Survo R - editorial computing environment for data analysis**
Presenter: **Kimmo Vehkalahti**, University of Helsinki, Finland
Co-authors: Reijo Sund
Survo R is a multiplatform, open source implementation of Survo computing environment, now fully incorporated into R. The earlier generations of Survo have been stand-alone implementations, each representing an integrated environment for statistical computing and related areas. The new Survo R raises the setting to a new level, utilizing the various possibilities of R. Features of Survo R include file-based data operations, flexible data preprocessing and manipulation tools, various plotting and printing facilities, a teaching friendly matrix interpreter, so-called editorial arithmetics for instant calculations, a powerful macro language, and plenty of statistical operations for multivariate data analysis and modelling. All this is glued together by an innovative text editor based GUI (originally invented in 1979) that allows us to freely mix commands, data and natural text encouraging towards reproducible research with ideas similar to literate programming. With Survo R, several new features and operations have been developed to achieve an efficient but seamless interplay with R from the GUI.

ES149 Room D1 COMPUTATIONAL STATISTICS II **Chair: Stephane Chretien**

E777: **Tests of fit for the lognormal distribution**
Presenter: **Apostolos Batsidis**, University of Ioannina, Greece
Co-authors: Polychronis Economou, George Tzavelas
The importance of lognormal distribution in formulating real phenomena justifies the necessity for reliable procedures for testing whether a dataset is coming from this distribution. Testing for the adequacy of the lognormal distribution as a model for positive measurements may be carried out by testing goodness-of-fit of the logarithmic data to the normal distribution. Hence, by applying this transformation to the data, all normality tests can be utilized for the purpose of testing the goodness-of-fit to the lognormal distribution. On the basis of the characterization property of the lognormal distribution which states that the Kullback-Leibler measure of divergence between a probability density function and its r-size weighted probability density function is symmetric only for the lognormal distribution, new goodness-of-fit tests for the lognormal distribution are proposed, without reducing the problem to that of testing normality. A simulation study examines the performance of the new procedures in comparison with existing goodness-of-fit tests. Finally, two well-known data sets are used to illustrate the methods developed.

E956: **Bias reduction studies in non-parametric regression with applications**
Presenter: **Cornelia Swanepoel**, North West University Potchefstroom Campus, South Africa
Co-authors: Marike Cockeran
The effect of three improvement methods on recently developed and well-established nonparametric kernel regression estimators is investigated. The improvement methods are applied to the Nadaraya-Watson estimator with cross-validation bandwidth selection as well as with plug-in bandwidth selection, a local linear estimator with plug-in bandwidth selection and a bias-corrected nonparametric estimator, using cross-validation bandwidth selection. The resulting regression estimators are evaluated by minimizing the mean integrated squared error. Two improvement methods are bootstrapped-based. Bagging is an acronym for bootstrap aggregating and is primarily a variance reduction tool. Bragging stands for bootstrap robust aggregating. A robust estimator is obtained by using the sample median over bootstrap estimates instead of the sample mean as in bagging. Boosting is the third improvement method and aims to reduce the bias component of the estimator iteratively. A selection of algorithms, involving the various improvement methods, estimators, models to be estimated, parametric set-ups and bandwidth selection methods (also involving various bootstrap strategies) are presented, displaying clear steps in terms of handling the data to obtain the goals of the study. Results are briefly summarized and illustrated, followed by conclusions and remarks regarding computational problems such as grid issues and bandwidth selection issues.

E1088: **A study of online and batch-wise updating of the EM algorithm for Gaussian mixtures**
Presenter: **Jochen Einbeck**, Durham University, UK
Co-authors: Daniel Bonetti
When estimating Gaussian mixture parameters from sequentially arriving data, it may be inefficient to run an entire EM algorithm, using the full set of available data, after each new influx of data. We consider simple update schemes which allow adjusting the previous EM output by the newly arrived information. Several different update schemes, allowing for online as well as batch-wise updating of (univariate or multivariate) mixture parameter estimates for sequentially recorded data, are investigated. The update schemes are compared in terms of computational performance and the ability to find the overall MLE. Interestingly, it turns out that suitable update schemes may reduce the risk of getting trapped in local maxima of the likelihood, opening the intriguing possibility of artificially performing EM in batches, even though all data are available. The methods are illustrated through simulated data as well as industrial (process control) applications.

E1307: **Detecting change points with a fusion penalty**
Presenter: **Jana Mareckova**, University of Konstanz, Germany
The fusion penalty, particularly an L_1-norm of successive differences of coefficients in time, is applied to detect change points in coefficients of a linear regression model. In the model, the coefficients are allowed to vary across time under the assumption that they change only a few times or not at all. This assumption is captured in an optimization problem by a fusion penalty. The objective function in the optimization problem might be a likelihood function or a sum of squared errors. The main advantage of the fusion penalty over the standard statistical tests which detect change points is the potential to detect a change point at the beginning or at the end of a time series without imposing any assumptions about the number of change points or their positions in the time series. In the simulation study, the introduced optimization problem with a fusion penalty is compared with two standard approaches of change point detection for several data generating processes (DGPs). The DGPs differ in a number of change points, positions of the change points and signal-to-noise ratios.

Authors Index